Bayesian combination of mechanistic modeling and machine learning (BaM³): improving clinical tumor growth predictions

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

¹Braunschweig Integrated Centre of Systems Biology and Helmholtz Center for
Infectious Research, 38106 Braunschweig, Germany

²KU Leuven, Department of Chemical Engineering, Celestijnenlaan 200F, 3001 Leuven,
Belgium

³Centre for Individualized Infection Medicine, 30625 Hannover, Germany

⁴Institute for Biochemistry, Biotechnology and Bioinformatics, Technische Universität
Braunschweig, 38106 Braunschweig, Germany

⁵Mathematics Department, Khalifa University, P.O. Box 127788, Abu Dhabi, UAE

⁶Center 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 (BaM³). We evaluate the proposed BaM³ 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

1 Biomedical problems are complex in nature, commonly involving several play-
 2 ers ranging from molecular to tissue scale [1, 2]. Such complexity is inherent
 3 in living systems and stems from the continuous interactions between their
 4 same individual parts and their environment [3]. When facing biomedical
 5 problems, and in particular cancer, scientists have been long supported by
 6 the use of mathematical modeling as a tool to identify causal relationships
 7 in the experimental and clinical data [4, 5, 6]. Mathematical models allow to
 8 propose and test biological hypotheses, analyze the sensitivity of observables
 9 with respect to biological parameters, and provide insights into the mecha-
 10 nistic details governing the phenomenon of interest [7]. Mathematical models
 11 can focus on the dynamics of biological systems irrespective of the spatial
 12 distribution of the underlying components (by employing, for instance, Ordi-
 13 nary Differential Equations). On the other hand, models can also take spatial

14 effects into account in a discrete (e.g. through Agent Based, Cellular Au-
15 tomata models) or continuous (e.g. through Partial Differential Equations)
16 framework [8, 9]. Although these models can be extremely powerful both
17 in predicting system responses and suggesting new experimental directions,
18 they require some knowledge of the underlying mechanisms of the analysed
19 system. Commonly, this knowledge is not complete, and only for a limited
20 portion of the involved variables the corresponding mechanistic interactions
21 are known. Therefore, although mathematical models often provide a good
22 description of a simplified version of the associated system dynamics, they
23 not always allow for accurate and universal predictions.

24 On the other hand, machine learning techniques are suitable to deal with
25 the inherent complexity of biomedical problems, but without caring for the
26 knowledge of the underlying interactions [10]. While mathematical models
27 rely on causality, statistical learning methods identify correlations among
28 data [11]. This approach allows to systemically process large amounts of
29 data and infer hidden patterns in biological systems. As a consequence,
30 machine learning-based techniques can provide valuable predictive accuracy
31 upon sufficient training, but do not typically allow for any mechanistic insight
32 into the investigated problem [12]. As a result, the overall understanding of
33 the fundamental system dynamics becomes almost impossible, as the chance
34 to generalize the 'learnt' system behavior.

35 In this work, we propose a novel Bayesian method to combine mathe-
36 matical modeling and statistical learning (BaM³), taking advantage of their
37 respective strengths. As a proof of concept, the proposed BaM³ method is
38 tested on a synthetic dataset of brain tumor growth. We analyze the method
39 performance in predicting two relevant clinical outcomes, namely tumor bur-
40 den and infiltration. When comparing predictions from the mechanistic
41 model with those from the BaM³ method we obtain improved predictions
42 for the vast majority of virtual patients. In addition, we apply the approach
43 to a dataset of patients suffering from Chronic Lymphocytic Leukemia. The
44 BaM³ method shows excellent agreement between the predicted clinical out-

45 put and the reported data.

