

Improving personalized tumor growth predictions using a Bayesian combination of mechanistic modeling and machine learning

Pietro Mascheroni¹, Symeon Savvopoulos², Juan Carlos López Alfonso¹, Michael Meyer-Hermann^{1,3,4} & Haralampos Hatzikirou^{5,6}✉

Abstract

Background In clinical practice, a plethora of medical examinations are conducted to assess the state of a patient's pathology producing a variety of clinical data. However, investigation of these data faces two major challenges. Firstly, we lack the knowledge of the mechanisms involved in regulating these data variables, and secondly, data collection is sparse in time since it relies on patient's clinical presentation. The former limits the predictive accuracy of clinical outcomes for any mechanistic model. The latter restrains any machine learning algorithm to accurately infer the corresponding disease dynamics.

Methods Here, we propose a novel method, based on the Bayesian coupling of mathematical modeling and machine learning, aiming at improving individualized predictions by addressing the aforementioned challenges.

Results We evaluate the proposed method on a synthetic dataset for brain tumor growth and analyze its performance in predicting two relevant clinical outputs. The method results in improved predictions in almost all simulated patients, especially for those with a late clinical presentation (>95% patients show improvements compared to standard mathematical modeling). In addition, we test the methodology in two additional settings dealing with real patient cohorts. In both cases, namely cancer growth in chronic lymphocytic leukemia and ovarian cancer, predictions show excellent agreement with reported clinical outcomes (around 60% reduction of mean squared error).

Conclusions We show that the combination of machine learning and mathematical modeling approaches can lead to accurate predictions of clinical outputs in the context of data sparsity and limited knowledge of disease mechanisms.

Plain language summary

Computational methods, such as machine learning and mathematical models, can help doctors and scientists to predict the likely course of a patient's disease, such as tumour growth in a person with cancer. These methods rely on various kinds of data, e.g. imaging, but these data are collected infrequently and do not take into account underlying disease mechanisms. Here, we present a method that combines machine learning and mathematical modeling to improve prediction of tumour growth. We demonstrate our approach on a simulated dataset for glioma and two real cohorts of patients with leukemia and ovarian cancer. Results from the method are in close agreement with actual clinical data for individual patients, suggesting its potential applicability in enabling accurate personalized clinical predictions.

¹Braunschweig Integrated Centre of Systems Biology and Helmholtz Centre for Infectious Research, Braunschweig, Germany. ²KU Leuven, Department of Chemical Engineering, Leuven, Belgium. ³Centre for Individualized Infection Medicine, Hannover, Germany. ⁴Institute for Biochemistry, Biotechnology and Bioinformatics, Technische Universität Braunschweig, Braunschweig, Germany. ⁵Mathematics Department, Khalifa University, Abu Dhabi, UAE. ⁶Centre for Information Services and High Performance Computing, TU Dresden, Dresden, Germany. ✉email: haralampos.hatzikirou@ku.ac.ae

Advances in patient care have led to the availability of large amounts of data, generated by typical examinations, such as blood sample analysis, clinical imaging (e.g., CT, MRI), and biopsy sampling, as well as by innovative ‘-omics’ sequencing techniques^{1,2}. Such clinical data are the cornerstone in the practice of personalized medicine and specifically in the field of oncology^{3,4}. However, this abundance of information comes with multiple issues related to data exploitation and synthesis towards the prediction of pathology dynamics. In particular, we identify the following two major challenges: (C1) First, knowledge of the regulatory mechanisms underlying clinical data is largely lacking, and (C2) second, patient data collection is usually sparse in time, since patient clinical visits/examinations are a limiting factor.

Regarding the challenge (C1), scientists have been long supported by the use of mathematical modeling as a tool to identify causal relationships in the experimental and clinical data, particularly in cancer treatment^{5–7}. Mathematical models allow to propose and test biological hypotheses, analyze the sensitivity of observables with respect to biological parameters, and provide insights into the mechanistic details governing the phenomenon of interest^{8–10}. Although these models can be extremely powerful both in predicting system responses and suggesting new experimental directions, they require adequate knowledge of the underlying biological mechanisms of the analyzed system. Typically, this knowledge is not complete, and only for a limited portion of the involved variables the corresponding mechanistic interactions are sufficiently known. Therefore, even though mathematical models provide a good description of a simplified version of the associated system dynamics, they do not always allow for accurate and quantitative predictions.

On the other hand, machine learning techniques are suitable to deal with the inherent complexity of biomedical problems, but without caring for the knowledge of the underlying interactions¹¹. While mathematical models rely on causality, statistical learning methods identify correlations among data¹². This approach allows to systemically process large amounts of data and infer hidden patterns in biological systems. As a consequence, machine learning-based techniques can provide valuable predictive accuracy upon sufficient training, but do not typically allow for any mechanistic insight into the investigated problem¹³. The overall understanding of the fundamental system dynamics becomes almost impossible, as the chance to generalize the ‘learned’ system behavior. The latter issue is further exacerbated by the (C2) challenge that has to be faced, related to the sparseness of clinical data. In particular for a single patient, such information is only available at a few time-points, corresponding to clinical presentation. To face the two mentioned challenges with the final aim of improving personalized predictions, we propose a novel—to the best of our knowledge—Bayesian method that combines mathematical modeling and statistical learning (BaM³). As a proof-of-concept, the proposed method is tested on a synthetic dataset of brain tumor growth. We analyze the performance of the new approach in predicting two relevant clinical outcomes, namely tumor burden and infiltration. When comparing predictions from the mechanistic model with those from the BaM³ method, we obtain improved predictions for the vast majority of virtual patients. We also apply the approach to a clinical dataset of patients suffering from chronic lymphocytic leukemia (CLL). The BaM³ method shows excellent agreement between the predicted clinical output and the reported data. Finally, as an additional test case, we show how the proposed methodology can be used to assess the time-to-relapse (TtR) in a dataset of ovarian cancer patients.

Methods

Formal definition of BaM³. We start by assuming a random variable (r.v.) triplet $(\mathbf{Y}, \mathbf{X}_m, \mathbf{X}_u)$ that denotes the system’s modelable \mathbf{X}_m , unmodelable \mathbf{X}_u variables/data (e.g., patient’s age or sex, results of different ‘-omics’ techniques, etc.) and the associated observed clinical outputs \mathbf{Y} . We then introduce t_0 as the clinical presentation time of a patient at which the patient-specific r.v. realizations $(\mathbf{X}_m = \mathbf{x}_m^*, \mathbf{X}_u = \mathbf{x}_u^*)$ are obtained. The overall goal of the method is to predict the patient’s clinical outputs by an estimate $\mathbf{Y} = \hat{\mathbf{y}}$ at a certain prediction time t_p . The true clinical outputs of the patient will be denoted as \mathbf{y} . Moreover, we consider the existence of an N -patient ensemble dataset $(\mathbf{y}, \mathbf{x}_m, \mathbf{x}_u)$. In this dataset all the variables (i.e., modelables, unmodelables, and clinical outputs) are recorded at the time of diagnosis t_d , which might differ from one patient to another. Both t_0 and t_d are calculated from the onset of the disease. We introduce two distinct times to account for the variability of the disease stage among different patients (t_d) and the time at which a specific patient is presented to the clinic (t_0) (see the corresponding Fig. S1).

The core idea of the method is to consider the predictions of the mathematical model $p(\mathbf{Y} = \hat{\mathbf{y}} | \mathbf{X}_m = \mathbf{x}_m^*)$ as an informative Bayesian prior of the posterior distribution $p(\mathbf{Y} = \hat{\mathbf{y}} | \mathbf{X}_m = \mathbf{x}_m^*, \mathbf{X}_u = \mathbf{x}_u^*)$. We can prove that:

$$\begin{aligned} p(\mathbf{Y} = \hat{\mathbf{y}} | \mathbf{X}_m = \mathbf{x}_m^*, \mathbf{X}_u = \mathbf{x}_u^*) &\propto \\ p(\mathbf{Y} = \hat{\mathbf{y}} | \mathbf{X}_m = \mathbf{x}_m^*) p(\mathbf{Y} = \hat{\mathbf{y}} | \mathbf{X}_u = \mathbf{x}_u^*). \end{aligned} \quad (1)$$

The implementation of the BaM³ method therefore reduces to the calculation of the aforementioned probability distributions. Although the prediction of the probability distribution function (pdf) of the clinical outputs is rather straightforward for the mathematical model, obtaining the pdf of the patient’s unmodelable data is not trivial. To retrieve the latter, we use a density estimator method upon the patient ensemble dataset to derive $p(\mathbf{Y}, \mathbf{X}_u)$, and then consider the patient-specific realization $\mathbf{X}_u = \mathbf{x}_u^*$. For further details about method derivation and estimators of performance, see Supplementary Note 1.

Testing the method on synthetic glioma growth. The equations of the selected mathematical model^{14,15} (‘full model’) describe the spatio-temporal dynamics of tumor cell density (c), oxygen concentration (n), and vascular density (v) in the context of glioma tumor growth. The full model includes the variation of cell motility and proliferation due to phenotypic plasticity of tumor cells induced by microenvironmental hypoxia^{16–20}. It also accounts for oxygen consumption by tumor cells, formation of new vessels due to tumor angiogenesis, and vaso-occlusion by compression from tumor cells^{15,21,22}. We generate $N = 500$ virtual patients by sampling the parameters of the full model from a uniform distribution over the available experimental range. We consider the tumor cell spatial density c to be the modelable variable. Moreover, we treat the integral over the tissue of oxygen concentration and vascular density, denoted as \bar{n} and \bar{v} , respectively, as the unmodelable quantities. Starting from the same initial conditions, we simulate the behavior of each virtual patient for 3 years, storing the values of all variables at each month. As sketched in Fig. 1, we use the modelable variable to setup a mathematical model. In particular, we take $c(x, t_0)$ at a specific time-point, the clinical presentation time t_0 , and use it as the initial condition for a Fisher-Kolmogorov equation^{23–27} (‘FK model’). We use this model to predict tumor behavior at a specific time in the future, the prediction time t_p . For each simulated patient we calculate the tumor size (TS) and infiltration width (IW). In parallel, for each patient we evaluate the diagnosis time t_d as a random number in the interval $[t_0 - 6, t_0 + 6]$ (in the unit

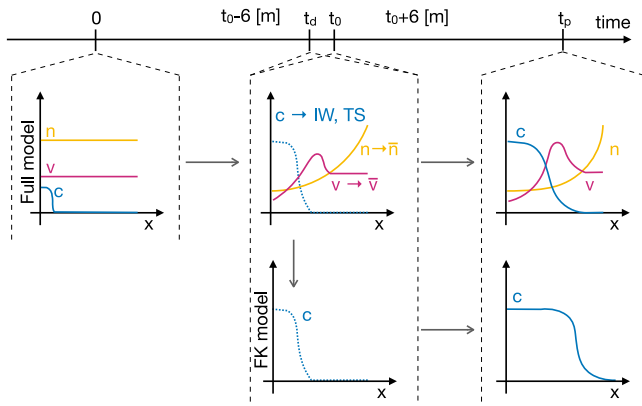


Fig. 1 Workflow for generating the synthetic data. The full model is initialized at $t = 0$ and used to simulate the spatio-temporal variation of tumor density c , oxygen concentration n , and functional tumor-associated vasculature v . For each virtual patient, two clinical outputs are tracked, i.e., the infiltration width (IW) and tumor size (TS), and two unmodelables recorded, i.e., the oxygen and vasculature integral over the tissue \bar{n} and \bar{v} , respectively. These quantities are generated at the time of clinical presentation t_0 every time the method is applied. To generate the patient ensemble, they are also generated at the diagnosis time t_d . The latter is assumed to be a random time between $t_0 \pm 6$ months. Then, the spatial profile of tumor concentration at time t_0 is used as the initial condition for the Fisher-Kolmogorov (FK) model, which is in turn used to simulate tumor growth until the prediction time t_p .

of months), and collect the values of modelables, unmodelables, and clinical outputs at this time to build the patient ensemble. Given the patient-specific modelable and unmodelable variables ($c(x, t_0)$ and \bar{n}, \bar{v} , respectively) at the clinical presentation time t_0 , the BaM³ method therefore produces the probability of observing the TS and IW at a specific prediction time t_p .

