Bayesian combination of mechanistic modeling and machine learning (BaM^3): improving clinical tumor growth predictions

Pietro Mascheroni[1], Symeon Savvopoulos[2], Juan Carlos López Alfonso[1], Michael Meyer-Hermann[1,3,4], and Haralampos Hatzikirou*[1,5,6]

[1]Braunschweig Integrated Centre of Systems Biology and Helmholtz Center for Infectious Research, 38106 Braunschweig, Germany
[2]KU Leuven, Department of Chemical Engineering, Celestijnenlaan 200F, 3001 Leuven, Belgium
[3]Centre for Individualized Infection Medicine, 30625 Hannover, Germany
[4]Institute for Biochemistry, Biotechnology and Bioinformatics, Technische Universität Braunschweig, 38106 Braunschweig, Germany
[5]Mathematics Department, Khalifa University, P.O. Box 127788, Abu Dhabi, UAE
[6]Center for Information Services and High Performance Computing, TU Dresden, 01062 Dresden, Germany

October 2, 2020

Abstract

Biomedical problems are highly complex and multidimensional. Commonly, only a small subset of the relevant variables can be modeled by

*To whom correspondence should be addressed.
E-mail: haralampos.hatzikirou@ku.ac.ae

virtue of mathematical modeling due to lack of knowledge of the involved phenomena. Although these models are effective in analyzing the approximate dynamics of the system, their predictive accuracy is generally limited. On the other hand, statistical learning methods are well-suited for quantitative reproduction of data, but they do not provide mechanistic understanding of the investigated problem. Herein, we propose a novel method, based on the Bayesian coupling of mathematical modeling and machine learning (BaM^3). We evaluate the proposed BaM^3 method on a synthetic dataset for brain tumor growth as a proof of concept and analyze its performance in predicting two major clinical outputs, namely tumor burden and infiltration. Combining these two approaches results in improved predictions in almost all simulated patients, especially for those with a late clinical presentation. In addition, we test the proposed methodology on a set of patients suffering from Chronic Lymphocytic Leukemia (CLL) and show excellent agreement with reported data.

Introduction

Biomedical problems are complex in nature, commonly involving several players ranging from molecular to tissue scale [1, 2]. Such complexity is inherent in living systems and stems from the continuous interactions between their same individual parts and their environment [3]. When facing biomedical problems, and in particular cancer, scientists have been long supported by the use of mathematical modeling as a tool to identify causal relationships in the experimental and clinical data [4, 5, 6]. 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 [7]. Mathematical models can focus on the dynamics of biological systems irrespective of the spatial distribution of the underlying components (by employing, for instance, Ordinary Differential Equations). On the other hand, models can also take spatial

effects into account in a discrete (e.g. through Agent Based, Cellular Automata models) or continuous (e.g. through Partial Differential Equations) framework [8, 9]. Although these models can be extremely powerful both in predicting system responses and suggesting new experimental directions, they require some knowledge of the underlying mechanisms of the analysed system. Commonly, this knowledge is not complete, and only for a limited portion of the involved variables the corresponding mechanistic interactions are known. Therefore, although mathematical models often provide a good description of a simplified version of the associated system dynamics, they not always allow for accurate and universal 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 interactions [10]. While mathematical models rely on causality, statistical learning methods identify correlations among data [11]. 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 problem [12]. As a result, the overall understanding of the fundamental system dynamics becomes almost impossible, as the chance to generalize the 'learnt' system behavior.

In this work, we propose a novel Bayesian method to combine mathematical modeling and statistical learning (BaM^3), taking advantage of their respective strengths. As a proof of concept, the proposed BaM^3 method is tested on a synthetic dataset of brain tumor growth. We analyze the method performance in predicting two relevant clinical outcomes, namely tumor burden and infiltration. When comparing predictions from the mechanistic model with those from the BaM^3 method we obtain improved predictions for the vast majority of virtual patients. In addition, we apply the approach to a dataset of patients suffering from Chronic Lymphocytic Leukemia. The BaM^3 method shows excellent agreement between the predicted clinical out-

put and the reported data.

Results

We introduce the notation and key ideas of the proposed methodology, leaving the full derivation and additional details to the Supplementary Information (SI). 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 *new* patient at which the patient specific r.v. realizations $(\mathbf{X}_m = \mathbf{x}_m^*, \mathbf{X}_u = \mathbf{x}_u^*)$ are obtained (see the scheme in Fig. 1). The overall goal of the method is to predict the patient's clinical output 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 a $N-$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). 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 $p(\mathbf{Y} = \hat{\mathbf{y}}|\mathbf{X}_m = \mathbf{x}_m^*, \mathbf{X}_u = \mathbf{x}_u^*)$. By using the arguments described in the SI, we can prove that:

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

The implementation of the BaM^3 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

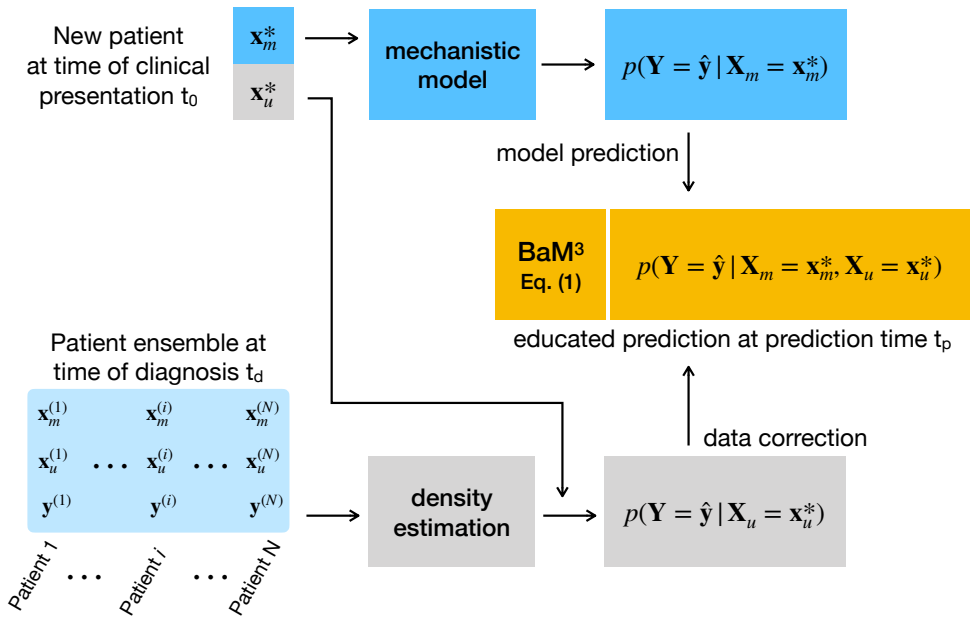


Figure 1: Schematic representation of the proposed methodology. The patient ensemble is constituted by modelable variables ($\mathbf{x}_m$), unmodelable variables ($\mathbf{x}_u$) and clinical observations $\mathbf{y}$ at the time of diagnosis t_d. Given the new patient presentation ($\mathbf{x}_m^*$,$\mathbf{x}_u^*$) at the clinical presentation time t_0, the goal of the suggested approach is to find an estimate for the clinical observables $\hat{\mathbf{y}}$ at the prediction time t_p. The modelable variables $\mathbf{x}_m^*$ are used to setup the mathematical model, and provide predictions for the clinical observables $p(\mathbf{Y} = \hat{\mathbf{y}}|\mathbf{X}_m = \mathbf{x}_m^*)$. The unmodelable variables $\mathbf{x}_u^*$ are used in a density estimation method to provide the probability distribution function (pdf) $p(\mathbf{Y} = \hat{\mathbf{y}}|\mathbf{X}_u = \mathbf{x}_u^*)$, which is used to correct the predictions from the mathematical model. The proposed approach (Bayesian combination of mathematical modeling and machine learning, BaM^3) results in a new pdf of the clinical observables at time t_p, combining the outputs of the mathematical model and density estimation method.

straightforward for the mathematical model (see Methods), 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 $\mathbf{X}_u = \mathbf{x}_u^*$.