46 Results

47 We introduce the notation and key ideas of the proposed methodology, leav-
 48 ing the full derivation and additional details to the Supplementary Informa-
 49 tion (SI). We start by assuming a random variable (r.v.) triplet $(\mathbf{Y}, \mathbf{X}_m, \mathbf{X}_u)$
 50 that denotes the system’s modelable $\mathbf{X}_m$, unmodelable $\mathbf{X}_u$ variables/data
 51 (e.g. patient’s age or sex, results of different ‘-omics’ techniques, etc.) and
 52 the associated observed clinical outputs $\mathbf{Y}$. We then introduce t_0 as the clin-
 53 ical presentation time of a *new* patient at which the patient specific r.v. re-
 54 alizations $(\mathbf{X}_m = \mathbf{x}_m^*, \mathbf{X}_u = \mathbf{x}_u^*)$ are obtained (see the scheme in Fig. 1). The
 55 overall goal of the method is to predict the patient’s clinical output by an
 56 estimate $\mathbf{Y} = \hat{\mathbf{y}}$ at a certain prediction time t_p . The true clinical outputs
 57 of the patient will be denoted as $\mathbf{y}$. Moreover, we consider the existence of
 58 a N -patient ensemble dataset $(\mathbf{y}, \mathbf{x}_m, \mathbf{x}_u)$. In this dataset all the variables
 59 (i.e. modelables, unmodelables and clinical outputs) are recorded at the time
 60 of diagnosis t_d , which might differ from one patient to another. Both t_0 and
 61 t_d are calculated from the onset of the disease. We introduce two distinct
 62 times to account for the variability of the disease stage among different pa-
 63 tients (t_d) and the time at which a specific patient is presented to the clinic
 64 (t_0). The core idea of the method is to consider the predictions of the math-
 65 ematical model $p(\mathbf{Y} = \hat{\mathbf{y}} | \mathbf{X}_m = \mathbf{x}_m^*)$ as an informative Bayesian prior of the
 66 posterior distribution $p(\mathbf{Y} = \hat{\mathbf{y}} | \mathbf{X}_m = \mathbf{x}_m^*, \mathbf{X}_u = \mathbf{x}_u^*)$. By using the arguments
 67 described in the SI, we can prove that:

$$68 \quad p(\mathbf{Y} = \hat{\mathbf{y}} | \mathbf{X}_m = \mathbf{x}_m^*, \mathbf{X}_u = \mathbf{x}_u^*) \propto p(\mathbf{Y} = \hat{\mathbf{y}} | \mathbf{X}_m = \mathbf{x}_m^*) p(\mathbf{Y} = \hat{\mathbf{y}} | \mathbf{X}_u = \mathbf{x}_u^*). \quad (1)$$

69 The implementation of the BaM³ method therefore reduces to the calculation
 70 of the aforementioned probability distributions. Although the prediction of
 71 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, BaM³) 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. :

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

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

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

This highlights that the entropy S_{method} produced by the proposed BaM³ 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 BaM³ prediction and the conditional entropy that reads [13]:

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

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 BaM³ 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 .

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

130 According to our previous notation, we have therefore $\mathbf{x}_m = c(x, t)$,
131 $\mathbf{x}_u = (\bar{n}, \bar{v})$ and $\mathbf{y} = (\text{TS}, \text{IW})$. Given the patient-specific modelable and
132 unmodelable variables ($c(\mathbf{x}, t_0)$ and $\bar{n}, \bar{v}$, respectively) at the clinical pre-
133 sentation time t_0 , the BaM³ method produces the probability of observing
134 certain clinical outputs (i.e. TS and IW) at a specific prediction time t_p .
135 The data-driven pdf in the equation (1) is obtained through a normal kernel
136 density estimator (KDE) [30, 31, 31, 32] trained with the patient ensemble.
137 Then, the first factor in equation (1) is calculated using the FK model for
138 each patient, running the simulations over the whole parameter range of cell
139 motility and proliferation rates. This produces a map of possible IW and
140 TS predicted from the mathematical model for each virtual patient (see the
141 Methods for further information about the KDE and modeling steps).

142 Improving the predictions of the mathematical model

143 Figs. 3A-C show the results of applying the BaM³ method to a representative
144 patient. We select a clinical presentation time $t_0 = 24$ months and a time
145 of prediction $t_p = 9$ months. The conditional probability $p(\mathbf{Y} = \hat{\mathbf{y}} | \mathbf{X}_m =$
146 $\mathbf{x}_m^*)$ obtained from the FK model is shown in Fig. 3A. Interestingly, the
147 prediction of the model in that particular case shows two peaks, one with
148 low TS and high IW and another with opposite properties. We calculate the
149 expected values of TS and IW from the pdf obtained with the FK model and
150 compare it to the 'true' values given by the full model. As shown in the plot,
151 for this patient the presence of a bimodal distribution shifts the expected
152 values far from the true ones. We enforce the BaM³ method making use
153 of the probability calculated from the KDE, shown in Fig. 3B. The latter
154 pdf takes into account the correlations between the clinical outputs and the
155 unmodelable variables present in the patient ensemble. For this patient,
156 the unmodelable distribution selects the probability mode closer to the true
157 IW and TS values, as displayed in Fig. 3C (another example for a different