Mathematical models for glioma growth. The system variables are the density of glioma cells $c(x, t)$, the concentration of oxygen $n(x, t)$, and the density of functional vasculature $v(x, t)$ ^{14,15}. For simplicity we consider a one-dimensional computational domain. We normalize the system variables to their carrying capacity and write the system as

$$\frac{\partial c}{\partial t} = D \frac{\partial^2}{\partial x^2} \left(\frac{\alpha}{\alpha_0} c \right) + b \frac{\beta}{\beta_0} c(1 - c), \tag{2}$$

$$\frac{\partial n}{\partial t} = D_n \frac{\partial^2 n}{\partial x^2} + h_1 v(n_0 - n) - h_2 cn, \tag{3}$$

$$\frac{\partial v}{\partial t} = D_v \frac{\partial^2 v}{\partial x^2} + g_1 \mathcal{H}(n - n_0)v(1 - v) - g_2 vc^\delta. \tag{4}$$

Here $\mathcal{H}(\cdot)$ is a sigmoidal function ($\mathcal{H}(x - x_0) = 1/(1 + \exp(b^*(x - x_0)))$, with $b > 0$ being a constant) allowing for tumor angiogenesis in hypoxic conditions, i.e., for $n < n_0$ where n_0 is the hypoxic oxygen threshold. Then, the functions $\alpha = \alpha(n)$ and $\beta = \beta(n)$ account for the dependence of cellular motility and proliferation on the oxygen level, respectively^{16,17,19}. They are defined as:

$$\alpha = \frac{\lambda_1 - n}{(\lambda_2 - 1)n + \lambda_1}, \tag{5}$$

$$\beta = \frac{\lambda_2 n}{(\lambda_2 - 1)n + \lambda_1}. \tag{6}$$

When the oxygen level is fixed to the maximum level $n = 1$ in the

tissue $\alpha = \alpha_0$ and $\beta = \beta_0$, so that the equation for c reduces to

$$\frac{\partial c}{\partial t} = D \frac{\partial^2 c}{\partial x^2} + bc(1 - c), \tag{7}$$

which we denote in the rest of the manuscript as the Fisher-Kolmogorov (FK) model for tumor cell density. We remark that Eq. (7) has been extensively used to predict untreated glioma kinetics based on patient-specific parameters from standard medical imaging procedures^{23–27}.

Eqs. (2)–(4) define an extended version of the FK equation, enriched with nonlinear glioma cell diffusion and proliferation terms. The latter terms depend on the oxygen concentration in the tumor microenvironment, which is in turn coupled to cell density through the oxygen consumption term. The functional vasculature density controls the supply of oxygen to the tissue. Blood vessel density increases due to tumor angiogenesis and decreases because of vaso-occlusion by high tumor cell density. The values of the parameters used in the simulations and their descriptions are given in Table S1. In addition, a typical full model simulation is shown in Fig. S3 for a representative patient.

We solve the system in Eqs. (2)–(4) by imposing the initial conditions:

$$c(x, 0) = c_0 \mathcal{H}(x - \epsilon) \text{ in } 0 \leq x \leq L, \tag{8}$$

$$n(x, 0) = n_0 \text{ in } 0 \leq x \leq L, \tag{9}$$

$$v(x, 0) = v_0 \text{ in } 0 \leq x \leq L, \tag{10}$$

where the positive parameters c_0, n_0 , and v_0 are the initial density of glioma cells spatially distributed in a segment of length ϵ , the density of functional tumor vasculature, and the oxygen concentration, respectively. Then, $L > 0$ is the length of the one-dimensional computational domain. In addition, we consider an isolated host tissue in which all system behaviors arise solely due to the interaction terms in Eqs. (2)–(4). This assumption results in no-flux boundary conditions of the form:

$$\frac{\partial c}{\partial x}(0, t) = \frac{\partial n}{\partial x}(0, t) = \frac{\partial v}{\partial x}(0, t) = 0, \tag{11}$$

$$\frac{\partial c}{\partial x}(L, t) = \frac{\partial n}{\partial x}(L, t) = \frac{\partial v}{\partial x}(L, t) = 0. \tag{12}$$

Both the full and FK models are used to calculate two clinical outputs, namely the tumor IW and TS. The IW at a specific time is defined by the difference between the points where glioma cell density is 80% and 2% of the maximum cellular density. In turn, the TS is obtained by integrating the spatial profile of tumor density and dividing it for the maximum value of the latter.

We run the full model and simulate the growth of the tumor for N patients, each one from a parameter set taken randomly from a uniform distribution over the parameter range. We run simulations for $N = 50, 100, 250$, and 500 with 10 repetitions within each N -case. To generate the patients, we vary five parameters in the list in Table S1, namely the tumor motility D , proliferation rate b , oxygen consumption h_1 , vascular formation, and occlusion rates g_1 and g_2 , respectively. Then, we use the tumor density at the time of clinical presentation, t_0 , as the initial condition for the FK model. The latter model is employed to generate predictions at the prediction time t_p . We also consider the unmodelable variables and clinical outputs at the diagnosis time t_d , taken randomly between $t_0 \pm 6$ months, to build the patient ensemble. Finally, we use the results of the full model in terms of clinical outputs as the ground truth to be compared with the predictions of the FK model alone and with the ones obtained by the BaM³ method.

Probability distribution from the FK model. As described in the previous sections, we take the spatial profile of tumor density at the clinical presentation time t_0 as the initial condition of the FK model. We use the latter mathematical model to run simulations over the whole parameter set for cell motility D and proliferation rate b . Then, we define the model-derived pdf as in the following. For each couple of clinical outputs IW^* and TS^* we calculate the area $A_\alpha(IW^*, TS^*)$ over the (IW, TS) plane as $A_\alpha = [(1 - \alpha)IW^* < IW < (1 + \alpha)IW^*], (1 - \alpha)TS^* < TS < (1 + \alpha)TS^*]$, where α is a given tolerance (here set $\alpha = 0.05$). Then, we calculate the pdf by normalizing A_α by the total area of predicted IW and TS values. We store the value of the probability for each patient at the different prediction times and use it to compute the expected value of the model pdf.

Probability distribution of the unmodelables from the full model. To retrieve the data-derived pdf we use a normal kernel density estimator (KDE)^{28,29}, which depends upon all the data points in the patient ensemble. Briefly, the method estimates the joint probability $p(IW, TS, \bar{n}, \bar{v})$ from which the ensemble entries are drawn through the sum of a kernel function over all the occurrences of the dataset. The kernel function is characterized by a hyperparameter, the bandwidth \tilde{h} , which we assume according to Silverman’s rule of thumb

$$\tilde{h}_i = \sigma_i \left[\frac{4}{(d+2)n} \right]^{\frac{1}{d+4}}, \quad i = 1, 2, \dots, d, \quad (13)$$

where d is the number of dimensions, n is the number of observations, and σ_i is the standard deviation of the i th variate³⁰. After calculating $p(IW, TS, \bar{n}, \bar{v})$, we specify the realization of a specific patient and calculate the value of $p(IW = IW^*, TS = TS^*, \bar{n} = \bar{n}^*, \bar{v} = \bar{v}^*)$ over the (IW, TS) space of the estimated clinical outputs.

Scoring glioma growth predictions. We calculate for each patient the relative errors d_m and d_b as described the main text. To assess how the BaM³ method has changed the prediction of the mathematical model, we compare the latter quantities: if $|d_b - d_m| \leq \epsilon d_m$, then there was no change; if $d_b > (1 + \epsilon)d_m$, then the method deteriorated the prediction of the model; if $d_b < (1 - \epsilon)d_m$, then the method improved the prediction of the model. Here, ϵ is a tolerance used for the comparison, taken to be $\epsilon = 0.05$.

Calculation of the effective variance. To calculate the effective variance s , we first calculate the mixed central moments Σ_{ij} of the pdf of interest according to the formula

$$\Sigma_{ij} = \int_{y_1} \int_{y_2} (y_1 - \mu_1)(y_2 - \mu_2)f(y_1, y_2) dy_1 dy_2, \quad (14)$$

where y_1 and y_2 are the clinical outputs (IW and TS , respectively) and μ_1, μ_2 the expected values of the corresponding variables. The elements of Σ form a symmetric two-dimensional matrix, for which we calculate the determinant. We define the effective variance s as the natural logarithm of the latter determinant. In Eq. (14) we consider $f(y_1, y_2)$ to be the pdf from the mathematical model or from the BaM³ method depending on whether we are interested in the effective variance s_m or s_b , respectively.

Pdf from the two-compartment model in CLL. Messmer and colleagues³¹ measured the fraction of labeled B-CLL cells in a cohort of 17 CLL patients that were administered deuterated water. They calibrated a two-compartment model on each patient and were able to reproduce the kinetics of labeled cells over a long time. We adopt their model and use it to generate the pdf for the

CLL example. The fraction of labeled cells over time is calculated through the expression

$$f(t) = h(t) + f(0)e^{\frac{-bt}{v_r}} + \frac{e^{\frac{-bt}{v_r}}(g(0) - h(t))}{v_r - 1} + \frac{e^{-bt}(h(t) - g(0))}{v_r - 1} \quad (15)$$

where $g(0)$ is the initial fraction of cells in the first compartment, b the fractional cell birth, v_r the relative size of the compartments, and $h(t)$ is the deuterated water concentration of the body over time. The latter is a function of the fractional daily water exchange f_w . We refer the interested reader to the supplementary information of Messmer et al.³¹ for a more detailed description of the model and a full account of the model parameters. In this work we focus on three quantities, namely b , v_r , and f_w , and run the model in Eq. (15) over the experimental range. This range was obtained by considering the patient-specific fitting performed by Messmer and colleagues and selecting the minimum and maximum values. We evaluate the fraction of labeled cells at day 50, f_{50} , and build the probability distribution from its histogram, by counting the number of occurrences of a given f_{50}^* for $\min(f_{50}) < f_{50}^* < \max(f_{50})$ and then normalizing the result. For the CLL example, all the patients start with the same initial fraction of labeled cells, set to zero.