Here we would like to point out that the only assumption made to derive expression (1) was the statistical independence of the modelable and unmodelable variables conditioned by the clinical outputs, i.e. :

$$p(\mathbf{X}_m, \mathbf{X}_u|\mathbf{Y}) = p(\mathbf{X}_m|\mathbf{Y})p(\mathbf{X}_u|\mathbf{Y}). \tag{2}$$

The impact of this assumption can be estimated by reformulating eq. (1) into the corresponding information theoretical quantities (see SI). Then, it can be proven that

$$S_{method}(\mathbf{Y}|\mathbf{X}_m, \mathbf{X}_u) \geq S_{true}(\mathbf{Y}|\mathbf{X}_m, \mathbf{X}_u). \tag{3}$$

This highlights that the entropy S_{method} produced by the proposed BaM^3 method is greater than the entropy S_{true} of the corresponding 'true' predictions without using the assumption (2). In particular, we can use the inequality that connects the mean square estimator (MSE) between reality and BaM^3 prediction and the conditional entropy that reads [13]:

$$\langle(\hat{\mathbf{y}}(\mathbf{x_m}, \mathbf{x_u}) - \mathbf{y})^2\rangle \geq \frac{e^{S_{method}(\mathbf{Y}|\mathbf{X}_m,\mathbf{X}_u)}}{2\pi e} \geq \frac{e^{S_{true}(\mathbf{Y}|\mathbf{X}_m,\mathbf{X}_u)}}{2\pi e}. \tag{4}$$

The above inequality implies that the assumption of statistical independence of the two sets of variables forces the lower bound of the MSE higher in comparison to a method that would not assume this. Please note that the above inequality is a generalization of the Cramer-Rao bound, which can be recovered by assuming a Gaussian distribution for the r.v. $(\mathbf{Y}|\mathbf{X}_m, \mathbf{X}_u)$.

Testing the BaM^3 method on synthetic glioma growth

As a proof of concept, we synthetically generate a dataset of glioma patients using a system of PDEs recently published [14]. Gliomas are aggressive brain

tumours generally associated with low survival rates [15]. One of the most important hallmarks of this type of tumors is its invasive behaviour, combined with a marked phenotypic plasticity and infiltrative morphology [16]. Such qualities fostered the development of several mathematical models to support clinicians in the treatment of the disease [17]. The equations of the selected mathematical model ('full model', in the following) 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, 18, 19, 20, 21]. 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 [22, 23, 24]. 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 to be the modelable variable. Moreover, we treat the integral over the tissue of oxygen concentration and vascular density, denoted as $\bar{n}, \bar{v}$ respectively, as the unmodelable quantities . Starting from the same initial conditions, we simulate the behavior of each virtual patient for three years, storing the values of all variables at each month.

As sketched in Fig. 2, we use the modelable variable (i.e. tumor cell density c) to setup a mathematical model. In particular, we take $c(\mathbf{x}, t_0)$ at a specific time point, the clinical presentation time t_0, and use it as the initial condition for a Fisher–Kolmogorov equation [25, 26, 27, 28, 29] ('FK model', in the following). 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 two clinical observables, i.e. the Tumor Size (TS) and Infiltration Width (IW), as detailed in the Methods. The first quantity is related to tumor burden, whereas the second accounts for tumor infiltration in the host tissue. 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 of months), and

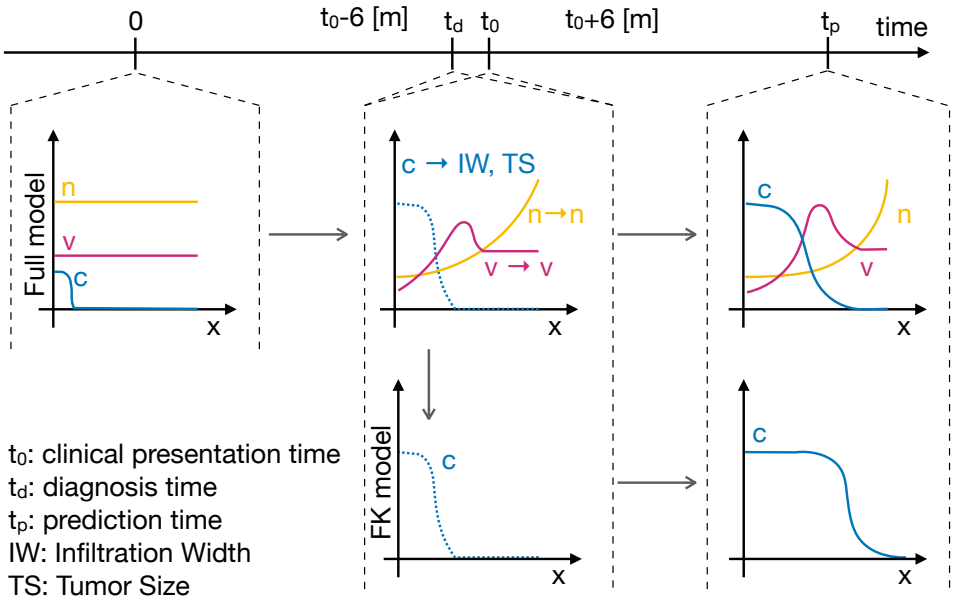


Figure 2: Workflow for generating the synthetic data. The full model is initialized at $t = 0$ and used to simulate the spatio-temporal variation of tumor density c, oxygen concentration n and functional tumor-associated vasculature v. For each virtual patient, two clinical outputs are tracked, i.e. 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.

collect the values of modelables, unmodelables and clinical outputs at this time to build the patient ensemble.

According to our previous notation, we have therefore $\mathbf{x}_m = c(x,t)$, $\mathbf{x}_u = (\bar{n}, \bar{v})$ and $\mathbf{y}$ =(TS, IW). Given the patient-specific modelable and unmodelable variables ($c(\mathbf{x}, t_0)$ and $\bar{n}, \bar{v}$, respectively) at the clinical presentation time t_0, the BaM^3 method produces the probability of observing certain clinical outputs (i.e. TS and IW) at a specific prediction time t_p. The data-driven pdf in the equation (1) is obtained through a normal kernel density estimator (KDE) [30, 31, 31, 32] trained with the patient ensemble. Then, the first factor in equation (1) is calculated using the FK model for each patient, running the simulations over the whole parameter range of cell motility and proliferation rates. This produces a map of possible IW and TS predicted from the mathematical model for each virtual patient (see the Methods for further information about the KDE and modeling steps).

Improving the predictions of the mathematical model

Figs. 3**A**-**C** show the results of applying the BaM^3 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 conditional probability $p(\mathbf{Y} = \hat{\mathbf{y}}|\mathbf{X}_m = \mathbf{x}_m^*)$ obtained from the FK model is shown in Fig. 3**A**. 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 BaM^3 method making use of the probability calculated from the KDE, shown in Fig. 3**B**. 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. 3**C** (another example for a different