Figure 3 (*previous page*): **A-C** Application of the BaM³ method to a representative patient. The clinical presentation time t_0 is of 24 months and the prediction time t_p is 9 months. **A** Pdf obtained from the FK model, plotted over the (TS, IW) space. The inset shows an enlargement of the probability region, in which the white arrow marks the error between the clinical output predicted by the full model (red dot) and the expected value of the distribution (black dot). **B** Data-driven pdf for the specific patient calculated through the kernel density estimator trained over the patient ensemble. **C** Probability distribution obtained from the BaM³ method. The dashed line shows the original probability obtained from the FK model. Relative errors between the clinical outputs obtained from the full model and the expected values given by the FK model (d_m), and after the application of the BaM³ method (d_b), for a clinical presentation time of 12 and 24 months (**D,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 BaM³ 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.

158 patient is given in Fig. S2 in the SI). These results evidence the ability of
 159 the proposed method to correct the predictions obtained by using exclusively
 160 the mathematical model, and to produce an expected value of the pdf that
 161 is closer to the ground truth.

162 We apply the BaM³ method for two clinical presentation times at 12 and
 163 24 months, and compare its outcomes with those provided by the full model
 164 at increasing prediction times (Fig. 3D,I). For each patient we calculate the
 165 relative error between the predicted clinical outputs obtained from the full
 166 model and the expected values of the pdf calculated from the FK model (d_m)
 167 and after implementing the BaM³ method (d_b). These nondimensional errors
 168 are calculated as

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

169 where $k = m, b$, $\langle y_i^k \rangle$ are the expected observable values ($i = \text{IW, TS}$) calcu-
 170 lated from the FK model and the BaM³ method, and y_i^r are the observable
 171 values obtained from the full model. We calculate the errors d_m and d_b for
 172 each patient at different prediction times. Then, we compare the correspond-
 173 ing errors and evaluate if the BaM³ method improved, deteriorated or left
 174 unchanged the prediction from the FK model, i.e. $d_b < d_m$, $d_b > d_m$ or
 175 $d_b \sim d_m$, respectively (see Methods). We denote the ratio of improved, un-
 176 changed and deteriorated cases with respect to the total number of simulated
 177 patients as S_i , S_u and S_d , respectively.

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

185 We also calculate the effective variance of the predictions as the logarithm
 186 of the determinant of the covariance matrices relative to the model and BaM³

187 pdfs, (identified by s_m , s_b respectively - see Methods). This quantity reflects
 188 the spreading of the pdfs over the (TS, IW) plane, with higher values denoting
 189 more uncertainty in the predictions. For both clinical presentation times
 190 (Fig. 3E,H) the BaM³ method provides thinner pdfs, more centered around
 191 their expected value with respect to the FK model-derived case.

192 Finally, the stacked bars in Fig. 3F,I show that BaM³ performs well at
 193 later prediction times and remarkably well (improvement ratio S_i close to 1),
 194 especially at the latest clinical presentation time $t_0 = 24$ months. For $t_0 =$
 195 12 months (Fig. 3F) the proposed method is not able to improve predictions
 196 until a prediction time of 6 months. Then, for $t_p = 6, 9$ and 12 month the
 197 advantages of using BaM³ over the FK model are unambiguous. On the
 198 other hand, for the clinical presentation time of 24 months (Fig. 3I) both
 199 S_u and S_d decrease significantly for prediction times equal or greater than
 200 3 months. The ratio of improved cases S_i reaches almost 100% at each of
 201 the last three prediction times, clearly overcoming the results of the FK
 202 model. The error bars in Fig. 3F,I denote the variability in the results that
 203 is obtained by replicating the study 10 times, each with $N = 500$ randomly
 204 generated patients. Fig. S3 in the SI shows similar results when decreasing
 205 the number of patients. Notably, the scores for $N = 500, 250, 100$ and 50 are
 206 very close, slightly improving with increasing the number of patients. The
 207 variability in the 10 replicates also decreases for higher values of N .