Pdf from the patients’ unmodelables in CLL. The data-derived pdf in the CLL example is obtained from four unmodelable quantities that are measured for each patient during the study. We consider all the possible combinations of unmodelables and calculate the mean squared error (MSE) for each case. The scatter plot in the same picture refers to the case in which the CD38 expression ($x_{u,1}$), age ($x_{u,2}$), growth rate of white blood cells ($x_{u,3}$), and V_H mutation status ($x_{u,4}$) are added consecutively with the specified order. As in the glioma example, we build the subdataset (y, \mathbf{x}_u) , where y and $\mathbf{x}_u = (x_{u,i})$ are the f_{50} and unmodelable variables of each patient, respectively, and apply the KDE using Silverman’s rule for the hyperparameters. The requested pdf, i.e., $p(Y = \hat{y} | \mathbf{X}_u = \mathbf{x}_u^*)$ is obtained by conditioning the probability from the KDE with the realizations of the unmodelables of the specific patient and calculating the result over the range of the estimated clinical output $\hat{y} = f_{50}$.

Mathematical model for ovarian cancer. We assume the total number of tumor cells T to be composed of the sensitive S and resistant R subpopulations. The latter are described by the following system of ordinary differential equations (ODEs):

$$\dot{S} = \gamma S - \delta S - \tau S, \quad (16)$$

$$\dot{R} = \gamma R - \lambda \delta R + \tau S, \quad (17)$$

where γ is the tumor net growth rate, $\delta = \delta(t)$ is the death rate induced by chemotherapy, τ is the mutation rate from sensitive to resistant cells, and λ is a factor that accounts for reduced death by therapy in resistant cells. As detailed in Fig. S12, the treatment is composed of three phases: first, the patients undergo different cycles of NACT; then, surgery is performed. The latter reduces the total tumor volume, irrespective of cells being sensitive or resistant, of a factor β . Finally, another series of chemotherapy cycles is performed. During chemotherapy, $\delta = \delta_0$, whereas we set this parameter to zero after chemotherapy and until tumor relapse. The latter condition occurs when T reaches the value T_R .

Equations (16) and (17) can be analytically integrated, and their results used to build the probability distribution of the clinical output—TtR, in this case. To obtain the pdf from the model, we calculate the time the tumor takes to reach the cell number at relapse T_R starting from the cell number after therapy.

We perform this calculation using the initial tumor cell number of each patient, and by varying both the initial fraction of S cells, x_0 , and the chemotherapy-induced death rate, δ_0 . We then obtain the patient-specific probability distribution from the histogram of TtR, similarly to what is done in the previous section for CLL. For x_0 , we select a range between 0.4 and 0.9, accounting for tumors with different initial degrees of intrinsic resistance³². For δ_0 , we first use a uniform distribution between 0.1 and 10 days⁻¹, accounting for a wide variation in death rates. The latter choice produces an almost flat distribution for the clinical output (see Fig. S13). To improve the mathematical model parametrization, we use the information about the tumor volume change after the first cycle of chemotherapy, which is included in the dataset. By fitting T obtained from Eqs. (16) and (17) to the observed volume change, we find a value of δ_0 for each patient in the dataset³². We take the mean value of these rates and use it to update the model pdf (see Fig. S14). We consider a range for δ_0 that is centered around its mean value across the patients, within an interval of $\pm 40\%$. Selecting other ranges provides similar results, however, a variation of 40% returns the lowest MSE. Analytical integration of Eqs. (16) and (17), as well as additional details about model parametrization are available in Supplementary Note 2.

Unmodelable variable for the ovarian cancer study. We build the data-derived pdf for the ovarian cancer example by exploiting the information about the age of the patients at diagnosis. Similarly to what done in the previous test cases, we first build the sub-dataset (TtR, A) by entering the information of each patient (here, A is the patient age). Then, we apply the KDE using Silverman's rule to estimate the bandwidth and calculate the joint probability $p(\text{TtR}, A)$. The data-derived pdf for each patient $p(\text{TtR}|A^*)$ is finally obtained over the domain of the clinical output TtR by considering the patient-specific age $A = A^*$.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Results

We introduce the key ideas of the proposed methodology in the context of brain tumor growth, leaving the full derivation of the equations and their general form to the Supplementary information (see Supplementary Note 1).

Gliomas are aggressive brain tumors generally associated with low survival rates³³. One of the most important hallmarks of this type of tumors is its invasive behavior, combined with a marked phenotypic plasticity and infiltrative morphology¹⁶. The clinical needs led to the development of several mathematical models to support clinicians in the treatment of the disease³⁴. As a first test case, we synthetically generate a dataset of glioma patients using a system of recently published^{14,15} partial differential equations (PDEs). This complex mathematical model ('full model', in the following) provides a set of *in silico* patients, which represents our synthetic reality and serves as a benchmark to evaluate the performance of the proposed BaM³ method. Our goal is to obtain a personalized prediction of the clinical observables of the patients, combining their 'modelable' and 'unmodelable' information. A simplified mathematical model (with respect to the full model used to generate the patients) is used to generate predictions of clinical outputs starting from the modelable variables. In turn, a machine learning algorithm produces predictions of the same clinical outputs leveraging on the information contained in the unmodelables. As displayed in Fig. 2, the core idea of the BaM³ method is to use the results of the mathematical model to guide the predictions of machine learning. In more technical terms, the

pdf obtained from the mathematical model ('model-derived pdf, in the following) works as a Bayesian prior that multiplies the pdf obtained from a nonparametric regression algorithm ('data-derived pdf'). The product of these two pdfs returns an estimate of the pdf for the clinical outputs of interest. More details about the formal definition of the BaM³ method and the mathematical details are available in the 'Methods' and Supplementary Note 1.

Improving predictions of synthetic glioma growth. For this first test case, we deal with two clinical observables, i.e., the TS and IW. The first quantity is related to tumor burden, whereas the second accounts for tumor infiltration in the host tissue. The modelable variable is the tumor cell density c , whereas we consider the amount of oxygen \bar{n} and vasculature \bar{v} in the tissue to be the patients' unmodelables (see the 'Methods'). Then, given the patient-specific modelable and unmodelable variables at the clinical presentation time t_0 , the BaM³ method produces the probability of observing certain values of the clinical outputs at a specific prediction time t_p (see Fig. 2). The data-derived pdf is obtained through a normal KDE^{28,29,35} incorporating the information about the patient ensemble. The latter is generated from the full model, at the diagnosis time t_d . Then, the model-derived pdf is calculated using the simplified mathematical model for each patient. In particular, we use the FK model²³ to produce a map of possible IW and TS starting from the tumor cell density of each virtual patient (see the 'Methods' for further information about the KDE and modeling steps).

Figure 3a–c shows the results of applying the BaM³ method to a representative patient. We select a clinical presentation time $t_0 = 24$ months and a time of prediction $t_p = 9$ months. The model-derived pdf obtained from the FK model is shown in Fig. 3a. Interestingly, the prediction of the model in that particular case shows two peaks, one with low TS and high IW and another with opposite properties. We calculate the expected values of TS and IW from the pdf obtained with the FK model and compare it to the 'true' values given by the full model. As shown in the plot, for this patient the presence of a bimodal distribution shifts the expected values far from the true ones. We enforce the BaM³ method making use of the probability calculated from the KDE, shown in Fig. 3b. The latter pdf takes into account the correlations between the clinical outputs and the unmodelable variables present in the patient ensemble. For this patient, the unmodelable distribution selects the probability mode closer to the true IW and TS values, as displayed in Fig. 3c (another example for a different patient is given in Fig. S4). These results evidence the ability of the proposed method to correct the predictions obtained by using exclusively the mathematical model, and to produce an expected value of the pdf that is closer to the ground truth.

We apply the BaM³ method for two clinical presentation times at 12 and 24 months, and compare its outcomes with those provided by the full model at increasing prediction times (Fig. 3d, i). For each patient, we calculate the relative error between the predicted clinical outputs obtained from the full model and the expected values of the pdf calculated from the FK model (d_m) and after implementing the BaM³ method (d_b). These nondimensional errors are calculated as

$$d_k = \sqrt{\frac{1}{L} \sum_{i=1}^{N_y} \left(1 - \frac{\langle y_i^k \rangle}{y_i^r} \right)^2} \quad (18)$$

where $k = m, b$, $\langle y_i^k \rangle$ are the expected observable values ($i = \text{IW}, \text{TS}$) calculated from the FK model and the BaM³ method, and y_i^r are the observable values obtained from the full model. We calculate the errors d_m and d_b for each patient at different prediction times. Then, we compare the corresponding errors and evaluate if the BaM³ method improved, deteriorated, or left

