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

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

[1]Braunschweig Integrated Centre of Systems Biology and Helmholtz Center for Infectious Research, 38106 Braunschweig, Germany
[2]Centre for Individualized Infection Medicine, 30625 Hannover, Germany
[3]Institute for Biochemistry, Biotechnology and Bioinformatics, Technische Universität Braunschweig, 38106 Braunschweig, Germany

May 6, 2020

Abstract

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

*To whom correspondence should be addressed.
E-mail: haralampos.hatzikirou@helmholtz-hzi.de

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 result in improved predictions in almost all simulated patients, especially for those with a late clinical presentation. We finally discuss some limitations of the proposed BaM^3 method and suggest future research directions.

1 Introduction

Biomedical problems are complex in nature, commonly involving several variables ranging from molecular to tissue scale [1, 2]. Such complexity is inherent in living systems, since they are composed by several continuously interacting components between them and also with their environment while responding to chemical and physical stimuli [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]. These models can focus on the dynamics of biological systems irrespective of the spatial distribution of the underlying components (as 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 seldom 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 (BaM^3) to couple mathematical modeling and statistical learning, 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 BaM^3 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.

The key idea of the method and why it works

1.1 Problem setting and notations

Let's assume a random variable (r.v.) triplet $(\mathbf{Y}, \mathbf{X}_m, \mathbf{X}_u) \in \mathbb{R}^M \times \mathbb{R}^K \times \mathbb{R}^L$ 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) and the

associated observed clinical outputs $\mathbf{Y}$. Let's also consider that at time t_0 a new patient is presented and clinical examinations imply the following r.v. realizations $(\mathbf{X}_m = \mathbf{x}_m^*, \mathbf{X}_u = \mathbf{x}_u^*)$. The overall goal is to predict the patient's clinical output by an estimate $\mathbf{Y} = \hat{\mathbf{y}}$. The true clinical outputs of the patient will be denoted as $\mathbf{y}$. Moreover, we consider the existence of a $N-$patient ensemble dataset $(\mathbf{y}, \mathbf{x}_m, \mathbf{x}_u) \in \mathbb{R}^{N\times M} \times \mathbb{R}^{N\times K} \times \mathbb{R}^{N\times L}$. In this dataset all the variables (i.e. modelables, unmodelables and clinical outputs) are recorded at the time of diagnosis t_d, and differ from patient to patient. In turn, we introduce t_0 as the clinical presentation time of a *new* patient at which the patient specific $\mathbf{x}_m^*$ and $\mathbf{x}_u^*$ are obtained (see the schema in Fig. 1**A**). The latter state should be distinguished from the ensemble diagnosis time t_d.

Statement of the main result

The overall goal is to estimate the probability distribution of the outputs $\hat{\mathbf{y}}$ upon the knowledge of patient's data $(\mathbf{x}_m^*, \mathbf{x}_u^*)$. The core idea is to consider the predictions of the mechanistic 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 Supplementary Information (SI), we can prove that:

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

Therefore, the implementation of the BaM^3 method reduces to the calculation of the aforementioned probability distributions. Although, the prediction probability distribution function (pdf) of the clinical outputs is rather straightforward for the mathematical model (see the SI), the probability distribution of the patient's unmodelable data based on outputs is not trivial. The key idea is to use a density estimator method upon the patient ensemble dataset to derive $p(\mathbf{Y}, \mathbf{X}_u)$, and then perform a nonparametric regression for the specific patient realization $\mathbf{X}_u = \mathbf{x}_u^*$.