208 We also calculate the prediction scores using the distribution mode to
 209 generate the scores instead of the expected value (see Fig. S4). When the
 210 pdfs display multiple maxima we consider the average of the relative errors
 211 between the values of the full model and the different peaks. The perfor-
 212 mance of the BaM³ method sensibly degrades with respect to using the ex-
 213 pected value. Improvement in predictions is observed only for later clinical
 214 presentation and prediction times.

215 In summary, the BaM³ method is able to correct the FK model predictions
 216 for most of the patients, particularly at later clinical presentation and predic-
 217 tion times. The improvement in the prediction occurs by (i) decreasing the

218 median relative error between expected observable values and ground truth,
 219 (ii) decreasing the rate at which the error increases with prediction time, and
 220 (iii) decreasing the variance associated with the probability distributions.

221 **Failure in improving TS model predictions reduces BaM³** 222 **performance**

223 Even though the BaM³ method performs well for the majority of patients,
 224 there are some cases for which it fails to improve the predictions of the
 225 mathematical model. We analyzed the failure cases by splitting the errors
 226 d_m and d_b into the two partial errors

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

227 where $k = m, b$, $\langle IW^k \rangle$ and $\langle TS^k \rangle$ are the expected values of the clinical
 228 outputs obtained from the mathematical model and BaM³ pdfs ($k = m, b$,
 229 respectively), and IW^r, TS^r are the values of these quantities from the full
 230 model. Fig. 4 shows how these partial errors are distributed over the presen-
 231 tation and prediction times. The dashed line in the plots highlights the neu-
 232 tral boundary, where the partial errors of the FK model and BaM³ method
 233 are equal. Above this line, the proposed BaM³ method deteriorates the
 234 model predictions, whereas under that line the BaM³ method improves pre-
 235 dictions. The red dots in the scatter plots represent the patients for which
 236 the BaM³ method fails ('failure cases', in the following). After a predic-
 237 tion time of one month, in which a characteristic pattern is not evident,
 238 the plots highlight that failure cases are generally associated with regions
 239 where the BaM³ method under-performs to the FK model with respect to
 240 TS ($\Delta TS_b > \Delta TS_m$). Interestingly, the same failure cases belong to regions
 241 in which $\Delta IW_b < \Delta IW_m$: the BaM³ method is improving the IW predictions
 242 and at the same time deteriorating the TS predictions. This happens for both
 243 $t_0 = 12$ and 24 months, however the number of failure cases is considerably

244 higher for the earlier presentation time. For the specific case under consider-
 245 ation, lower performance of the BaM³ method is therefore associated to its
 246 inability in correcting the FK model predictions for TS, with a tendency that
 247 improves for the later presentation time due to strong corrections for IW.

248 **Transient behavior of the unmodelable distribution is** 249 **associated to limited improvements**

250 To investigate the reasons for the poor performance of the BaM³ method in
 251 improving the predictions for one of the clinical observables, we analyzed the
 252 behavior of the pdf arising from density estimation, i.e. $p(\mathbf{Y} = \hat{\mathbf{y}} | \mathbf{X}_u = \mathbf{x}_u^*)$
 253 in Equation (1). Fig. 5 shows the temporal evolution of $p(\mathbf{Y} = \hat{\mathbf{y}} | \mathbf{X}_u = \mathbf{x}_u^*)$
 254 for different clinical presentation times t_0 . As done previously for $t_0 = 12$
 255 and 24 months, each entry of the ensemble of patients from which the pdf is
 256 drawn is evaluated at a diagnosis time t_d taken between ± 6 months around
 257 the presentation time t_0 . Fig. 5A shows a plot of the unmodelable pdf for a
 258 representative patient over the clinical output space. From a pdf that covers
 259 a limited region in the (IW, TS) plane, the probability distribution spreads
 260 over a broader area as the presentation time increases. The center of mass
 261 of the distribution, however, tends to converge to a more specific region as
 262 time progresses. This is more evident in Fig. 5B,C, showing the marginal
 263 probabilities for IW and TS, calculated from the distribution in Fig. 5A.
 264 The marginal distributions become broader for both IW and TS, but in the
 265 first case their peak stabilizes at later t_0 times. On the contrary, the peak of
 266 the marginal probability for TS moves towards larger values at higher times.
 267 To quantify this behavior across the different patients we then evaluated
 268 the degree of overlap between the marginal probabilities at two subsequent
 269 t_0 . Results from this calculation are plotted in Fig. 5D,E for the overlap
 270 between the distributions at presentation times t_0 of 12 and 18 months and
 271 between 24 and 30 months. Here, the degree of overlap is calculated as the
 272 area of overlap for the IW and TS marginal distributions. Values close to one