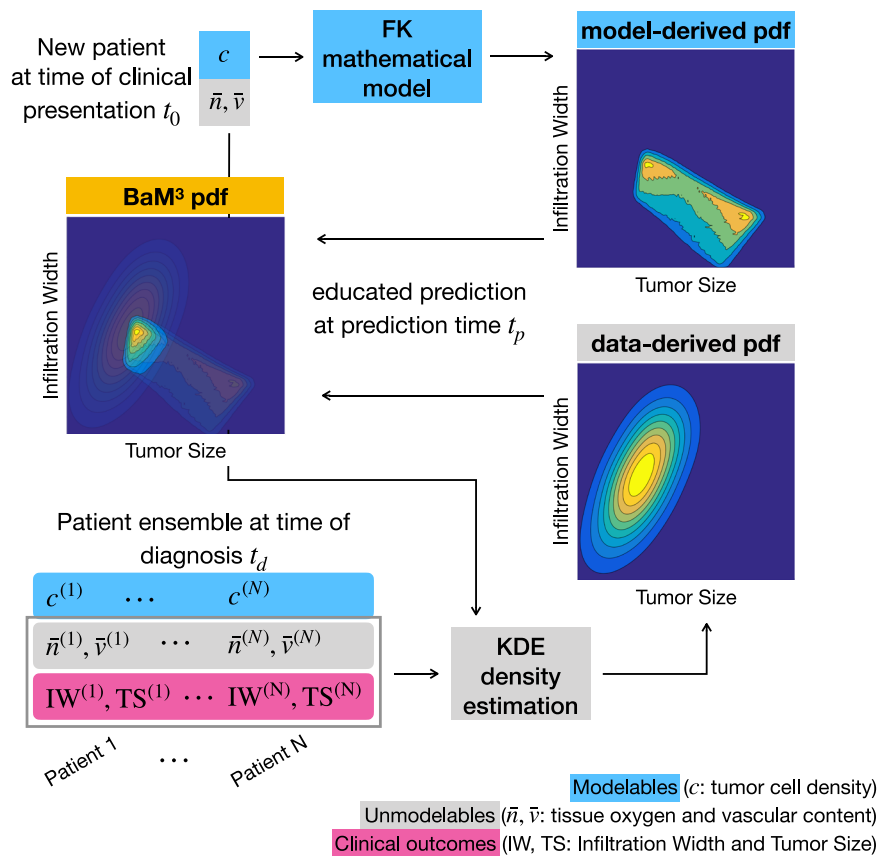


Fig. 2 Application of the BaM³ methodology to the prediction of glioma growth. Given a new patient characterized by modelable (c) and unmodelable (\bar{n}, \bar{v}) variables at the clinical presentation time t_0 , the goal of the BaM³ approach is to find an estimate for the clinical observables IW and TS at the prediction time t_p . The modelable variable is used in the FK mathematical model to generate the model-driven pdf. In turn, the unmodelable variables are injected into a density estimation method, which contains the information about the patient ensemble at the diagnosis time t_d . This step provides the data-driven pdf, which is used to correct the predictions from the mathematical model. Application of the BaM³ method results in a new pdf for the clinical observables at the prediction time t_p .

unchanged the prediction from the FK model, i.e., $d_b < d_m$, $d_b > d_m$, or $d_b \sim d_m$, respectively (see the ‘Methods’). We denote the ratio of improved, unchanged, and deteriorated cases with respect to the total number of simulated patients as S_i , S_u , and S_d , respectively.

Both the relative errors d_m and d_b increase for increasing prediction times, as shown in Fig. 3d, g for the two clinical presentation times considered. However, after applying the BaM³ method the errors decrease, especially at later times. In general, it is possible to notice an improvement both in terms of median values and sparseness of the data. Interestingly, the relative error obtained from the BaM³ method increases at a lower rate if compared to the relative error obtained from the FK model.

We also calculate the effective variance of the predictions as the logarithm of the determinant of the covariance matrices relative to the model and BaM³ pdfs, (identified by s_m and s_b , respectively; see the ‘Methods’). This quantity reflects the spreading of the pdfs over the (TS, IW) plane, with higher values denoting more uncertainty in the predictions. For both clinical presentation times (Fig. 3e, h), the BaM³ method provides thinner pdfs, more centered around their expected value with respect to the FK model-derived case.

Finally, the stacked bars in Fig. 3f, i show that BaM³ performs well at later prediction times, and especially remarkably well (improvement ratio S_i close to 1) at the latest clinical presentation time $t_0 = 24$ months. For $t_0 = 12$ months (Fig. 3f), the proposed method is not able to improve predictions until a prediction time

of 6 months. Then, for $t_p = 6, 9,$ and 12 months the advantages of using BaM³ over the FK model are unambiguous. On the other hand, for the clinical presentation time of 24 months (Fig. 3i) both S_u and S_d decrease significantly for prediction times equal to or greater than 3 months. The ratio of improved cases S_i reaches almost 100% at each of the last three prediction times, clearly overcoming the results of the FK model. The error bars in Fig. 3f, i denote the variability in the results that is obtained by replicating the study 10 times, each with $N = 500$ randomly generated patients. Fig. S5 shows similar results when decreasing the number of patients. Notably, the scores for $N = 500, 250, 100,$ and 50 are very close, slightly improving with increasing the number of patients. The variability in the 10 replicates also decreases for higher values of N .

We also calculate the prediction scores using the distribution mode to generate the scores instead of the expected value (see Fig. S6). When the pdfs display multiple maxima we consider the average of the relative errors between the values of the full model (i.e., the synthetic reality) and the different peaks. The performance of the BaM³ method sensibly degrades with respect to using the expected value. Improvement in predictions is observed only for later clinical presentation and prediction times.

In summary, the BaM³ method is able to correct the FK model predictions for most of the patients, particularly at later clinical presentation and prediction times. The improvement in the prediction occurs by: (i) decreasing the median relative error between expected observable values and ground truth; (ii)

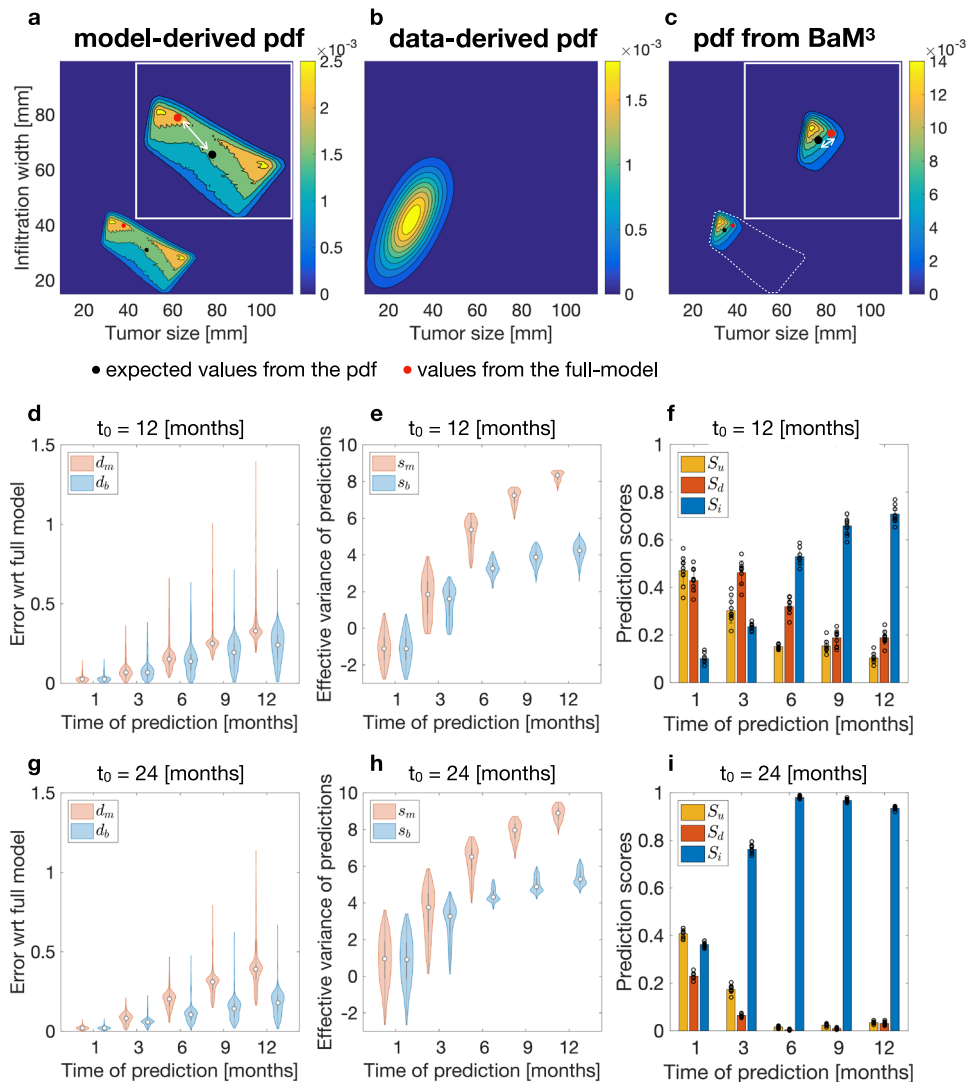


Fig. 3 BaM³ outperforms mathematical modeling for an in silico patient cohort. **a, b** Application of the BaM³ method to a representative patient. The clinical presentation time t_0 is of 24 months and the prediction time t_p is 9 months. **a** Pdf obtained from the FK model, plotted over the (TS, IW) space. The inset shows an enlargement of the probability region, in which the white arrow marks the error between the clinical output predicted by the full model (red dot) and the expected value of the distribution (black dot). **b** Data-derived pdf for the specific patient calculated through the kernel density estimator trained over the patient ensemble. **c** Probability distribution obtained from the BaM³ method. The dashed line shows the original probability obtained from the FK model. Relative errors between the clinical outputs obtained from the full model and the expected values given by the FK model (d_m), and after the application of the BaM³ method (d_b), for a clinical presentation time of 12 and 24 months ((**d**) and (**g**), respectively). Effective variance of the predictions for the corresponding clinical presentation times ((**e**) and (**h**), respectively). Prediction score, as the ratio of cases for which the BaM³ method improved (S_i), did not change (S_d), or deteriorated (S_d) the model predictions. The error bars show the standard deviation of the results by replicating the method over 10 different sets (black circles) of $N = 500$ patients ((**f**) and (**i**), respectively).

decreasing the rate at which the error increases with prediction time; and (iii) decreasing the variance associated with the probability distributions.

BaM³ performance depends on the clinical output. Even though the BaM³ method performs well for the majority of patients, there are some cases for which it fails to improve the predictions of the mathematical model. We analyze the failure cases by splitting the errors d_m and d_b into the two partial errors

$$\Delta IW_k = \sqrt{\left(1 - \frac{\langle IW^k \rangle}{IW^r}\right)^2}, \quad \Delta TS_k = \sqrt{\left(1 - \frac{\langle TS^k \rangle}{TS^r}\right)^2} \quad (19)$$

where $k = m, b$, $\langle IW^k \rangle$, and $\langle TS^k \rangle$ are the expected values of the clinical outputs obtained from the mathematical model and BaM³

pdfs ($k = m$ and b , respectively), and IW^r and TS^r are the values of these quantities from the full model. Figure 4 shows how these partial errors are distributed over the presentation and prediction times. The dashed line in the plots highlights the neutral boundary, where the partial errors of the FK model and BaM³ method are equal. Above this line, the proposed BaM³ method deteriorates the model predictions, whereas under that line the BaM³ method improves predictions. The red dots in the scatter plots represent the patients for which the BaM³ method fails ('failure cases', in the following). After a prediction time of 1 month, in which a characteristic pattern is not evident, the plots highlight that failure cases are generally associated with regions where the BaM³ method under-performs to the FK model with respect to TS ($\Delta TS_b > \Delta TS_m$). Interestingly, the same failure cases belong to regions in which $\Delta IW_b < \Delta IW_m$: the BaM³ method is