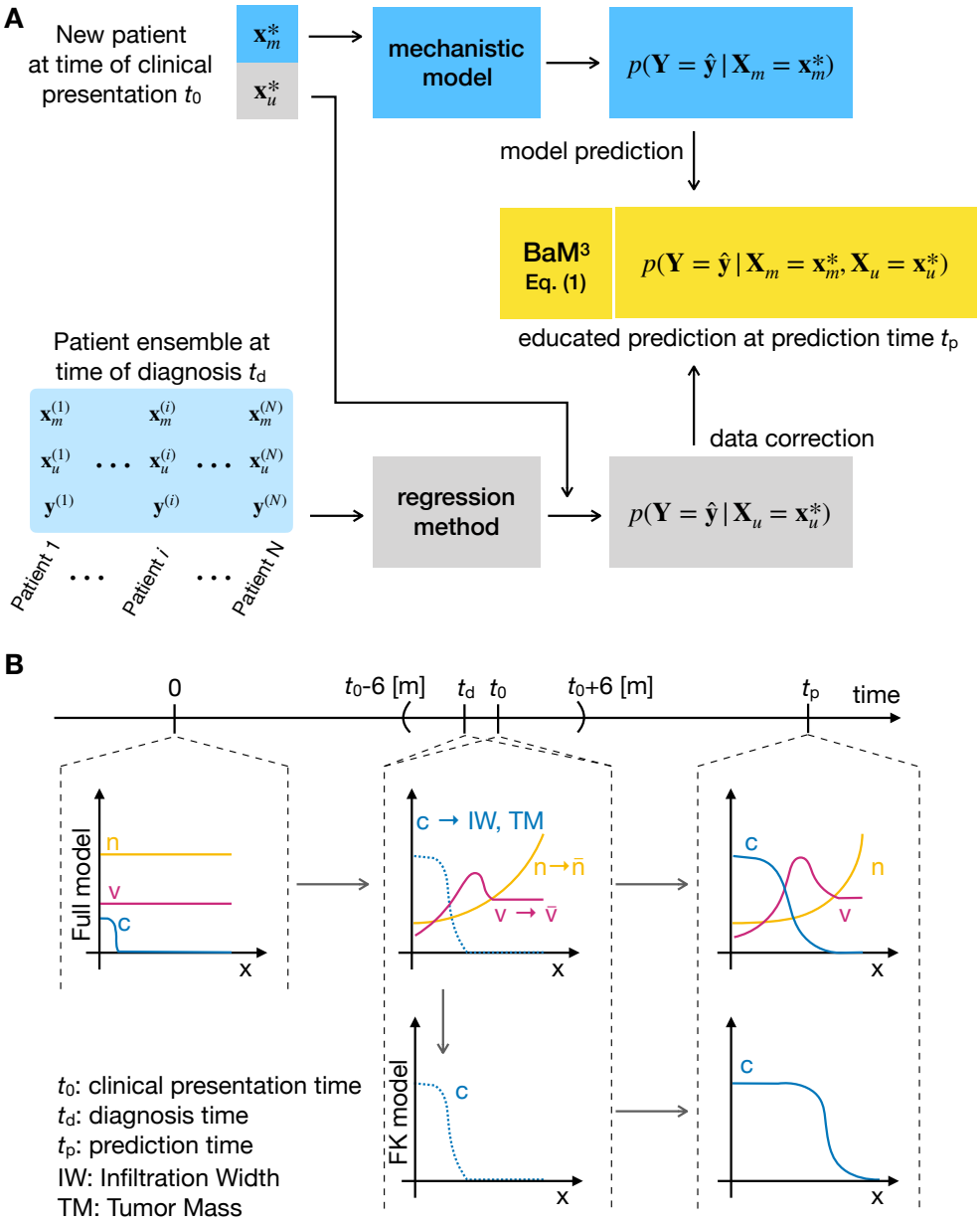
A
New patient at time of clinical presentation t_0
$\mathbf{x}_m^*$
$\mathbf{x}_u^*$
mechanistic model
$p(\mathbf{Y} = \hat{\mathbf{y}} \,|\, \mathbf{X}_m = \mathbf{x}_m^*)$
model prediction
BaM³ Eq. (1)
$p(\mathbf{Y} = \hat{\mathbf{y}} \,|\, \mathbf{X}_m = \mathbf{x}_m^*, \mathbf{X}_u = \mathbf{x}_u^*)$
educated prediction at prediction time t_p
data correction
Patient ensemble at time of diagnosis t_d
$\mathbf{x}_m^{(1)}$ $\mathbf{x}_m^{(i)}$ $\mathbf{x}_m^{(N)}$
$\mathbf{x}_u^{(1)}$ $\mathbf{x}_u^{(i)}$ $\mathbf{x}_u^{(N)}$
$\mathbf{y}^{(1)}$ $\mathbf{y}^{(i)}$ $\mathbf{y}^{(N)}$
Patient 1
Patient i
Patient N
regression method
$p(\mathbf{Y} = \hat{\mathbf{y}} \,|\, \mathbf{X}_u = \mathbf{x}_u^*)$
B
0
t_0-6 [m]
t_d
t_0
t_0+6 [m]
t_p
time
Full model
n
v
c
x
c → IW, TM
n→n̄
v → v̄
FK model
t_0: clinical presentation time
t_d: diagnosis time
t_p: prediction time
IW: Infiltration Width
TM: Tumor Mass

Figure 1 *(previous page)*: **A** A 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 time of observation 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 introduced in a nonparametric regression method to correct the predictions from the mathematical model, providing the probability distribution function (pdf) $p(\mathbf{Y} = \hat{\mathbf{y}}|\mathbf{X}_u = \mathbf{x}_u^*)$. 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 regression method. **B** 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 normalized Tumor Mass (TM), 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 and diagnosis t_d. The latter is assumed to be a random time between $t_0 \pm 6$ months. In addition, the spatial profile of tumor concentration at time t_0 is used as the initial condition for the Fisher–Kolmogorov (FK) model. The FK model is then used to simulate tumor growth until the prediction time t_p and outputs compared to those from the mathematical model used to generate the synthetic reality (full model).