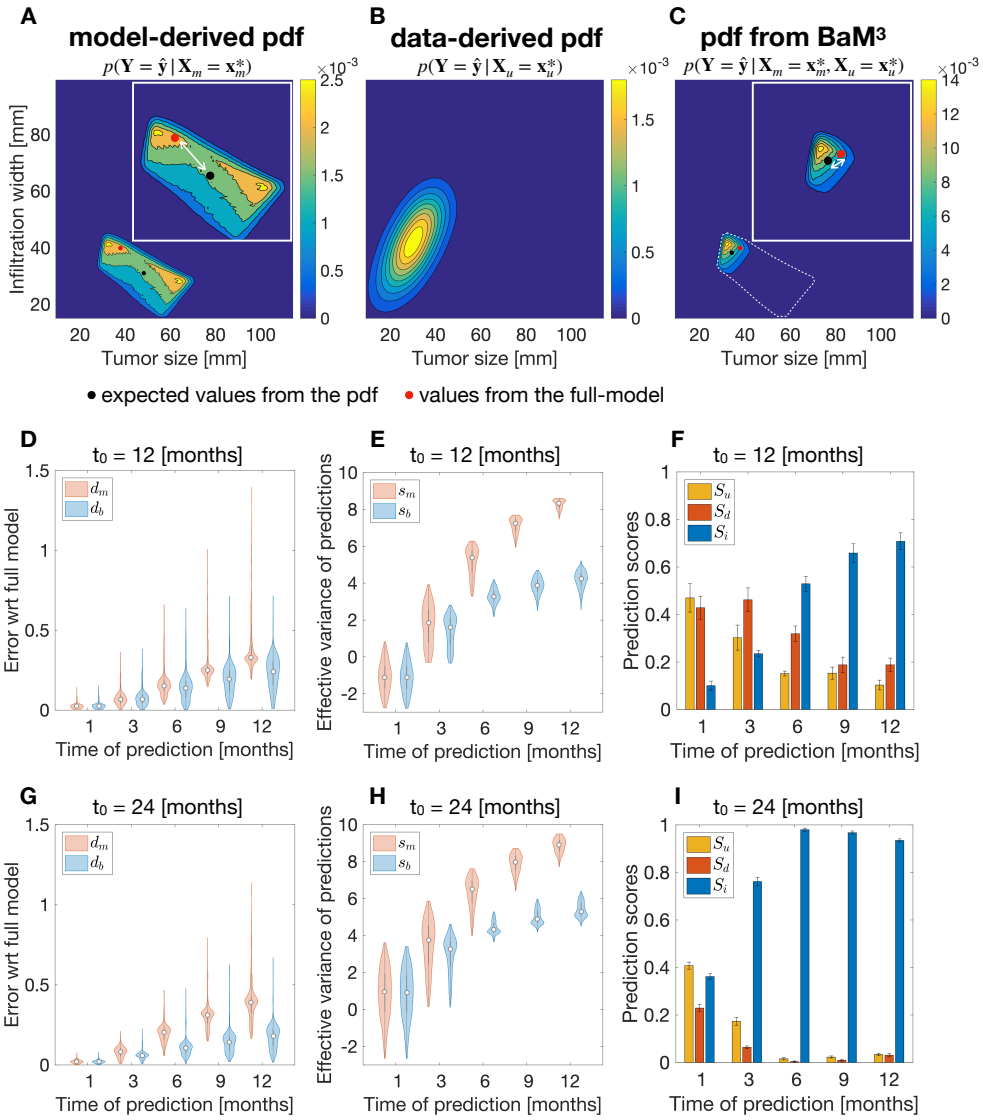

A
model-derived pdf
$p(\mathbf{Y} = \hat{\mathbf{y}} \,|\, \mathbf{X}_m = \mathbf{x}_m^*)$
B
data-derived pdf
$p(\mathbf{Y} = \hat{\mathbf{y}} \,|\, \mathbf{X}_u = \mathbf{x}_u^*)$
C
pdf from BaM3
$p(\mathbf{Y} = \hat{\mathbf{y}} \,|\, \mathbf{X}_m = \mathbf{x}_m^*, \mathbf{X}_u = \mathbf{x}_u^*)$
Infiltration width [mm]
Tumor size [mm]
expected values from the pdf
values from the full-model
D
E
F
G
H
I
t_0 = 12 [months]
t_0 = 24 [months]
Error wrt full model
Effective variance of predictions
Prediction scores
Time of prediction [months]
d_m
d_b
s_m
s_b
S_u
S_d
S_i

Figure 3 *(previous page)*: **A-C** Application of the BaM^3 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-driven pdf for the specific patient calculated through the kernel density estimator trained over the patient ensemble. **C** Probability distribution obtained from the BaM^3 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 BaM^3 method (d_b), for a clinical presentation time of 12 and 24 months (**D**,**G**, respectively). Effective variance of the predictions for the corresponding clinical presentation times (**E**,**H**, respectively). Prediction score, as the ratio of cases for which the BaM^3 method improved (S_i), did not change (S_u) or deteriorated (S_d) the model predictions. The error bars show the variation of the results by replicating the method over 10 different sets of N=500 patients.

patient is given in Fig. S2 in the SI). 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 BaM^3 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. 3**D**,**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 BaM^3 method (d_b). These nondimensional errors are calculated as

$$d_k = \sqrt{\frac{1}{L}\sum_{i=1}^{N_y}\left(1-\frac{\langle y_i^k\rangle}{y_i^r}\right)^2} \tag{5}$$

where $k = m, b$, $\langle y_i^k\rangle$ are the expected observable values (i =IW, TS) calculated from the FK model and the BaM^3 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 BaM^3 method improved, deteriorated or left 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 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. 3**D**,**G** for the two clinical presentation times considered. However, after applying the BaM^3 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 BaM^3 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 BaM^3

pdfs, (identified by s_m, s_b respectively - see 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. 3**E**,**H**) the BaM^3 method provides thinner pdfs, more centered around their expected value with respect to the FK model-derived case.

Finally, the stacked bars in Fig. 3**F**,**I** show that BaM^3 performs well at later prediction times and remarkably well (improvement ratio S_i close to 1), especially at the latest clinical presentation time $t_0 = 24$ months. For $t_0 =$ 12 months (Fig. 3**F**) the proposed method is not able to improve predictions until a prediction time of 6 months. Then, for $t_p = 6$, 9 and 12 month the advantages of using BaM^3 over the FK model are unambiguous. On the other hand, for the clinical presentation time of 24 months (Fig. 3**I**) both S_u and S_d decrease significantly for prediction times equal 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. 3**F**,**I** denote the variability in the results that is obtained by replicating the study 10 times, each with $N = 500$ randomly generated patients. Fig. S3 in the SI 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. S4). When the pdfs display multiple maxima we consider the average of the relative errors between the values of the full model and the different peaks. The performance of the BaM^3 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 BaM^3 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) decreasing the rate at which the error increases with prediction time, and (iii) decreasing the variance associated with the probability distributions.

Failure in improving TS model predictions reduces BaM^3 performance

Even though the BaM^3 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 analyzed the failure cases by splitting the errors d_m and d_b into the two partial errors

$$\Delta \mathrm{IW}_k = \sqrt{\left(1 - \frac{\langle \mathrm{IW}^k \rangle}{\mathrm{IW}^r}\right)^2}, \quad \Delta \mathrm{TS}_k = \sqrt{\left(1 - \frac{\langle \mathrm{TS}^k \rangle}{\mathrm{TS}^r}\right)^2} \tag{6}$$

where $k = m, b$, $\langle \mathrm{IW}^k \rangle$ and $\langle \mathrm{TS}^k \rangle$ are the expected values of the clinical outputs obtained from the mathematical model and BaM^3 pdfs ($k = m, b$, respectively), and $\mathrm{IW}^r, \mathrm{TS}^r$ are the values of these quantities from the full model. Fig. 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 BaM^3 method are equal. Above this line, the proposed BaM^3 method deteriorates the model predictions, whereas under that line the BaM^3 method improves predictions. The red dots in the scatter plots represent the patients for which the BaM^3 method fails ('failure cases', in the following). After a prediction time of one month, in which a characteristic pattern is not evident, the plots highlight that failure cases are generally associated with regions where the BaM^3 method under-performs to the FK model with respect to TS ($\Delta \mathrm{TS}_b > \Delta \mathrm{TS}_m$). Interestingly, the same failure cases belong to regions in which $\Delta \mathrm{IW}_b < \Delta \mathrm{IW}_m$: the BaM^3 method is 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 BaM^3 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 BaM^3 method in improving the predictions for one of the clinical observables, we analyzed the behavior of the pdf arising from density estimation, i.e. $p(\mathbf{Y} = \hat{\mathbf{y}}|\mathbf{X}_u = \mathbf{x}_u^*)$ in Equation (1). Fig. 5 shows the temporal evolution of $p(\mathbf{Y} = \hat{\mathbf{y}}|\mathbf{X}_u = \mathbf{x}_u^*)$ for different clinical presentation times t_0. As done previously for $t_0 = 12$ and 24 months, each entry of the ensemble of patients from which the pdf is drawn is evaluated at a diagnosis time t_d taken between ± 6 months around the presentation time t_0. Fig. 5**A** 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. 5**B**,**C**, showing the marginal probabilities for IW and TS, calculated from the distribution in Fig. 5**A**. 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. 5**D**,**E** for the overlap between the distributions at presentation times t_0 of 12 and 18 months and between 24 and 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