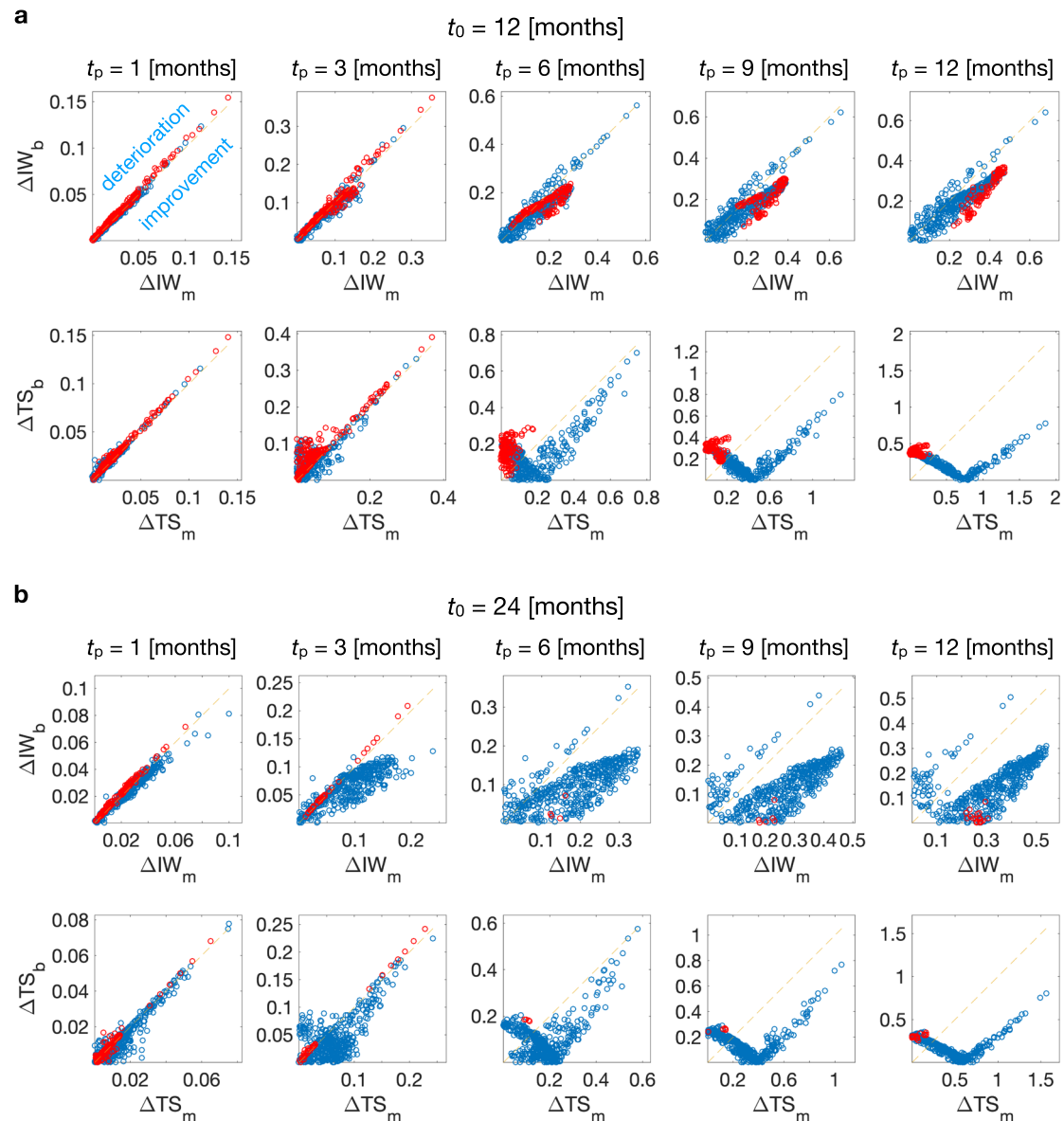


Fig. 4 Evaluating BaM³ performances depending on clinical outputs. Scatter plots of the partial errors ΔIW_i and ΔTS_i , over different prediction times t_p ($i = m$ and b , for the quantity related to the mechanistic model or the BaM³ method, respectively). Results for presentation times of $t_0 = 12$ months (**a**) and $t_0 = 24$ months (**b**) are shown. Each red dot corresponds to a patient for which the BaM³ method fails in improving the FK model predictions. Regions over the dashed line correspond to areas in which the BaM³ method deteriorates the FK model predictions, while regions under that line show areas in which it improves the FK predictions.

improving the IW predictions and at the same time deteriorating the TS predictions. This happens for both $t_0 = 12$ and 24 months, however, the number of failure cases is considerably higher for the earlier presentation time. For the specific case under consideration, lower performance of the BaM³ method is therefore associated to its inability in correcting the FK model predictions for TS, with a tendency that improves for the later presentation time due to strong corrections for IW.

Transient behavior of the unmodelable distribution is associated to limited improvements. To investigate the reasons for the poor performance of the BaM³ method in improving the predictions for one of the clinical observables, we analyze the behavior of the pdf arising from density estimation, i.e., the data-derived pdf. Figure 5 shows the temporal evolution of this quantity for different clinical presentation times t_0 . Figure 5a shows a plot of the

unmodelable pdf for a representative patient over the clinical output space. From a pdf that covers a limited region in the (IW, TS) plane, the probability distribution spreads over a broader area as the presentation time increases. The center of mass of the distribution, however, tends to converge to a more specific region as time progresses. This is more evident in Fig. 5b, c, showing the marginal probabilities for IW and TS, calculated from the distribution in Fig. 5a. The marginal distributions become broader for both IW and TS, but in the first case their peak stabilizes at later t_0 times. On the contrary, the peak of the marginal probability for TS moves towards larger values at higher times. To quantify this behavior across the different patients, we then evaluated the degree of overlap between the marginal probabilities at two subsequent t_0 . Results from this calculation are plotted in Fig. 5d, e for the overlap between the distributions at presentation times t_0 of 12 and 18 months and between 24 and

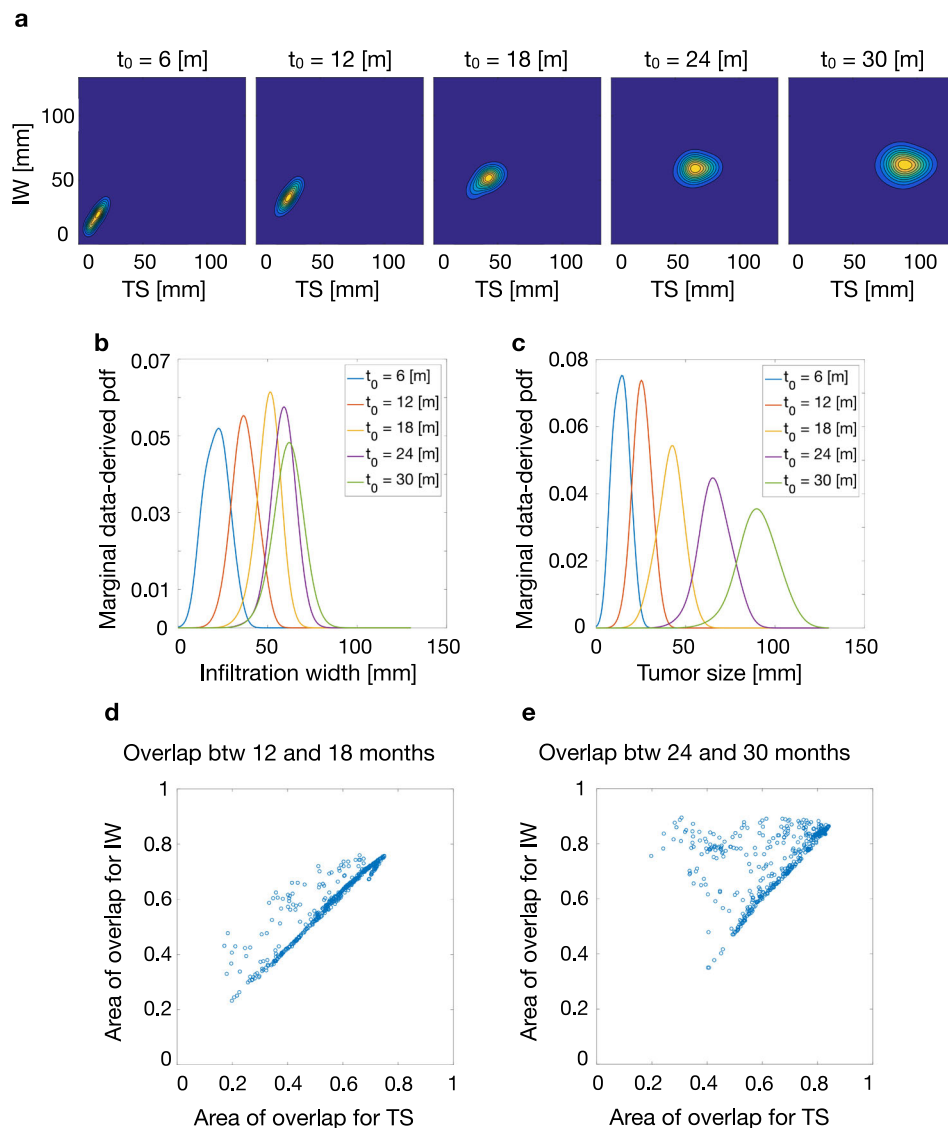


Fig. 5 Temporal evolution of the unmodelable pdf from the KDE step. **a** Data-derived pdf for a representative patient at different clinical presentation times, plotted over the (IW,TS) space. Marginal probabilities from the plots in **(a)**, showing the evolution of the probability distribution for IW **(b)** and TS **(c)**. Scatter plots representing the degree of overlap between the marginal probabilities for IW and TS between clinical presentation times t_0 of 12 and 18 months **(d)** and of 24 and 30 months **(e)**.

30 months. Here, the degree of overlap is calculated as the area of overlap for the IW and TS marginal distributions. Values close to one represent maximum overlap, whereas values near zero are associated to poor overlap between the two marginal pdfs. In a rough approximation, when this overlap score is high the marginal pdf is close to a steady state (since the pdf has not moved over time), and vice versa. For the earlier times in Fig. 5d, the patients are mostly scattered along a line of increasing IW, and TS, with points where the overlap is poor (close to 0.4 in certain regions). On the other hand, for the later times in Fig. 5e the patients are shifted towards higher values of overlap. Moreover, a horizontal line of high overlap for the IW output is visible for a large patient ensemble, pointing to a stabilization towards a steady state for the IW at later presentation times. This explains the lower performances of BaM³ method at $t_0 = 12$ months, since the pdf from the KDE that should correct the model predictions is projecting the model pdf over (IW,TS) values that are outdated, far from the steady state. The situation improves for the case of $t_0 = 24$ months; even though the correction of the BaM³ method for the TS might be wrong in some cases, the pdf for the IW has

stabilized and points towards the correct value. In most of the cases, the correction for the IW outperforms the one for the TS, which leads to a general improvement of predictions by the BaM³ method.