Here we would like to point out that the only assumption made to derive the expression 1 was the statistical independence of the following:

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

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 evidences 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). This can be naively understood as the BaM^3 estimator having a larger variance than an estimator not involving that assumption of independence, i.e. less precision.

Rationale

We would like to provide some arguments on our rational, i.e. why BaM^3 improves prediction performance. As a cost function that quantifies prediction performance, we consider the mean square error (MSE) given by $(\hat{\mathbf{y}} - \mathbf{y})^2$. As stated above, the main idea is to use the mechanistic model predictions as an informative prior. In turn, the unmodelable data should contribute to the corresponding likelihood term. The corresponding Bayesian formula, after the assumption (2), leads to Eq. (1) which can be alternatively written as:

$$p(\mathbf{Y}|\mathbf{X}_m, \mathbf{X}_u) \propto p_{unmodel} \times p_{model}, \tag{4}$$

where the prior is $p_{model} = p(\mathbf{Y}|\mathbf{X}_m)$, and the corresponding unmodelable-based likelihood is $p_{unmodel} = p(\mathbf{Y}|\mathbf{X}_u)$.

To illustrate the potential of the proposed BaM^3 method, let's assume an estimator of clinical outputs based on the full input dataset $(\mathbf{x}_m^*, \mathbf{x}_u^*)$,

e.g. using a certain machine learning method. Any estimator is dictated by the Cramer-Rao inequality [13] which provides a lower bound for the MSE:

$$\langle(\hat{\mathbf{y}} - \mathbf{y})^2\rangle \geq \frac{1}{\mathrm{I}_{full}}, \tag{5}$$

where $\mathrm{I}_{full} = \mathrm{I}(\hat{\mathbf{y}})$ is the corresponding Fisher Information (FI) of the full dataset estimator. Thus, even for unbiased estimators this lower bound cannot be less than inverse of FI which is always a finite number.

In the BaM^3 method, we can improve the above lower bound by adding an informative prior based on mathematical modeling. In particular, by using the van Trees inequality [14], known as Bayesian Cramer-Rao bound, and assuming an increasing model-based FI denoted as I_{model}, we can deduce that the lower bound vanishes:

$$\langle(\hat{\mathbf{y}} - \mathbf{y})^2\rangle \geq \lim_{\mathrm{I}_{model}\to\infty} \frac{1}{\mathrm{I}_{umodel} + \mathrm{I}_{model}} = 0.$$

The expression $\mathrm{I}_{umodel} \leq \mathrm{I}_{full}$ denotes the FI that unmodelable data achieve. The last formula shows that the proposed BaM^3 method has the potential to minimize the MSE lower bound and outscore any data-driven method by harnessing the benefits of a "good" mathematical model.

Testing the BaM^3 method on synthetic tumor data

As a proof of concept, we synthetically generated a dataset of patients using a system of PDEs recently published [15]. We present below the main steps of the derivation, leaving the complete details in the SI. The equations of the selected mathematical model as use case ("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, 17, 18, 19, 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 [21, 22, 23]. We generate a total of $N = 675$ 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 ($K = N_h$, where N_h is the numerical discretization of the spatial domain). 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 ($L = 2$). Starting from the same initial conditions, we simulate the behavior of each virtual patient for 3 years, storing the variable values at each month.

As sketched in Fig. 1**B**, 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 observation time t_0, and use it as the initial condition for a Fisher–Kolmogorov equation [24, 25, 26, 27, 28] ("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 observables ($M = 2$), i.e. the normalized Tumor Mass (TM) and Infiltration Width (IW), as detailed in the SI. The first quantity is related to tumor size, whereas the second quantifies accounts for tumor infiltration in the host tissue. Finally, 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 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 $\mathbf{x}_m = [c(x_i, t), i = 1, \ldots, N_h]$, $\mathbf{x}_u = (\bar{n}, \bar{v})$ and $\mathbf{y}$ =(TM, IW). At a new patient presentation at a prediction time t_p, the goal is to calculate the probability of observing certain clinical outputs (i.e. TM and IW), given the patient-specific modelable ($c(\mathbf{x}, t_0)$) and unmodelable ($\bar{n}, \bar{v}$) variables at the clinical presentation time t_0. The data-driven pdf in the equation (1) is obtained through a normal kernel density estimator [29, 30] 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 TM predicted from the mathematical model for each virtual patient (see SI for further information about the regression and model steps).