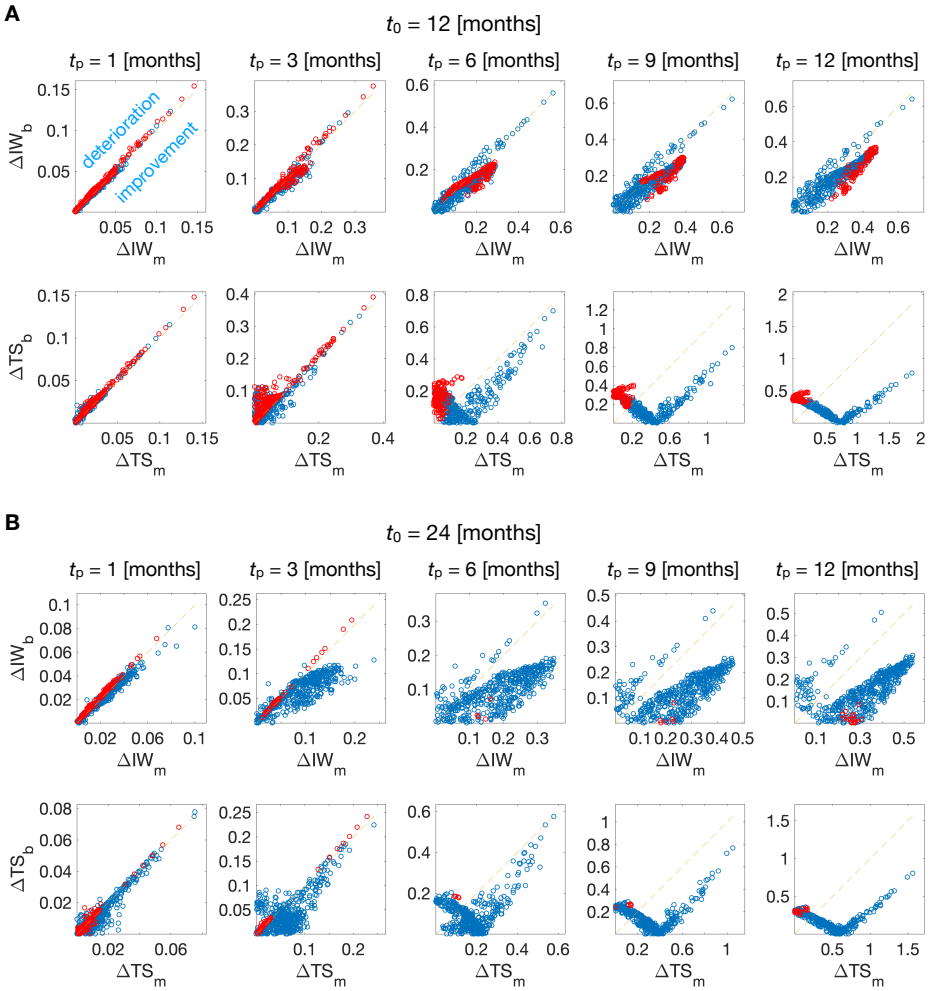


Figure 4: Scatter plots of the partial errors $\Delta\mathrm{IW}_i, \Delta\mathrm{TS}_i$ over different prediction times t_p ($i = m, b$, for the quantity related to the mechanistic model or the BaM^3 method, respectively). Results for presentation times of $t_0 = 12$ months (**A**) and $t_0 = 24$ months (**B**) are shown. Each red dot corresponds to a patient for which the BaM^3 method fails in improving the FK model predictions. Regions over the dashed line correspond to areas in which the BaM^3 method deteriorates the FK model predictions, while regions under that line show areas in which it improves the FK predictions.

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 viceversa. For the earlier times in Fig. 5**D** the patients are mostly scattered along a line of increasing IW_r and TS_r with points where the overlap is poor (close to 0.4 in certain regions). On the other hand, for the later case in Fig. 5**E** 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 BaM^3 method performances 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 BaM^3 method for the TS might be wrong is 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 BaM^3 method.

Patients at the extreme of the parameter space challenge the method

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 BaM^3 method performs poorly for those patients that are at the extremes of the parameter space. 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. S5).

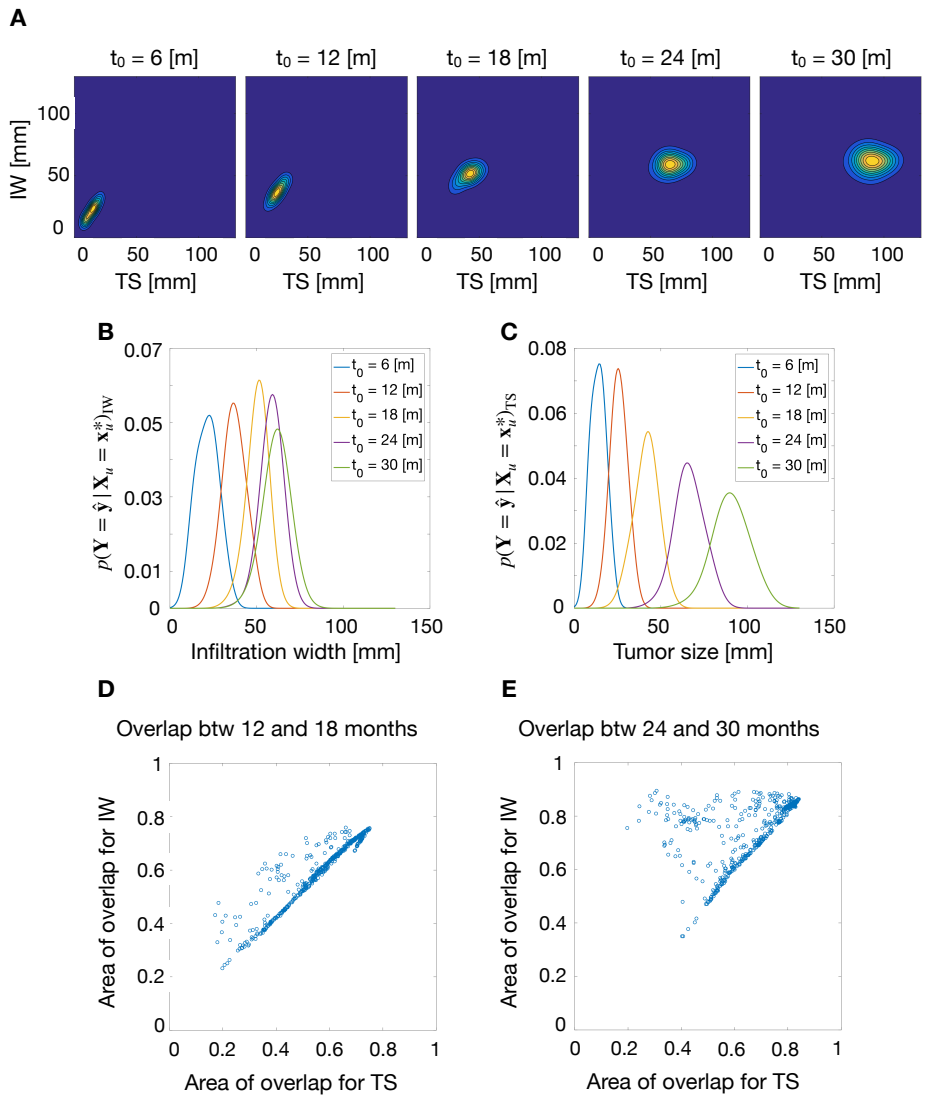


Figure 5: Temporal evolution of the unmodelable pdf from the KDE step. **A** Data-driven pdf for a representative patient at different clinical presentation times, plotted over the (IW,TS) space. Marginal probabilities from the plots in **A**, showing the evolution of the probability distribution for 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**).