Outlier patients challenge the method's performance. To explain the different behavior for IW and TS, we investigate the distribution of the failure cases over the full model parameter space. In general, the BaM³ method performs poorly for those patients that are at the extremes of the parameter space, who represent outlier patients. When plotting the patients in a scatter plot over cell motility and proliferation rate (Fig. 6), the points with high motility–high proliferation rates and high motility–low proliferation rates witness the highest number of failure cases for both clinical presentation times t_0 of 12 and 24 months. We checked for the distribution of failure cases also for the other model parameters, but no particular pattern was evident (see Fig. S7). Notably, patients falling into these high motility–high/low proliferation regions show the highest values for IW and TS (see ref. ¹⁴ and Fig. S8). Highly invasive and massive neoplasms are inadequately described by the pdf from the KDE, as

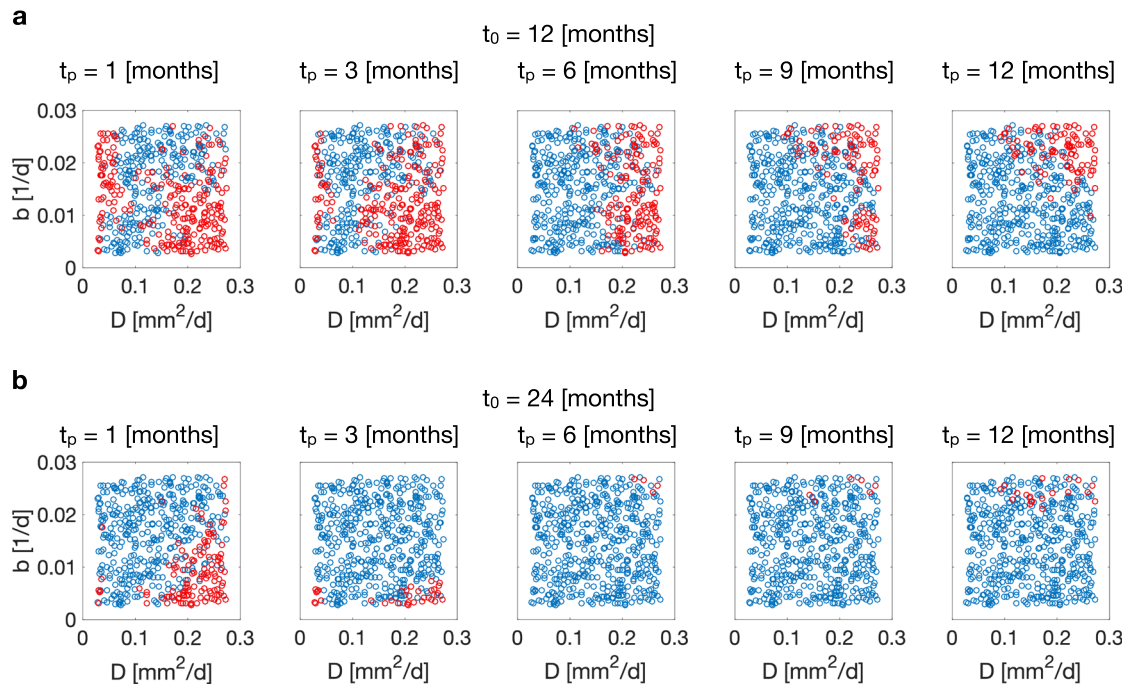


Fig. 6 Outlier patients challenge BaM³ performance. Scatter plots showing the distribution of failure cases (red dots) over the parameter space at different presentation (**a**, **b**) and prediction times. D : tumor cell motility; b : tumor cell proliferation rate; t_0 : clinical presentation time; t_p : prediction time.

they represent the extreme cases of the probability distribution. As a result, the FK model performs better in predicting the clinical outcomes with respect to the BaM³ method, since in the latter the correction from the dataset points towards smaller values of IW and, especially, TS.

Applying the method to real CLL patients: the effect of unmodelables. In addition to the proof-of-concept applied to in silico data, we test the BaM³ methodology on a cohort of real patients suffering from CLL. This cancer involves B cells and is characterized by the accumulation of lymphocytes in the blood, bone marrow, and secondary lymphoid tissues³⁶. In the past, CLL was considered to be a homogeneous disease of minimally self-renewing B cells, which accumulate due to a faulty apoptotic mechanism. This view was questioned by recent findings, suggesting a more heterogeneous neoplastic population continuously interacting with its microenvironment^{37–40}. Accumulation of leukemic cells occurs because of survival signals originating from the external environment and interacting with leukemic cells through a variety of receptors. The nature of this cross-talk with the environment is a current matter of research, featuring in vitro as well as in vivo experiments. One of the most significant experiments involving human patients was that of Messmer et al.³¹. Messmer and his co-workers inferred the kinetics of B-CLL cells from a group of patients through non-invasive labeling and mathematical modeling. Their investigation was quite thorough and involved the collection of several quantities related to patients' personal data (gender, age, etc.) and status of the disease (years since diagnosis, treatments, mutation status, etc.). They measured the fraction of neoplastic labeled cells in the blood of the patients, and fitted an ODE compartmental model to the dynamics that they observed. The model included three parameters, i.e., the daily water exchange rate (f_w), the B cell birth rate (b), and the relative size of the blood compartment (v_r).

We use the same model as Messmer and colleagues as the input for the BaM³ method, but discard the patient-specific fitting

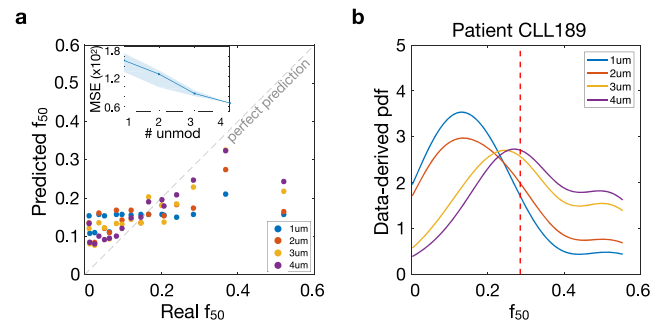


Fig. 7 Role of unmodelables in CLL patients. **a** Predicted vs measured values of the fraction of labeled cells at day 50. The colors denote the use of an increasing number of unmodelables in the BaM³ method. The inset shows the decrease in mean squared error (MSE) for the BaM³ predictions when we consider all the possible combinations of unmodelables. The solid line in the inset refers to the unmodelables used for the scatter plot. Here we included, in order, the CD38 expression, age, growth rate of white blood cells, and V_H mutation status as unmodelables. **b** Variation of the data-derived pdf obtained from density estimation when an increasing number of unmodelables is considered for a specific patient. The dashed red line highlights the measured value of f_{50} in this patient.

provided in their publication. Our aim is to show that, even when an individualized model parametrization is unknown, coupling the information given by the unmodelables can provide good patient-specific predictions. To accomplish this, we run simulations over uniform parameter ranges to obtain the pdf of the labeled cell fraction at day 50 (f_{50}), which is also the sole modelable variable in this dataset (see Fig. S9). Then, we incrementally select one to four unmodelable variables from the patients' dataset and build the data-derived pdf using the same KDE method as in the previous in silico example. The BaM³ method couples the two prediction distributions to obtain the pdf for the clinically relevant output (see Fig. S10). We show the

results of this procedure in Fig. 7a, where we compare the BaM³ predicted values against the patient f_{50} values reported in ref. 31. The fraction of labeled cells predicted by the BaM³ method agrees well with the reported data, especially when we increase the number of unmodelables used for density estimation. The inset shows how the MSE of BaM³ predictions decreases after considering all the possible combinations of unmodelables. Figure 7b shows how the probability distribution generated from the KDE changes for a representative patient. As the number of unmodelables increases, the mode of the distribution shifts towards the correct value of f_{50} , here denoted by a red dashed line. From Fig. 7a it is also possible to note that, even if the majority of points lies close to the perfect prediction line, the predictions of a few patients are significantly mismatched with respect to the corresponding real values. This occurs because these patients belong to the extremes of the parametric space (see Fig. S11). Patients characterized by outliers in their parametrization are under-represented in the modelable pdf due to the uniform sampling of the parameter space, and it is challenging for the data-derived correction to improve predictions for them.

Prediction of the time-to-relapse on a real ovarian cancer patient cohort: the importance of adequate model parametrization. To provide another application of the BaM³ method to a real scenario, we consider the case of patient response to therapy in high-grade serous ovarian cancer (HGSOC). This type of cancer is the most common epithelial ovarian cancer subtype, accounting for 70–80% of ovarian cancer-related deaths^{41,42}. In addition, due to treatment resistance, the 5-year survival rate in HGSOC is less than 50%^{43,44}. Indeed, the contribution of resistance mechanisms to tumor relapse after therapy is currently an active matter of research, recently backed up by evolutionary studies^{45–47}.

We start from the clinical dataset provided in a recent publication³², and elaborate a strategy to predict the TtR in ovarian cancer patients that makes use of the BaM³ methodology. The database of patients consists of 20 individuals, which are subject to the following treatment schedule (see Fig. S13). First, the patients receive neoadjuvant chemotherapy (NACT), consisting of different cycles of carboplatin and paclitaxel chemotherapy. Then, a surgery is performed, followed by other cycles of adjuvant chemotherapy. We propose a low-dimensional mathematical model to predict tumor TtR after treatment for each patient, which takes into account the presence of two cell subtypes. In particular, we include cells that are sensitive or resistant to chemotherapy. In addition, we consider the age of the patient at diagnosis as the unmodelable quantity used by the density estimator. Full details of the model and methodology are available in the corresponding sections of the ‘Methods’ and Supplementary Note 2.