Improving the predictions of the mathematical model

Figs. 2**A**-**C** show the results of applying the proposed BaM^3 method to a representative patient, considering the amount of oxygen and vasculature present in the tissue as unmodelable variables $(\bar{n}, \bar{v})$. For this case, we selected a time of observation $t_0 = 2$ years and a time of prediction $t_p = 9$ months after. The conditional probability $p(\mathbf{Y} = \hat{\mathbf{y}}|\mathbf{X}_m = \mathbf{x}_m^*)$ obtained from the FK model is shown in Fig. 2**A**. Interestingly, the prediction of the model in that particular case shows two peaks, one with low TM and high IW and the other one with the opposite behavior. We then calculate the expected values of TM and IW from the distribution obtained with the FK model and compare it to the "true" values given by the full model. As shown in the inset, for this patient the presence of a bimodal distribution shifts the expected values far from the true ones. We multiply each point in the (TM, IW) space by the probability calculated in the regression step, shown in Fig. 2**B**. The latter probability 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 for the probability mode closer to the true IW and TM values, as displayed in Fig. 2**C**. These results evidence the ability of the BaM^3 method to correct the predictions obtained by using exclusively the mathematical model, producing an expected value of the pdf that is closer to the ground truth.

Then, we apply the BaM^3 method to each virtual patient in the dataset for two clinical presentation times at 1 and 2 years, and compare its outcomes with those provided by the full model at increasing prediction times

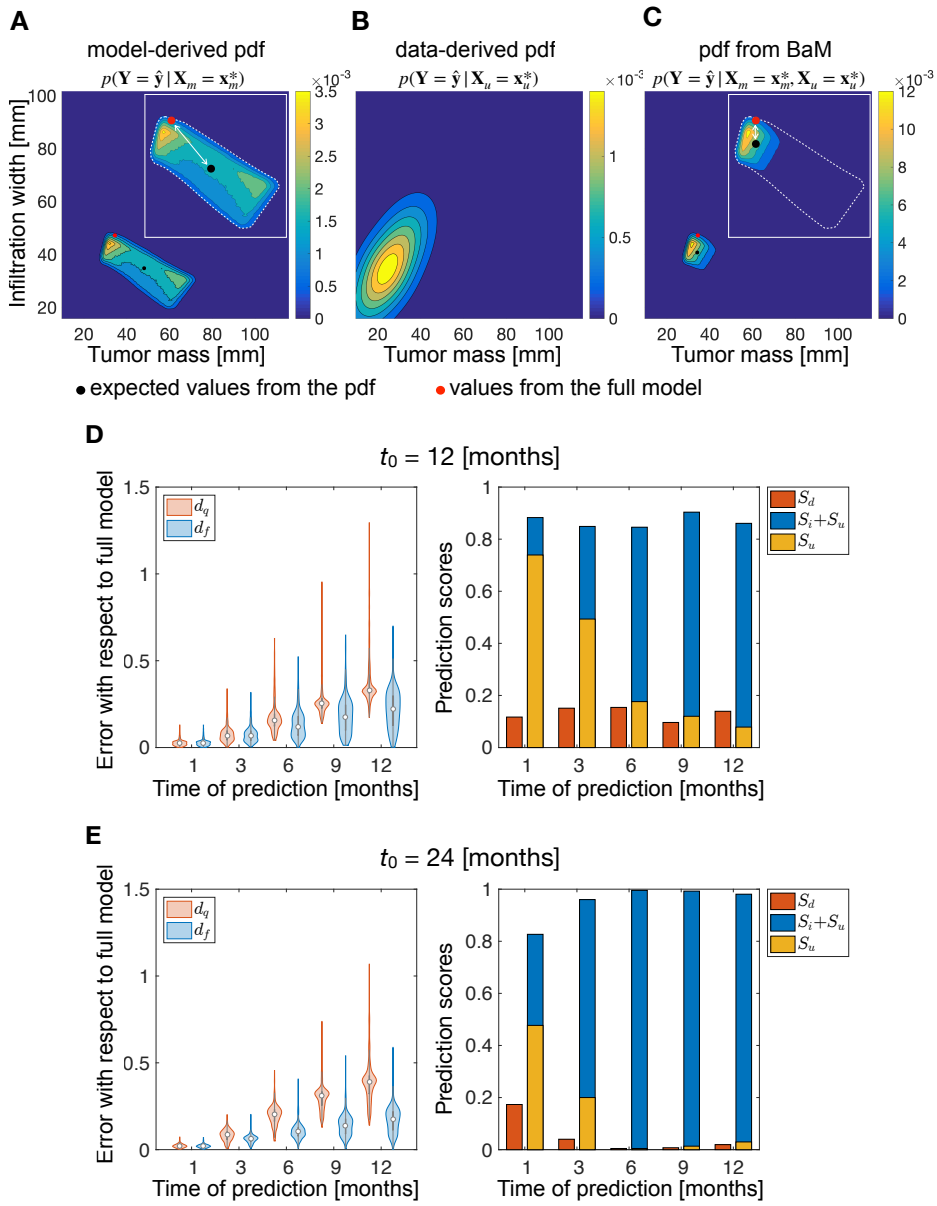
A
model-derived pdf
$p(\mathbf{Y} = \hat{\mathbf{y}} \,|\, \mathbf{X}_m = \mathbf{x}_m^*)$
B
data-derived pdf
$p(\mathbf{Y} = \hat{\mathbf{y}} \,|\, \mathbf{X}_u = \mathbf{x}_u^*)$
C
pdf from BaM
$p(\mathbf{Y} = \hat{\mathbf{y}} \,|\, \mathbf{X}_m = \mathbf{x}_m^*, \mathbf{X}_u = \mathbf{x}_u^*)$
Infiltration width [mm]
Tumor mass [mm]
expected values from the pdf
values from the full model
D
t_0 = 12 [months]
Error with respect to full model
Prediction scores
Time of prediction [months]
d_q
d_f
S_d
S_i+S_u
S_u
E
t_0 = 24 [months]