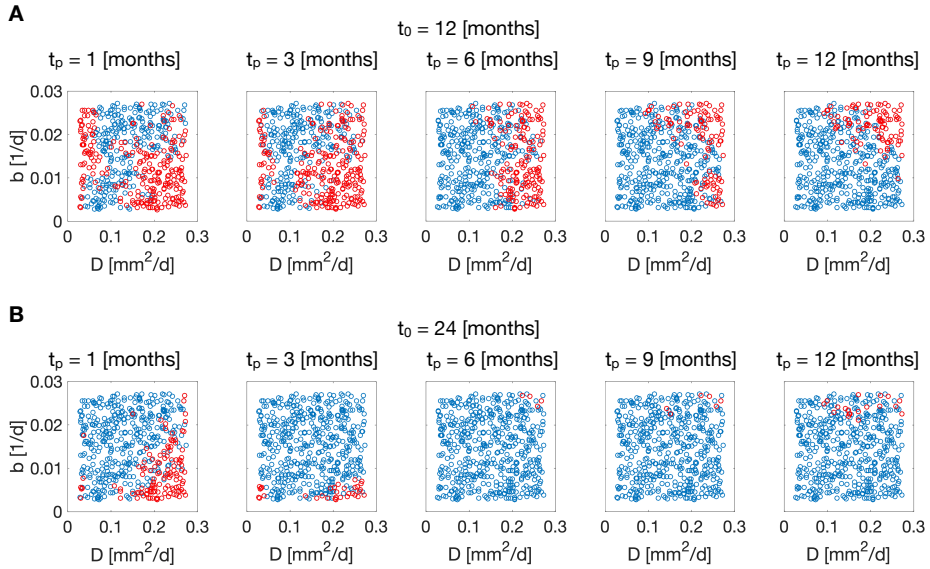


Figure 6: Scatter plots showing the distribution of failure cases (red dots) over the parameter space at different presentation and prediction times. D: tumor cell motility; b: tumor cell proliferation rate; t_0: clinical presentation time; t_p: prediction time.

Notably, patients falling into these high motility - high/low proliferation regions show the highest values for IW and TS (see [14] and Fig. S6). Highly invasive and massive neoplasms are inadequately described by the pdf from the KDE, as 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 BaM^3 method, since in the latter the correction from the dataset points towards smaller values of IW and, especially, TS.

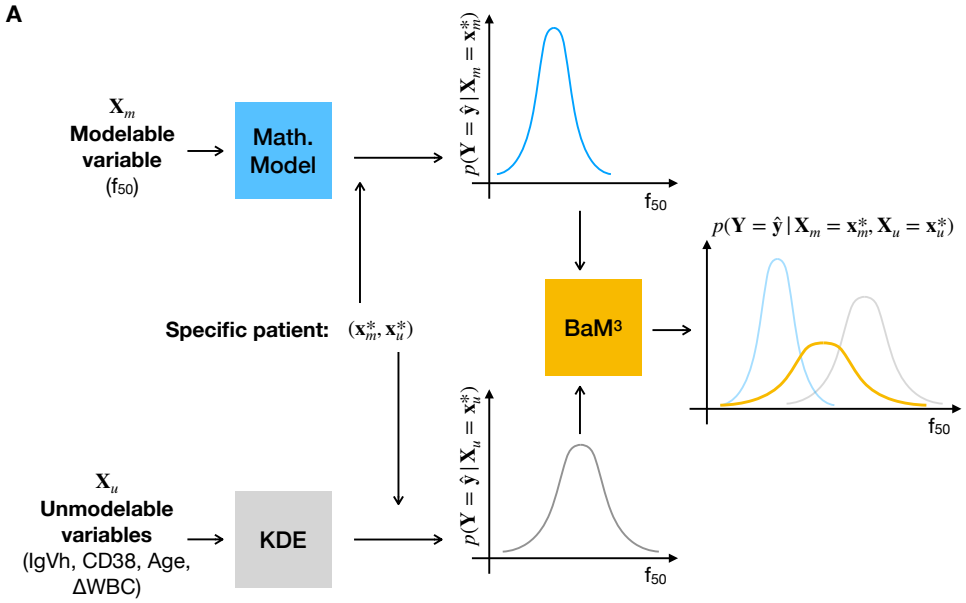
Applying the method to real CLL patients

In addition to the proof of concept applied to *in silico* data, we test the BaM^3 methodology in the context of Chronic Lymphocytic Leukemia (CLL). This cancer involves B cells and is characterized by the accumulation of lymphocytes in the blood, bone marrow and secondary lymphoid tissues [33]. In

the past, CLL was considered as 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 [34, 35, 36, 37]. 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 [38]. Messmer and his co-workers inferred the kinetics of CLL-B cells from a group of patients through non-invasive labeling, using a dynamic mathematical model. 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 used an ODE compartmental model to interpolate the fraction of labeled cells in the blood of the patients. The latter 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 run simulations over the fitted 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 the Methods and Fig. 7**A**). Then, we select four unmodelable variables from the patients' dataset and build the data-driven pdf using the same KDE method as in the previous *in silico* example. After that, we enter the patient specific data. In particular, we assumed the same initial labeled fraction f_{50}^0 for all the patients, i.e. $\mathbf{x}_m^* = f_{50}^0$. For the unmodelable variables, we equipped the KDE method using the four values of the unmodelables specific to a given patient. The BaM^3 method composed the two prediction distributions to obtain the pdf for the clinically relevant output (see Fig. S7).

We show the results of this procedure in Fig. 7**B**, where we compare


A
$\mathbf{X}_m$
Modelable variable
(f_{50})
Math. Model
$p(\mathbf{Y} = \hat{\mathbf{y}} \mid \mathbf{X}_m = \mathbf{x}_m^*)$
f_{50}
$p(\mathbf{Y} = \hat{\mathbf{y}} \mid \mathbf{X}_m = \mathbf{x}_m^*, \mathbf{X}_u = \mathbf{x}_u^*)$
Specific patient: $(\mathbf{x}_m^*, \mathbf{x}_u^*)$
BaM³
f_{50}
$\mathbf{X}_u$
Unmodelable variables
(IgVh, CD38, Age, ΔWBC)
KDE
$p(\mathbf{Y} = \hat{\mathbf{y}} \mid \mathbf{X}_u = \mathbf{x}_u^*)$
f_{50}

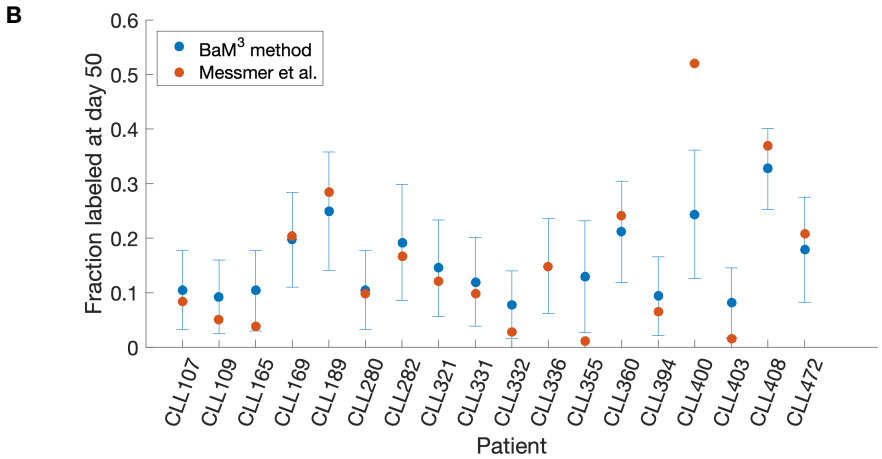

B
BaM³ method
Messmer et al.
Fraction labeled at day 50
0
0.1
0.2
0.3
0.4
0.5
0.6
CLL107
CLL109
CLL165
CLL169
CLL189
CLL280
CLL282
CLL321
CLL331
CLL332
CLL336
CLL355
CLL360
CLL394
CLL400
CLL403
CLL408
CLL472
Patient