As in the previous sections, the pdf from the mathematical model is obtained by simulating the latter over the parameter space. In this case, we focus on two parameters, namely the initial fraction of sensitive cells x_0 and the death rate induced by chemotherapy δ_0 . First, we consider a uniform distribution of both parameters. We assume x_0 between 0.4 and 0.9, in agreement with the degree of variability reported in the publication from which we take the dataset³². Since we lack any information about δ_0 , we select a wide range, from 0.1 and 10 days⁻¹. This results in an almost uniform pdf from the mathematical model, as shown in Fig. S13. In this condition, the pdf from the model enters the BaM³ as an uninformative prior in the Bayesian framework, leaving predictions to rely only on the pdf generated from the density estimation of the unmodelables. Note that, in these settings, BaM³ reduces to nonparametric

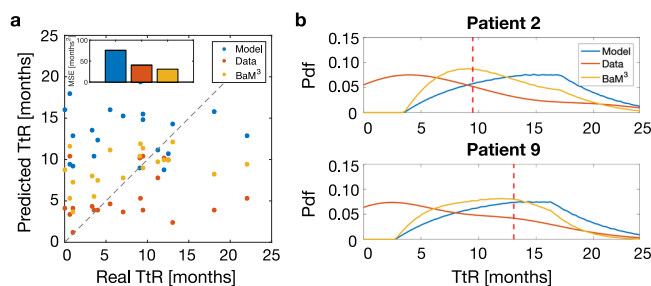


Fig. 8 BaM³ effectively couples predictions from machine learning and mathematical modeling in ovarian cancer patients. **a** Predicted vs real TtR calculated from the mode of the probability distribution obtained from the mathematical model (Model), density estimation (Data), and the BaM³ method. As shown in the inset, the MSE are of 75.674, 40.447, and 30.895 months², respectively. **b** Probability distributions obtained for two reference patients. The red dashed line denotes the real value of TtR for the specific patient. TtR: time-to-relapse of the tumor after therapy.

regression²⁹. We calculate the MSE using the mode of the distributions in the uninformative case, denoted as MSE_{un} , and find $MSE_{un} = 38.901$ months². As a next step, we use the additional information provided in the dataset to improve the parametrization of the mathematical model. Indeed, the dataset reports the tumor volume before and after the first cycle of therapy, as measured from clinical imaging³². We fit the value of δ_0 for each patient and take the mean of all these values as the center of another uniform distribution. We apply the BaM³ method using the newly generated pdf from the model and obtain a lower MSE, i.e., $MSE_{fit} = 30.895$ months² (see Fig. S14). Better parametrization results, therefore, in improved performance of the method. In addition, by applying BaM³ to a better parametrized model allows to obtain improved predictions, as shown in Fig. 8. The scatter plot in Fig. 8a shows reduced errors in BaM³ predictions with respect to the ones from the mathematical model or density estimation alone. Also, Fig. 8b displays the outcome of the method for two representative patients. In both cases, the pdf arising from BaM³ has its mode closer to the real TtR (dashed line), with respect to the modes obtained from the model or density estimation pdfs. This shows the potential of the BaM³ method, which is able to perform better than the single techniques upon which it is based.

Discussion

In the last few years, mathematical modeling and machine learning have emerged as promising methodologies in the biomedical field^{48–50}. However, several challenges persist and limit the prediction accuracy of both approaches. Among these issues, we identified the lack of knowledge of the mechanisms that govern the system under study (C1), and the paucity of time points at which patient information is available (C2), to significantly limit the performance of both mathematical models and machine learning techniques. In this work, we presented a method (BaM³) to couple mathematical modeling and density estimation in a Bayesian framework. The goal of BaM³ is to improve personalized tumor burden prediction in a clinical setting. This coupling allows to address the aforementioned (C1) and (C2) challenges, by exploiting the strengths of the respective methodologies and integrating them in a complementary path.

In particular, our proposed method aspires to solve a dire problem in personalized medicine that is related to the limited time-points of patient data collection. This implies that data

assimilation methods, such as Kalman filters or particle filters^{51,52} that require multiple data time-points integrated to a mechanistic model cannot be generally used. To this regard, the BaM³ method can be regarded as a one-step data assimilation method. Compared to other methodologies that combine outputs from mathematical models and measured data—such as Bayesian Melding⁵³, History Matching⁵⁴, Bayesian Model Calibration⁵⁵, or Approximate Bayesian Computation⁵⁶—the BaM³ method is not interested in parameter estimation to better calibrate the mathematical model. Instead, its goal is to improve predictions of mathematical models empowering them with knowledge from variables that are not usually considered (the ‘unmodelables’, in our framework). This is done without the exact knowledge of the parameters of the mathematical model; indeed, we calculated the pdfs of the modelable variables using a uniform sampling of the parameter space. Better estimation of the model parameters improves the outcomes of the method (as shown in the ovarian cancer test case), but it is not required for the methodology to be applied.

First, we tested the BaM³ method on a synthetic dataset of patients focusing on tumor growth dynamics. Our approach was able to improve the predictions of a FK model for the majority of the virtual patients, with significant improvements at later clinical presentation times. In addition, we tested the proposed methodology on two clinical datasets related to cancer, concerning tumor growth in leukemia and ovarian cancer patients. We compared the outcomes of the BaM³ method to the reported data and found excellent predictive capability. When analyzing the cases for which the performance of the BaM³ method was not optimal, we came across some limitations that should be addressed when applying the methodology to real cases.

The first limitation regards the selection of the proper unmodelable variables. These are quantities that cannot be easily mathematically modeled, but can be correlated to the patient clinical outputs. For our proof-of-concept we selected only a few unmodelables, but in principle multiple quantities could be considered at the same time. Moreover, the most important unmodelables could be selected in a process of feature selection similar to the ones usually adopted in machine learning, providing better accuracy for the predictions^{57,58}. We note as well that the method is open to progress in knowledge: should an unmodelable variable become modelable because of an increased understanding of the biological mechanisms, this variable can change side and become modelable.

One should also propose an adequate mathematical model that describes the dominant dynamics of the disease, as shown in the last case for ovarian cancer. A better parametrization of the model facilitates the work of density estimation, considerably reducing prediction errors. Not only better model parametrizations, but also mathematical models that encompass a suitable amount of mechanisms about the phenomenon that is modeled are advocated. In the case of ovarian cancer, we show in Supplementary Note 2 that a simplified model (with respect to the two cell populations presented in the ‘Methods’) is not able to provide good predictions when used in the context of the BaM³ method (see also Fig. S15).

Care must be taken with the selection of the metric that should be improved by the BaM³ method. For the *in silico* case, for example, considering the expected value of the final pdf resulted in better method performance when compared to selecting the pdf mode (see Figs. 3 and S5). This was probably due to the very similar natures of the FK and full models. Indeed, for lower clinical presentation times the FK model is already ‘primed’ towards the correct solution (in terms of outcomes of the full model); applying the BaM³ method might result in adding noise

to the FK prediction, degrading the final prediction. However, for some patients the FK model provides pdfs with multiple local maxima, sometimes far away from the full model values. In these instances (see Fig. S3), the BaM³ method is able to correct for the correct mode, shifting the pdf to the correct values. Therefore, a good practice would be to try multiple pdf metrics and test the BaM³ method on each of them. This would result in a more thorough understanding of the problem, eventually allowing for better predictions.

Another important issue is that the correlations between unmodelables and clinical outputs should be persistent over time, evolving on a timescale that is faster than the dynamics of the problem. In our synthetic dataset this was partially accomplished at later clinical prediction times, especially for the case of the tumor infiltration width. Indeed, the unmodelable variables need to provide as much time-invariant information as possible on the clinical output variables, implying an equilibrated pdf. Such data can be, for instance, from genetic origin (such as mutations) or from other variables with slow characteristic evolution time. We stress that it is the probability distribution of the unmodelables that has to be close to equilibrium; note that this does not require the value of the unmodelable variables to reach a constant value but the values should be drawn from a steady-state distribution.

We see room for improvement also concerning the selection of the density estimation method. We adopted a well-known form of nonparametric estimation through kernel density estimation, but other approaches could be tailored to a specific problem—especially when high-dimensional datasets come into play⁵⁹. Moreover, introducing density estimation methods to be able to integrate categorical variables would greatly benefit the technique, especially in biomedical problems (e.g., it would be extremely beneficial to include the grade of a tumor, or the particular sequence of therapies that a patient has undergone). The modularity of the BaM³ method makes it extremely versatile, allowing one to change the density estimation step, the modeling part, or both of them at the same time to improve the final prediction scores.

Care should also be taken to generate pdfs that are able to cope with outliers. In our proof-of-concept we generated the probability distributions considering the same weight for every patient, irrespective of his position in the parameter space. Techniques able to identify these extreme cases and to improve their contribution to the final pdfs should be implemented for a better method performance⁶⁰.

In summary, we can identify three main actions that could be undertaken when these limitations hamper the predictive capabilities of BaM³: (i) one should look for ways to improve the mathematical model, designing it to be as informative as possible; (ii) then, an effort should be put to constrain the model by a robust choice of parameters; finally, (iii) extreme care should be devoted to the selection of the most informative unmodelable variables.

We conclude by stating that the proposed method is not restricted to oncology. The core problem concerning clinical predictions is that data are heterogeneous and sparse in time along with lack of full mechanistic knowledge. Therefore, a vast variety of medical problems could be addressed by using the BaM³ approach. For instance, predicting the fate of renal grafts by using pre- and post-transplantation data is a prime application of our proposed methodology.

Data availability

All data generated or analyzed during this study are included in this published article, its References, and Supplementary information files. In particular, the dataset for chronic lymphocytic leukemia case is published in Table 1 of ref. ³¹. For the ovarian cancer case,

we use the data available in Supplementary Table 1 of ref.³². Note that, according to the Good Practice of Secondary Data Analysis (GPS) in Germany⁶¹, there is no need for additional ethics approval/consent when public domain data are analyzed.

Code availability

All source codes and analysis are publicly available through the Zenodo platform (<https://zenodo.org/record/4964592>)⁶².

Received: 23 April 2021; Accepted: 8 July 2021;