Figure 2 *(previous page)*: **A-C** Application of the BaM^3 method to a representative patient. The time of observation is 2 years and the prediction time is 9 months later, considering the integral over the tissue of the oxygen concentration ($\bar{n}$) and vasculature density ($\bar{v}$) as unmodelable variables. **A** Probability distribution obtained from the FK model, plotted over the (IW, TM) space. The inset shows an enlargement of the probability region, in which the white arrow marks the error between the observable values 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 by multiplying the probabilities displayed in **A** and **B** for each point in the (TM, IW) space. The dashed line in the inset shows the original probability obtained from the FK model. Quantification of model performance at a clinical presentation time t_0 of 1 (**D**) and 2 (**E**) years. (*left*) Relative errors between the values of clinical outputs obtained from the full model and the expected values given by FK model (d_q) and after the application of the BaM^3 method (d_f).(*right*) 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 prediction times range from 1 to 12 months.

(Fig. 2**D**,**E**). 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_q) and after implementing the BaM^3 method (d_f). These nondimensional errors are calculated as

$$d_k = \sqrt{\frac{1}{L}\sum_{i=1}^{N_y}\left(1-\frac{\langle y_i^k\rangle}{y_i^r}\right)^2} \tag{6}$$

where $k = q, f$, $\langle y_i^k\rangle$ are the expected observable values (i =IW, TM) calculated from the FK model and the BaM^3 method ($k = q, f$, respectively), and y_i^r are the observable values obtained from the full model. We calculate the errors d_q and d_f for each patient at different prediction times. Then, we compare the corresponding errors for each patient and evaluate if the BaM^3 method improved, deteriorated or left unchanged the prediction from the FK model, i.e. $d_f < d_q$, $d_f > d_q$ or $d_f \sim d_q$, respectively (see SI for further details). 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_q and d_f augment for increasing prediction times, as shown in the left side of Fig. 2**D**,**E** for the two different 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 variance 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.

The stacked bars on the right side of Fig. 2**D**,**E** show that the correction in estimated observables performs well at early prediction times and remarkably well at later times (improvement ratio S_i close to 1), especially at the latest presentation time t_0 (Fig. 2**E**). For a presentation time of 1 year (Fig. 2**D**) there is a fraction of cases in which the correction deteriorates the prediction or does not change it considerably ($S_d \sim 0.2$ and nonzero S_u). Still, the correction performs good, as the sum of the cases in which the approach

has improved or not changed the prediction is above 80% in all cases. For the clinical presentation time of 2 years (Fig. 2**E**) 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. In summary, the BaM^3 method is able to correct the FK model predictions for most of the patients, particularly at later presentation and prediction times. The improvement in the prediction occurs by (i) decreasing the median relative error between expected observable values and ground truth and (ii) decreasing the rate at which the error increases with prediction time.