Figure 7 *(previous page)*: **A** Schematic of approach used for the CLL patients. The mathematical model provides the pdf for the clinical outcome ($\mathbf{Y}$) having as input the modelable variable ($\mathbf{X}_m$). Note that in this case both the clinical output and the modelable variable correspond to the fraction of labeled cells at day 50 (f_{50}). At the same time, KDE provides the pdf for the clinical output using a set of unmodelable variables ($\mathbf{X}_u$). After entering the data for a specific patient ($\mathbf{x}_m^*, \mathbf{x}_u^*$), the BaM^3 method is applied. **B** Comparison between the expected values of the pdf obtained from the BaM^3 method (blue dots) and the reported data (orange dots) [38]. Error bars denote the standard deviation of the pdf.

the BaM^3 predicted values against the patient f_{50} values in [38]. The fraction of labeled cells predicted by the BaM^3 method agrees well with the reported data. The majority of data points lies within the standard deviation estimated in the generated distributions from BaM^3. However, there is a remarkable error with BaM^3 predicting the behavior of patients CLL355 and CLL400. This might occur because these patients belong to the extremes of the parametric space (see Fig. S8). Patients characterized by outliers in their parametriziation are under-represented in the modellable pdf due to the uniform sampling of the parameter space.

Discussion

In the last few years, machine learning has emerged as a promising methodology in the biomedical field [39, 40]. Leveraging the improved quality and quantity of the available biomedical data, novel algorithms are being developed to increase the prediction accuracy of machine learning methods. Even if the latter become more and more accurate, the black-box nature of the prediction process limits its transparency and accountability [41]. The urge for interpretability advocates for novel approaches, able to shed light on the

machine decision processes [42, 43, 44]. In this work, we presented a novel method (BaM^3) to couple mathematical modeling and density estimation in a Bayesian framework. By unifying these complementary paths, the mechanistic understanding of mathematical models is empowered with the versatility of statistical methods.

Compared to other methodologies that combine outputs from mathematical models and measured data - such as Bayesian Melding [45], History Matching [46], Bayesian Model Calibration [47] or Approximate Bayesian Computation [48] - the BaM^3 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 might improve the outcomes of the method, but it is not required for the methodology to be applied.

The BaM^3 method was tested 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 a dataset of CLL patients and compared the outcomes of the BaM^3 method to the reported data. When analyzing the cases for which the performance of the BaM^3 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 took only two unmodelables for the first example and four for the second. However, 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 [49, 50]. In addition, 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.

Care must be taken with the selection of the metric that should be improved by the BaM^3 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 Fig. 3 and Fig. S4). 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 BaM^3 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. S2) the BaM^3 method is able to correct for the right peak, shifting the pdf to the correct values. Therefore, a good practice would be to try multiple pdf metrics and test the BaM^3 method on each of them. This would result in a more thorough picture 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 unmodelables variables need to provide as much as possible time-invariant information 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 play [51]. The modularity of the BaM^3 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.

In addition, care should 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 performance [52].

Finally, it is important to state 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 BaM^3 approach. For instance, predicting the fate of renal grafts by using pre- and post transplantation data is a prime application of our proposed methodology.

Methods

The full mathematical tumor model

Following [14], we propose a mathematical model that describes the growth of vascularised gliomas focusing on the interplay between the migration/proliferation dichotomy and vaso-occlusion at the margin of viable tumour tissue ('full model', in the main text). 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 $v(x,t)$. For simplicity we consider a one-dimensional computational domain. The variables are non-dimensional and the equations read:

$$\frac{\partial c}{\partial t} = D\frac{\partial^2}{\partial x^2}\left(\frac{\alpha}{\alpha_0}c\right) + b\frac{\beta}{\beta_0}c(1-c), \tag{7}$$

$$\frac{\partial n}{\partial t} = D_n\frac{\partial^2 n}{\partial x^2} + h_1 v(n_0 - n) - h_2 cn, \tag{8}$$

$$\frac{\partial v}{\partial t} = D_v\frac{\partial^2 v}{\partial x^2} + g_1\mathcal{H}(n-n_0)v(1-v) - g_2 vc^{\delta}. \tag{9}$$

Here $\mathcal{H}(\cdot)$ is a sigmoid function 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, 18, 20]. They are defined as:

$$\alpha = \frac{\lambda_1 - n}{(\lambda_2 - 1)n + \lambda_1}, \tag{10}$$

$$\beta = \frac{\lambda_2 n}{(\lambda_2 - 1)n + \lambda_1}. \tag{11}$$

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

$$\frac{\partial c}{\partial t} = D\frac{\partial^2 c}{\partial x^2} + bc(1-c), \tag{12}$$

which we denote in the rest of the manuscript as the Fisher-Kolmogorov (FK) model for tumor cell density. We remark that equation (12) has been extensively used to predict untreated glioma kinetics based on patient-specific parameters from standard medical imaging procedures [25, 26, 27, 28, 29].

Eq. (7)-(9) 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 tumour microenvironment, which is in turn coupled to cell density through the oxygen consumption term. The functional vascular 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 is given in Tab. S1.

We solve the system in (7)-(9) by imposing the initial conditions:

$$c(x,0) = c_0\mathcal{H}(x-\varepsilon) \text{ in } 0 \leq x \leq L, \tag{13}$$

$$n(x,0) = n_0 \text{ in } 0 \leq x \leq L, \tag{14}$$

$$v(x,0) = v_0 \text{ in } 0 \leq x \leq L, \tag{15}$$

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 tumour 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 behaviours arise solely due to the interaction terms in Eq. (7)-(9). This assumption results in no-flux boundary conditions of the form:

$$\frac{\partial c}{\partial x}(0,t) = \frac{\partial n}{\partial x}(0,t) = \frac{\partial v}{\partial x}(0,t) = 0, \tag{16}$$

$$\frac{\partial c}{\partial x}(L,t) = \frac{\partial n}{\partial x}(L,t) = \frac{\partial v}{\partial x}(L,t) = 0. \tag{17}$$

Both the full and FK models are used to calculate two clinical outputs, namely the tumor Infiltration Width (IW) and Tumor Size (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 Tab. 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. We use the latter model 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 BaM^3 method.

Probability distribution from the FK model

As described above, 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 pdf $p(\mathbf{Y} = \hat{\mathbf{y}}|\mathbf{X}_m = \mathbf{x}_m^*)$ as in the following. For each couple of clinical outputs $\hat{y}_1^*$ and $\hat{y}_2^*$ (i.e. $\hat{y}_1^* = \mathrm{IW}^*$, $\hat{y}_2^* = \mathrm{TS}^*$) we calculate the area $A_\alpha(\hat{y}_1^*, \hat{y}_2^*)$ over the $(\hat{y}_1, \hat{y}_2)$ plane as $A_\alpha = [(1-\alpha)\hat{y}_1^* < \hat{y}_1 < (1+\alpha)\hat{y}_1^*), (1-\alpha)\hat{y}_2^* < \hat{y}_2 < (1+\alpha)\hat{y}_2^*)]$, where α is a given tolerance (here set equal to $\alpha = 0.05$). Then, we calculate the pdf by dividing A_α by the total area of predicted $\hat{y}_1$ and $\hat{y}_2$ 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-driven pdf $p(\mathbf{Y} = \hat{\mathbf{y}}|\mathbf{X}_u = \mathbf{x}_u^*)$ we use a normal kernel density estimator (KDE) [30, 31] which depends upon all the data points in the patient ensemble. Briefly, the method estimates the shape of the probability $p(\mathbf{Y}, \mathbf{X}_u)$ from which the ensemble entries are drawn through the sum of a kernel function over all the occurrences of the dataset (see the scheme in Fig. S9). The kernel function is characterized by a hyper-parameter, the bandwidth h, which we assume according to Silverman's rule

of thumb

$$h_i = \sigma_i \left[\frac{4}{(d+2)n} \right]^{\frac{1}{d+4}}, \quad i = 1, 2, \ldots d, \tag{18}$$

where d is the number of dimensions, n is the number of observations, and σ_i is the standard deviation of the i-th variate [53]. After calculating $p(\mathbf{Y}, \mathbf{X}_u)$, we specify the realization of a specific patient and calculate the value of $p(\mathbf{Y} = \hat{\mathbf{y}} | \mathbf{X}_u = \mathbf{x}_u^*)$ over the $(\hat{y}_1 = \text{IW}, \hat{y}_2 = \text{TS})$ space of the estimated clinical outputs.