Published online: 29 July 2021

References

- Raghupathi, W. & Raghupathi, V. Big data analytics in healthcare: promise and potential. *Health Inf. Sci. Syst.* **2**, 3 (2014).
- Fahr, P., Buchanan, J. & Wordsworth, S. A review of the challenges of using biomedical big data for economic evaluations of precision medicine. *Appl. Health Econ. Health Policy* **17**, 443–452 (2019).
- Kalia, M. Personalized oncology: recent advances and future challenges. *Metabolism* **62**, S11–S14 (2013).
- Cruz, A. & Peng, W. K. Perspective: cellular and molecular profiling technologies in personalized oncology. *J. Pers. Med.* **9**, 44 (2019).
- Greenspan, H. Models for the growth of a solid tumor by diffusion. *Stud. Appl. Math.* **51**, 317–340 (1972).
- Gatenby, R. A. & Gawlinski, E. T. A reaction-diffusion model of cancer invasion. *Cancer Res.* **56**, 5745–5753 (1996).
- Byrne, H. M. Dissecting cancer through mathematics: from the cell to the animal model. *Nat. Rev. Cancer* **10**, 221–230 (2010).
- Altrock, P. M., Liu, L. L. & Michor, F. The mathematics of cancer: integrating quantitative models. *Nat. Rev. Cancer* **15**, 730–745 (2015).
- Araujo, R. P. & McElwain, D. S. A history of the study of solid tumour growth: the contribution of mathematical modelling. *Bull. Math. Biol.* **66**, 1039–1091 (2004).
- Preziosi, L. *Cancer Modelling and Simulation* (CRC Press, 2003).
- Zitnik, M. et al. Machine learning for integrating data in biology and medicine: principles, practice, and opportunities. *Inf. Fusion* **50**, 71–91 (2019).
- Alber, M. et al. Integrating machine learning and multiscale modeling—perspectives, challenges, and opportunities in the biological, biomedical, and behavioral sciences. *npj Digit. Med.* **2**, 1–11 (2019).
- Baker, R. E., Pena, J.-M., Jayamohan, J. & Jérusalem, A. Mechanistic models versus machine learning, a fight worth fighting for the biological community? *Biol. Lett.* **14**, 20170660 (2018).
- Alfonso, J. et al. Why one-size-fits-all vaso-modulatory interventions fail to control glioma invasion: in silico insights. *Sci. Rep.* **6**, 37283 (2016).
- Mascheroni, P. et al. On the impact of chemo-mechanically induced phenotypic transitions in gliomas. *Cancers* **11**, 716 (2019).
- Giese, A., Bjerkvig, R., Berens, M. & Westphal, M. Cost of migration: invasion of malignant gliomas and implications for treatment. *J. Clin. Oncol.* **21**, 1624–1636 (2003).
- Giese, A. et al. Dichotomy of astrocytoma migration and proliferation. *Int. J. Cancer* **67**, 275–282 (1996).
- Heddleston, J. M., Li, Z., McLendon, R. E., Hjelmeland, A. B. & Rich, J. N. The hypoxic microenvironment maintains glioblastoma stem cells and promotes reprogramming towards a cancer stem cell phenotype. *Cell Cycle* **8**, 3274–3284 (2009).
- Hatzikirou, H., Basanta, D., Simon, M., Schaller, K. & Deutsch, A. ‘Go or Grow’: the key to the emergence of invasion in tumour progression? *Math. Med. Biol.* **29**, 49–65 (2012).
- Joseph, J. V. et al. Hypoxia enhances migration and invasion in glioblastoma by promoting a mesenchymal shift mediated by the *hif1 α* -*zeb1* axis. *Cancer Lett.* **359**, 107–116 (2015).
- Stylianopoulos, T. et al. Causes, consequences, and remedies for growth-induced solid stress in murine and human tumors. *Proc. Natl Acad. Sci. USA* **109**, 15101–15108 (2012).
- Stylianopoulos, T. et al. Coevolution of solid stress and interstitial fluid pressure in tumors during progression: implications for vascular collapse. *Cancer Res.* **73**, 3833–3841 (2013).
- Murray, J. *Mathematical Biology II: Spatial Models and Biomedical Applications* (Springer New York, 2001).
- Swanson, K. R. et al. Quantifying the role of angiogenesis in malignant progression of gliomas: in silico modeling integrates imaging and histology. *Cancer Res.* **71**, 7366–7375 (2011).
- Swanson, K. R., Alvord, E. C. & Murray, J. Virtual brain tumours (gliomas) enhance the reality of medical imaging and highlight inadequacies of current therapy. *Br. J. Cancer* **86**, 14–18 (2002).
- Harpold, H. L., Alvord Jr., E. C. & Swanson, K. R. The evolution of mathematical modeling of glioma proliferation and invasion. *J. Neuropathol. Exp. Neurol.* **66**, 1–9 (2007).
- Hawkins-Daarud, A., Rockne, R. C., Anderson, A. R. & Swanson, K. R. Modeling tumor-associated edema in gliomas during anti-angiogenic therapy and its impact on imageable tumor. *Front. Oncol.* **3**, 66 (2013).
- Peter, D. H. Kernel estimation of a distribution function. *Commun. Stat. Theory Methods* **14**, 605–620 (1985).
- Gramacki, A. *Nonparametric Kernel Density Estimation and its Computational Aspects* (Springer, 2018).
- Läuter, H. Silverman, bw: Density estimation for statistics and data analysis. chapman & hall, london–new york 1986, 175 pp., £ 12.-. *Biom. J.* **30**, 876–877 (1988).
- Messmer, B. T. et al. In vivo measurements document the dynamic cellular kinetics of chronic lymphocytic leukemia b cells. *J. Clin. Invest.* **115**, 755–764 (2005).
- Kozłowska, E. et al. Mathematical modeling predicts response to chemotherapy and drug combinations in ovarian cancer. *Cancer Res.* **78**, 4036–4044 (2018).
- Ohgaki, H. & Kleihues, P. Epidemiology and etiology of gliomas. *Acta Neuropathol.* **109**, 93–108 (2005).
- Alfonso, J. et al. The biology and mathematical modelling of glioma invasion: a review. *J. R. Soc. Interface* **14**, 20170490 (2017).
- Chen, Y.-C. Modal regression using kernel density estimation: a review. *Wiley Interdiscip. Rev. Computat. Stat.* **10**, e1431 (2018).
- Kipps, T. J. et al. Chronic lymphocytic leukaemia. *Nat. Rev. Dis. Primers* **3**, 1–22 (2017).
- Chiorazzi, N., Rai, K. R. & Ferrarini, M. Chronic lymphocytic leukemia. *N. Engl. J. Med.* **352**, 804–815 (2005).
- Puente, X. S. & López-Otin, C. The evolutionary biography of chronic lymphocytic leukemia. *Nat. Genet.* **45**, 229–231 (2013).
- Kikushige, Y. et al. Self-renewing hematopoietic stem cell is the primary target in pathogenesis of human chronic lymphocytic leukemia. *Cancer Cell* **20**, 246–259 (2011).
- Savvopoulos, S., Misener, R., Panoskaltis, N., Pistikopoulos, E. N. & Mantalaris, A. A personalized framework for dynamic modeling of disease trajectories in chronic lymphocytic leukemia. *IEEE Trans. Biomed. Eng.* **63**, 2396–2404 (2016).
- Wright, A. A. et al. Neoadjuvant chemotherapy for newly diagnosed, advanced ovarian cancer: Society of Gynecologic Oncology and American Society of Clinical Oncology Clinical Practice Guideline. *Gynecol. Oncol.* **143**, 3–15 (2016).
- Bowtell, D. D. et al. Rethinking ovarian cancer II: reducing mortality from high-grade serous ovarian cancer. *Nat. Rev. Cancer* **15**, 668–679 (2015).
- Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2015. *CA Cancer J. Clin.* **65**, 5–29 (2015).
- Vergote, I. et al. Neoadjuvant chemotherapy or primary surgery in stage IIIc or IV ovarian cancer. *N. Engl. J. Med.* **363**, 943–953 (2010).
- Castellarin, M. et al. Clonal evolution of high-grade serous ovarian carcinoma from primary to recurrent disease. *J. Pathol.* **229**, 515–524 (2013).
- Cooke, S. L. et al. Genomic analysis of genetic heterogeneity and evolution in high-grade serous ovarian carcinoma. *Oncogene* **29**, 4905–4913 (2010).
- Salomon-Perzyński, A., Salomon-Perzyńska, M., Michalski, B. & Skrzypulec-Plinta, V. High-grade serous ovarian cancer: the clone wars. *Arch. Gynecol. Obstet.* **295**, 569–576 (2017).
- Brady, R. & Enderling, H. Mathematical models of cancer: when to predict novel therapies, and when not to. *Bull. Math. Biol.* **81**, 3722–3731 (2019).
- Kononenko, I. Machine learning for medical diagnosis: history, state of the art and perspective. *Artif. Intell. Med.* **23**, 89–109 (2001).
- Kumar, A., Bi, L., Kim, J. & Feng, D. D. Machine learning in medical imaging. In *Biomedical Information Technology* 2nd edn (ed. Feng, D. D.) 167–196 (Elsevier, 2020).
- Meinhold, R. J. & Singpurwalla, N. D. Understanding the Kalman filter. *Am. Stat.* **37**, 123–127 (1983).
- Ristic, B., Arulampalam, S. & Gordon, N. *Beyond the Kalman Filter: Particle Filters for Tracking Applications* (Artech House, 2003).
- Poole, D. & Raftery, A. E. Inference for deterministic simulation models: the bayesian melding approach. *J. Am. Stat. Assoc.* **95**, 1244–1255 (2000).
- Andrianakis, I. et al. Bayesian history matching of complex infectious disease models using emulation: a tutorial and a case study on hiv in uganda. *PLoS Comput. Biol.* **11**, e1003968 (2015).
- Kennedy, M. C. & O’Hagan, A. Bayesian calibration of computer models. *J. R. Stat. Soc. B* **63**, 425–464 (2001).

56. Csilléry, K., Blum, M. G., Gaggiotti, O. E. & François, O. Approximate bayesian computation (ABC) in practice. *Trends Ecol. Evol.* **25**, 410–418 (2010).
57. Guyon, I. & Elisseeff, A. An introduction to variable and feature selection. *J. Mach. Learn. Res.* **3**, 1157–1182 (2003).
58. Cai, J., Luo, J., Wang, S. & Yang, S. Feature selection in machine learning: a new perspective. *Neurocomputing* **300**, 70–79 (2018).
59. Scott, D. W. *Multivariate Density Estimation: Theory, Practice, and Visualization* (John Wiley & Sons, 2015).
60. Rousseeuw, P. J. & Leroy, A. M. *Robust Regression and Outlier Detection* Vol. 589 (John Wiley & Sons, 2005).
61. Swart, E. et al. Gute praxis sekundärdatenanalyse (GPS): leitlinien und empfehlungen. *Das Gesundheitswesen* **77**, 120–126 (2015).
62. Mascheroni, P. Bam3-method [Online]. <https://doi.org/10.5281/zenodo.4964592>. V1.0 (2021).

Acknowledgements

The authors gratefully acknowledge the funding support of the Helmholtz Association of German Research Centers-Initiative and Networking Fund for the project on Reduced Complexity Models (ZT-I-0010). H.H. and P.M. acknowledge the funding support of MicMode-I2T (01ZX1710B) by the Federal Ministry of Education and Research (BMBF). H.H. is supported by SYSIMIT (01ZX1308D) and MulticellML (01ZX1707C) by the Federal Ministry of Education and Research (BMBF). Finally, H.H. would like to thank the Volkswagenstiftung for its support within the “Life?” program (96732).

Author contributions

Conceptualization: H.H.; study design: P.M., S.S., J.C.L.A. and H.H.; software: P.M., S.S., and J.C.L.A.; formal analysis: P.M., S.S., J.C.L.A., M.M.-H. and H.H.; writing-original draft preparation: P.M.; writing-review and editing: P.M., S.S., J.C.L.A., M.M.-H. and H.H.; supervision: M.M.-H. and H.H.; funding acquisition: M.M.-H. and H.H.

Funding

Open Access funding enabled and organized by Projekt DEAL.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s43856-021-00020-4>.

Correspondence and requests for materials should be addressed to H.H.

Peer review information *Communications Medicine* thanks the anonymous reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021, corrected publication 2021