Failure in improving TM 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_q and d_f into the two partial errors

$$\Delta \mathrm{IW}_k = \sqrt{\left(1 - \frac{\langle \mathrm{IW}^k \rangle}{\mathrm{IW}^r}\right)^2}, \quad \Delta \mathrm{TM}_k = \sqrt{\left(1 - \frac{\langle \mathrm{TM}^k \rangle}{\mathrm{TM}^r}\right)^2} \tag{7}$$

where $k = q, f$, $\langle \mathrm{IW}^k \rangle$ and $\langle \mathrm{TM}^k \rangle$ are the expected values of the clinical outputs obtained from the mathematical model and BaM^3 pdfs ($k = q, f$, respectively), and $\mathrm{IW}^r, \mathrm{TM}^r$ are the "real" values of these quantities from the full model. Fig. 3 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 Fig. 3**A**), 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 TM ($\Delta\text{TM}_f > \Delta\text{TM}_q$). Interestingly, the same failure cases belong to regions in which $\Delta\text{IW}_f < \Delta\text{IW}_q$: the BaM^3 method is improving the IW predictions and at the same time deteriorating the TM predictions. This happens for both $t_0 = 12$ and 24 months, however the number of failure cases is considerably higher for the lower presentation time. Lower performance of the BaM^3 method is therefore associated to its inability in correcting the FK model predictions for TM, 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 the regression step, i.e. $p(\mathbf{Y} = \hat{\mathbf{y}}|\mathbf{X}_u = \mathbf{x}_u^*)$ in Equation (1). Fig. 4 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. 4**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,TM) 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. 4**B**, showing the marginal probabilities for IW (left) and TM (right) calculated from the distribution in Fig. 4**A**. The marginal distributions become broader for both IW and TM, but in the first case their peak stabilizes at later t_0 times. On the contrary,

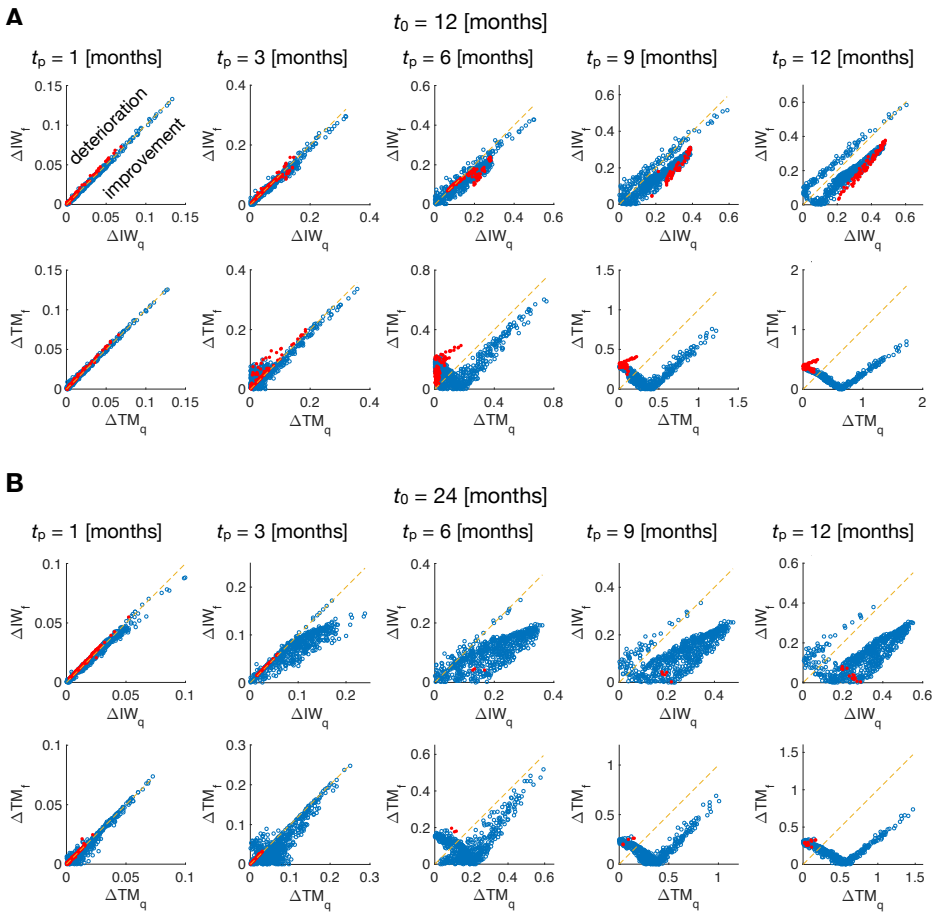

Figure 3: Scatter plots of the partial errors $\Delta\mathrm{IW}_i, \Delta\mathrm{TM}_i$ over different prediction times t_p ($i = q, f$, for the quantity related to the mechanistic model or the BaM^3 method, respectively). Results for the presentation times of $t_0 = 12$ months (**A**) and $t_0 = 24$ months (**B**) are shown. Each red dot corresponds to a patient for which the BaM^3 method fails in improving the FK model predictions. The first row in **A** and **B** compares the partial error relative to the IW, whereas the second compares the error for the TM. Regions over the dashed line highlight 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.