Scoring glioma growth predictions

We calculate the pdfs that appear in equation 1 and calculate for each patient the relative errors d_m and d_b as described the main text. To assess how the BaM^3 method has changed the prediction of the mathematical model we compare the latter quantities: if $|d_b - d_m| \leq \varepsilon d_m$, then there was no change; if $d_b > (1+\varepsilon)d_m$, then the method deteriorated the prediction of the model; if $d_b < (1-\varepsilon)d_m$, then the method improved the prediction of the model. Here, ε is a tolerance used for the comparison, taken to be equal to $\varepsilon = 0.05$.

Calculation of the effective variance

To calculate the effective variance s shown in Fig. 4 we first calculate the mixed central moments Σ_{ij} of the pdf of interest according to the formula

$$\Sigma_{ij} = \int_{y_1} \int_{y_2} (y_1 - \mu_1)(y_2 - \mu_2) f(y_1, y_2) \, dy_1 dy_2, \tag{19}$$

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. (19) we consider $f(y_1, y_2)$ to be the pdf from the mathematical model or from the BaM^3 method depending on whether we are interested in the effective variance s_m or s_b, respectively.

Pdf from the two-compartment model

Messmer and colleagues [38] measured the fraction of labelled B-CLL cells in a court of 18 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

$$f(t) = h(t) + f(0)e^{\frac{-bt}{v_r}} + \frac{e^{\frac{-bt}{v_r}}\left(g(0) - h(t)\right)}{v_r - 1} + \frac{e^{-bt}\left(h(t) - g(0)\right)}{v_r - 1} \tag{20}$$

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. [38] 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. (20) over all their experimental range. We evaluate the fraction of labeled cells at day 50, f_{50}, and build the distribution 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, so that they all share the same modelable pdf.

Pdf from the patients' unmodelables

The data-driven pdf in the CLL example is obtained from four quantities that are measured for each patient during the study, namely the $\mathrm{V_H}$ mutation status ($x_{u,1}$), CD38 expression ($x_{u,2}$), the age ($x_{u,3}$), and the growth rate of white blood cells ($x_{u,4}$). As in the glioma example we build the sub-dataset $(y, \mathbf{x}_u)$, 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 $\hat{y} = f_{50}$.

Code availability

The code used to generate the simulations and the plots is available at `https://github.com/PiMasGit/BaM3-method`

Author contributions

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

Acknowledgments

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). HH and PM 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, HH would like to thank the Volkswagenstiftung for the its support within the "Life?" programm (96732).

References

[1] Helen V Warner, Nikita Sivakumar, Shayn M Peirce, and Matthew J Lazzara. Multiscale computational models of cancer. *Current Opinion in Biomedical Engineering*, 2019.

[2] John H Phan, Chang F Quo, Chihwen Cheng, and May Dongmei Wang. Multiscale integration of-omic, imaging, and clinical data in biomedical informatics. *IEEE reviews in biomedical engineering*, 5:74–87, 2012.

[3] Bruce Alberts, Dennis Bray, Karen Hopkin, Alexander D Johnson, Julian Lewis, Martin Raff, Keith Roberts, and Peter Walter. *Essential cell biology*. Garland Science, 2013.

[4] HP Greenspan. Models for the growth of a solid tumor by diffusion. *Studies in Applied Mathematics*, 51(4):317–340, 1972.

[5] Robert A Gatenby and Edward T Gawlinski. A reaction-diffusion model of cancer invasion. *Cancer research*, 56(24):5745–5753, 1996.

[6] Helen M Byrne. Dissecting cancer through mathematics: from the cell to the animal model. *Nature Reviews Cancer*, 10(3):221–230, 2010.

[7] Philipp M Altrock, Lin L Liu, and Franziska Michor. The mathematics of cancer: integrating quantitative models. *Nature Reviews Cancer*, 15(12):730–745, 2015.

[8] Robyn P Araujo and DL Sean McElwain. A history of the study of solid tumour growth: the contribution of mathematical modelling. *Bulletin of mathematical biology*, 66(5):1039–1091, 2004.

[9] Luigi Preziosi. *Cancer modelling and simulation*. CRC Press, 2003.

[10] Marinka Zitnik, Francis Nguyen, Bo Wang, Jure Leskovec, Anna Goldenberg, and Michael M Hoffman. Machine learning for integrating data

in biology and medicine: Principles, practice, and opportunities. *Information Fusion*, 50:71–91, 2019.

[11] Mark Alber, Adrian Buganza Tepole, William R Cannon, Suvranu De, Salvador Dura-Bernal, Krishna Garikipati, George Karniadakis, William W Lytton, Paris Perdikaris, Linda Petzold, et al. Integrating machine learning and multiscale modeling—perspectives, challenges, and opportunities in the biological, biomedical, and behavioral sciences. *npj Digital Medicine*, 2(1):1–11, 2019.

[12] Ruth E Baker, Jose-Maria Pena, Jayaratnam Jayamohan, and Antoine Jérusalem. Mechanistic models versus machine learning, a fight worth fighting for the biological community? *Biology letters*, 14(5):20170660, 2018.

[13] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*. Wiley-Interscience, USA, 2006.

[14] JCL Alfonso, A Köhn-Luque, T Stylianopoulos, F Feuerhake, A Deutsch, and H Hatzikirou. Why one-size-fits-all vaso-modulatory interventions fail to control glioma invasion: in silico insights. *Scientific reports*, 6:37283, 2016.

[15] Hiroko Ohgaki and Paul Kleihues. Epidemiology and etiology of gliomas. *Acta neuropathologica*, 109(1):93–108, 2005.

[16] A Giese, R Bjerkvig, ME Berens, and M Westphal. Cost of migration: invasion of malignant gliomas and implications for treatment. *Journal of clinical oncology*, 21(8):1624–1636, 2003.

[17] JCL Alfonso, K Talkenberger, M Seifert, B Klink, A Hawkins-Daarud, KR Swanson, H Hatzikirou, and A Deutsch. The biology and mathematical modelling of glioma invasion: a review. *Journal of the Royal Society Interface*, 14(136):20170490, 2017.

[18] Alf Giese, Melinda A Loo, Nhan Tran, Dorothy Haskett, Stephen W Coons, and Michael E Berens. Dichotomy of astrocytoma migration and proliferation. *International journal of cancer*, 67(2):275–282, 1996.

[19] John M Heddleston, Zhizhong Li, Roger E McLendon, Anita B Hjelmeland, and Jeremy N Rich. The hypoxic microenvironment maintains glioblastoma stem cells and promotes reprogramming towards a cancer stem cell phenotype. *Cell cycle*, 8(20):3274–3284, 2009.

[20] Haralampos Hatzikirou, David Basanta, Matthias Simon, K Schaller, and Andreas Deutsch. 'go or grow': the key to the emergence of invasion in tumour progression? *Mathematical medicine and biology: a journal of the IMA*, 29(1):49–65, 2012.

[21] Justin V Joseph, Siobhan Conroy, Kirill Pavlov, Pallavi Sontakke, Tushar Tomar, Ellie Eggens-Meijer, Veerakumar Balasubramaniyan, Michiel Wagemakers, Wilfred FA den Dunnen, and Frank AE Kruyt. Hypoxia enhances migration and invasion in glioblastoma by promoting a mesenchymal shift mediated by the hif1α–zeb1 axis. *Cancer letters*, 359(1):107–116, 2015.

[22] Triantafyllos Stylianopoulos, John D Martin, Vikash P Chauhan, Saloni R Jain, Benjamin Diop-Frimpong, Nabeel Bardeesy, Barbara L Smith, Cristina R Ferrone, Francis J Hornicek, Yves Boucher, et al. Causes, consequences, and remedies for growth-induced solid stress in murine and human tumors. *Proceedings of the National Academy of Sciences*, 109(38):15101–15108, 2012.

[23] Triantafyllos Stylianopoulos, John D Martin, Matija Snuderl, Fotios Mpekris, Saloni R Jain, and Rakesh K Jain. Coevolution of solid stress and interstitial fluid pressure in tumors during progression: implications for vascular collapse. *Cancer research*, 73(13):3833–3841, 2013.

[24] Pietro Mascheroni, Juan Carlos López Alfonso, Maria Kalli, Triantafyllos Stylianopoulos, Michael Meyer-Hermann, and Haralampos

Hatzikirou. On the impact of chemo-mechanically induced phenotypic transitions in gliomas. *Cancers*, 11(5):716, 2019.

[25] JD Murray. *Mathematical biology II: spatial models and biomedical applications*. Springer New York, 2001.

[26] Kristin R Swanson, Russell C Rockne, Jonathan Claridge, Mark A Chaplain, Ellsworth C Alvord, and Alexander RA Anderson. Quantifying the role of angiogenesis in malignant progression of gliomas: in silico modeling integrates imaging and histology. *Cancer research*, 71(24):7366–7375, 2011.

[27] Kristin R Swanson, Ellsworth C Alvord, and JD Murray. Virtual brain tumours (gliomas) enhance the reality of medical imaging and highlight inadequacies of current therapy. *British journal of cancer*, 86(1):14–18, 2002.

[28] Hana LP Harpold, Ellsworth C Alvord Jr, and Kristin R Swanson. The evolution of mathematical modeling of glioma proliferation and invasion. *Journal of Neuropathology & Experimental Neurology*, 66(1):1–9, 2007.

[29] Andrea Hawkins-Daarud, Russell C Rockne, Alexander RA Anderson, and Kristin R Swanson. Modeling tumor-associated edema in gliomas during anti-angiogenic therapy and its impact on imageable tumor. *Frontiers in oncology*, 3:66, 2013.

[30] Hill Peter D. Kernel estimation of a distribution function. *Communications in Statistics-Theory and Methods*, 14(3):605–620, 1985.

[31] Artur Gramacki. *Nonparametric kernel density estimation and its computational aspects*. Springer, 2018.

[32] Yen-Chi Chen. Modal regression using kernel density estimation: A review. *Wiley Interdisciplinary Reviews: Computational Statistics*, 10(4):e1431, 2018.

[33] Thomas J Kipps, Freda K Stevenson, Catherine J Wu, Carlo M Croce, Graham Packham, William G Wierda, Susan O'brien, John Gribben, and Kanti Rai. Chronic lymphocytic leukaemia. *Nature reviews Disease primers*, 3(1):1–22, 2017.

[34] Nicholas Chiorazzi, Kanti R Rai, and Manlio Ferrarini. Chronic lymphocytic leukemia. *New England Journal of Medicine*, 352(8):804–815, 2005.

[35] Xose S Puente and Carlos López-Otín. The evolutionary biography of chronic lymphocytic leukemia. *Nature Genetics*, 45(3):229–231, 2013.

[36] Yoshikane Kikushige, Fumihiko Ishikawa, Toshihiro Miyamoto, Takahiro Shima, Shingo Urata, Goichi Yoshimoto, Yasuo Mori, Tadafumi Iino, Takuji Yamauchi, Tetsuya Eto, et al. Self-renewing hematopoietic stem cell is the primary target in pathogenesis of human chronic lymphocytic leukemia. *Cancer cell*, 20(2):246–259, 2011.

[37] Symeon Savvopoulos, Ruth Misener, Nicki Panoskaltsis, Efstratios N Pistikopoulos, and Athanasios Mantalaris. A personalized framework for dynamic modeling of disease trajectories in chronic lymphocytic leukemia. *IEEE Transactions on Biomedical Engineering*, 63(11):2396–2404, 2016.

[38] Bradley T Messmer, Davorka Messmer, Steven L Allen, Jonathan E Kolitz, Prasad Kudalkar, Denise Cesar, Elizabeth J Murphy, Prasad Koduru, Manlio Ferrarini, Simona Zupo, et al. In vivo measurements document the dynamic cellular kinetics of chronic lymphocytic leukemia b cells. *The Journal of clinical investigation*, 115(3):755–764, 2005.

[39] Igor Kononenko. Machine learning for medical diagnosis: history, state of the art and perspective. *Artificial Intelligence in medicine*, 23(1):89–109, 2001.

[40] Ashnil Kumar, Lei Bi, Jinman Kim, and David Dagan Feng. Machine learning in medical imaging. In *Biomedical Information Technology*, pages 167–196. Elsevier, 2020.

[41] Cynthia Rudin. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, 1(5):206–215, 2019.

[42] David J Albers, Matthew E Levine, Andrew Stuart, Lena Mamykina, Bruce Gluckman, and George Hripcsak. Mechanistic machine learning: how data assimilation leverages physiologic knowledge using bayesian inference to forecast the future, infer the present, and phenotype. *Journal of the American Medical Informatics Association*, 25(10):1392–1401, 2018.

[43] Chiara Nicolò, Cynthia Périer, Melanie Prague, Carine Bellera, Gaëtan MacGrogan, Olivier Saut, and Sébastien Benzekry. Machine learning and mechanistic modeling for prediction of metastatic relapse in early-stage breast cancer. *JCO Clinical Cancer Informatics*, 4:259–274, 2020.

[44] Maziar Raissi and George Em Karniadakis. Hidden physics models: Machine learning of nonlinear partial differential equations. *Journal of Computational Physics*, 357:125–141, 2018.

[45] David Poole and Adrian E Raftery. Inference for deterministic simulation models: the bayesian melding approach. *Journal of the American Statistical Association*, 95(452):1244–1255, 2000.

[46] Ioannis Andrianakis, Ian R Vernon, Nicky McCreesh, Trevelyan J McKinley, Jeremy E Oakley, Rebecca N Nsubuga, Michael Goldstein, and Richard G White. Bayesian history matching of complex infectious disease models using emulation: a tutorial and a case study on hiv in uganda. *PLoS Comput Biol*, 11(1):e1003968, 2015.

[47] Marc C Kennedy and Anthony O'Hagan. Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(3):425–464, 2001.

[48] Katalin Csilléry, Michael GB Blum, Oscar E Gaggiotti, and Olivier François. Approximate bayesian computation (abc) in practice. *Trends in ecology & evolution*, 25(7):410–418, 2010.

[49] Isabelle Guyon and André Elisseeff. An introduction to variable and feature selection. *Journal of machine learning research*, 3(Mar):1157–1182, 2003.

[50] Jie Cai, Jiawei Luo, Shulin Wang, and Sheng Yang. Feature selection in machine learning: A new perspective. *Neurocomputing*, 300:70–79, 2018.

[51] David W Scott. *Multivariate density estimation: theory, practice, and visualization*. John Wiley & Sons, 2015.

[52] Peter J Rousseeuw and Annick M Leroy. *Robust regression and outlier detection*, volume 589. John wiley & sons, 2005.

[53] H Läuter. Silverman, bw: Density estimation for statistics and data analysis. chapman & hall, london–new york 1986, 175 pp.,£ 12.—. *Biometrical Journal*, 30(7):876–877, 1988.