the peak of the marginal probability for TM 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. 4**C**, for the overlap between the distributions at $t_0 = 12$ and 18 months (left) and between $t_0 = 24$ and 30 months (right). Here, IW_r and TM_r are calculated as the area of overlap for the IW and TM marginal distributions, respectively. Values close to one represent maximum overlap, whereas values close to 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 $t_0 = 12 - 18$ months case in Fig. 4**C**(left) the patients are mostly scattered along a line of increasing IW_r and TM_r with points where the overlap is poor (close to 0.4 in certain regions). On the other hand, in the case of $t_0 = 24 - 30$ months (right) 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. The data-driven pdf from the regression step changes over time and stabilizes 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 regression step that should correct the model predictions is projecting the model pdf over (IW,TM) 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 TM might be wrong is some cases, the pdf for the IW has stabilized and points towards the corrected value. In most of the cases the correction for the IW outperforms the one for the TM which leads to a general improvement of predictions by the BaM^3 method.

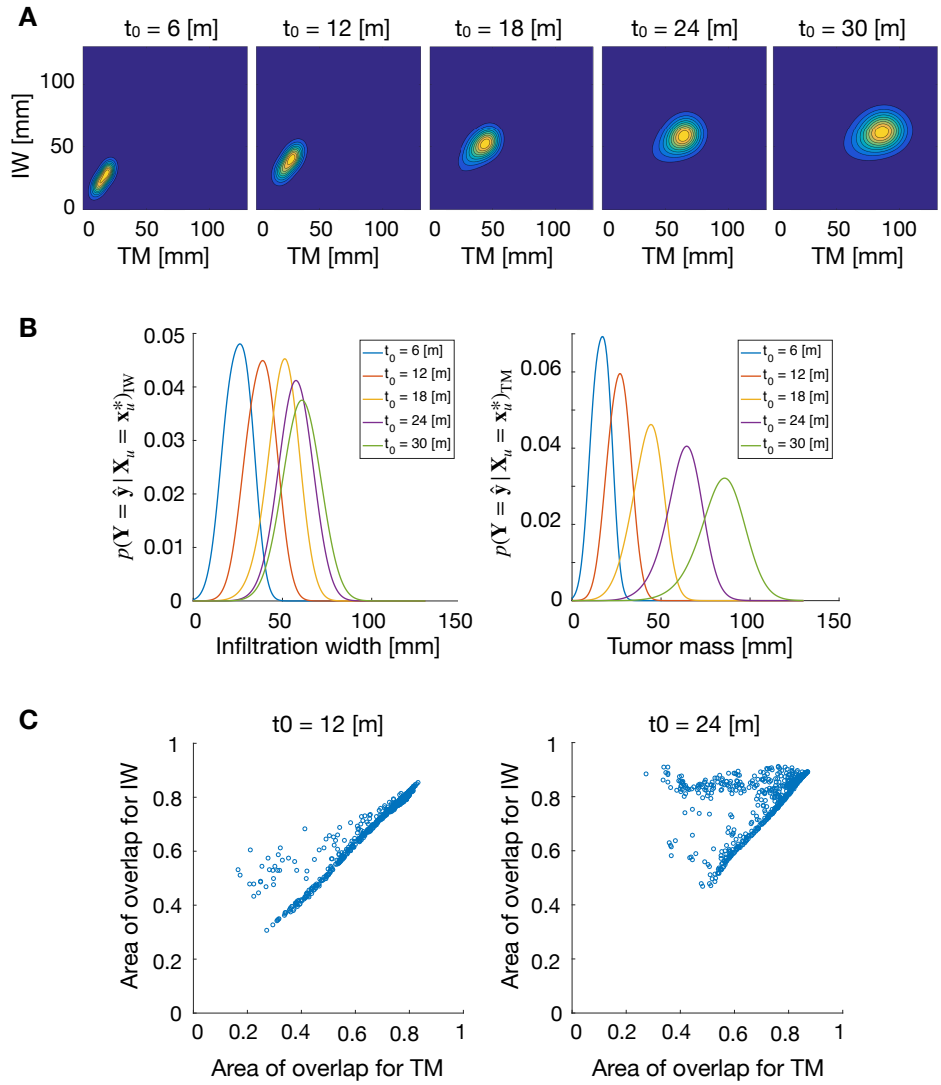

Figure 4: Temporal evolution of the unmodelable pdf from the regression step. **A** Data-driven pdf for a representative patient at different clinical presentation times, plotted over the (IW,TM) space. **B** Marginal probabilities from the plots in **A**, showing the evolution of the probability distribution for IW (left) and TM (right). **C** Scatter plots representing the degree of overlap between the marginal probabilities for IW and TM between the prediction times $t_0 = 12$ and 18 months (right) and $t_0 = 24$ and 30 months (left).

Patients at the extremes of the parameter space are the most challenging

To conclude the analysis, we investigate the distribution of the failure cases over the full model parameter space. Patients for which the BaM^3 method deteriorates the model prediction are represented as red dots in Fig. 5. Here each patient is identified by the couple of parameters describing cell motility and proliferation rate. We checked for the distribution of failure cases also for the other model parameters, but no particular pattern was evident (see Figure S3 in the SI). In general, the BaM^3 method performs poorly for those patients that are at the extremes of the parameter space. In particular the top-right and bottom-right parts of the plots, i.e. patients with high motility - high proliferation rates and high motility - low proliferation rates, witness the highest number of failure cases for both $t_0 = 12$ and 24 months. Notably, patients falling into these parameter regions show the highest values for IW and TM (see [15] and Figure S4). Highly invasive and massive neoplasms are inadequately described by the regression step, 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, TM.

Discussion

In the last few years, machine learning has emerged as a promising methodology in the biomedical field [31, 32]. 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 [33]. The urge for interpretability advocates for novel approaches, able to shed light

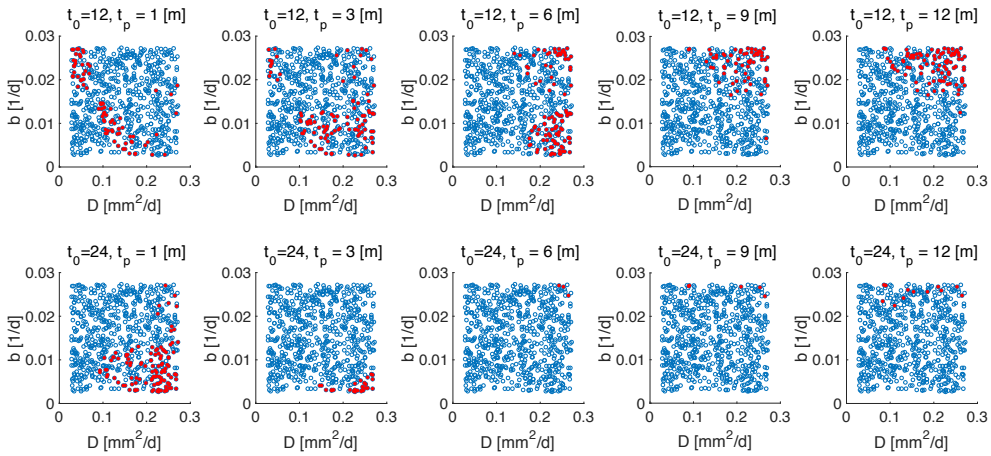

Figure 5: Scatter plots showing the distribution of failure cases (red dots) over the parameter space at different presentation and prediction times. D: tumor cell motility; b: tumor cell proliferation rate; t_0: clinical presentation time; t_p: prediction time.

on the machine decision processes [34, 35, 36]. In this work, we presented a novel method (BaM^3) to couple mathematical modeling and nonparametric regression in a Bayesian framework. By unifying these complementary paths, the mechanistic understanding of mathematical models is empowered with the versatility of regression methods. 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 Fisher-Kolmogorov model for the majority of the virtual patients, with significant improvements at later clinical presentation times. When analyzing the cases for which the BaM^3 method did not improve the results of the mathematical model, we came across some limitations that should be addressed when applying the BaM^3 method 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, 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.

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 regression method. We adopted a well known form of nonparametric regression through kernel density estimation, but other approaches could be tailored to a specific problem - especially when high-dimensional datasets come into play. The modularity of the BaM^3 method makes it extremely versatile, allowing one to change the regression step, the modeling part or both of them at the same time to improve the final prediction scores.

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.

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] Harald Cramér. *Mathematical methods of statistics*, volume 43. Princeton university press, 1999.

[14] Richard D Gill, Boris Y Levit, et al. Applications of the van trees inequality: a bayesian cramér-rao bound. *Bernoulli*, 1(1-2):59–79, 1995.

[15] 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.

[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] 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.

[18] 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.

[19] 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.

[20] 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.

[21] 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.

[22] 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.

[23] 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.

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

[25] 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.

[26] 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.

[27] 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.

[28] 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.

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

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

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

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

[33] 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.

[34] 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.

[35] 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.

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