Figure 4: Scatter plots of the partial errors $\Delta IW_i, \Delta TS_i$ over different prediction times t_p ($i = m, b$, for the quantity related to the mechanistic model or the BaM³ method, respectively). Results for presentation times of $t_0 = 12$ months (**A**) and $t_0 = 24$ months (**B**) are shown. Each red dot corresponds to a patient for which the BaM³ method fails in improving the FK model predictions. Regions over the dashed line correspond to areas in which the BaM³ method deteriorates the FK model predictions, while regions under that line show areas in which it improves the FK predictions.

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. 5D 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. 5E the patients are shifted towards higher values of overlap. Moreover, a horizontal line of high overlap for the IW output is visible for a large patient ensemble, pointing to a stabilization towards a steady state for the IW at later presentation times. This explains the lower BaM³ 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 BaM³ method for the TS might be wrong in some cases, the pdf for the IW has stabilized and points towards the correct value. In most of the cases the correction for the IW outperforms the one for the TS which leads to a general improvement of predictions by the BaM³ method.

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 BaM³ 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 BaM³ method, since in the latter the correction from the dataset points towards smaller values of IW and, especially, TS.

Applying the method to real CLL patients

In addition to the proof of concept applied to *in silico* data, we test the BaM³ 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. 7A). 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 BaM³ 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. 7B, where we compare

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 BaM³ method is applied. **B** Comparison between the expected values of the pdf obtained from the BaM³ method (blue dots) and the reported data (orange dots) [38]. Error bars denote the standard deviation of the pdf.

the BaM³ predicted values against the patient f_{50} values in [38]. The fraction of labeled cells predicted by the BaM³ method agrees well with the reported data. The majority of data points lies within the standard deviation estimated in the generated distributions from BaM³. However, there is a remarkable error with BaM³ 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 parametrization 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

362 machine decision processes [42, 43, 44]. In this work, we presented a novel
 363 method (BaM³) to couple mathematical modeling and density estimation in a
 364 Bayesian framework. By unifying these complementary paths, the mechanis-
 365 tic understanding of mathematical models is empowered with the versatility
 366 of statistical methods.

367 Compared to other methodologies that combine outputs from mathemat-
 368 ical models and measured data - such as Bayesian Melding [45], History
 369 Matching [46], Bayesian Model Calibration [47] or Approximate Bayesian
 370 Computation [48] - the BaM³ method is not interested in parameter esti-
 371 mation to better calibrate the mathematical model. Instead, its goal is to
 372 improve predictions of mathematical models empowering them with knowl-
 373 edge from variables that are not usually considered (the 'unmodelables', in
 374 our framework). This is done without the exact knowledge of the parameters
 375 of the mathematical model - indeed we calculated the pdfs of the modelable
 376 variables using a uniform sampling of the parameter space. Better estimation
 377 of the model parameters might improve the outcomes of the method, but it
 378 is not required for the methodology to be applied.

379 The BaM³ method was tested on a synthetic dataset of patients focusing
 380 on tumor growth dynamics. Our approach was able to improve the predic-
 381 tions of a FK model for the majority of the virtual patients, with significant
 382 improvements at later clinical presentation times. In addition, we tested
 383 the proposed methodology on a dataset of CLL patients and compared the
 384 outcomes of the BaM³ method to the reported data. When analyzing the
 385 cases for which the performance of the BaM³ method was not optimal, we
 386 came across some limitations that should be addressed when applying the
 387 methodology to real cases.

388 The first limitation regards the selection of the proper unmodelable vari-
 389 ables. These are quantities that cannot be easily mathematically modeled,
 390 but can be correlated to the patient clinical outputs. For our proof of con-
 391 cept, we took only two unmodelables for the first example and four for the
 392 second. However, in principle multiple quantities could be considered at the

393 same time. Moreover, the most important unmodelables could be selected
 394 in a process of feature selection similar to the ones usually adopted in ma-
 395 chine learning, providing better accuracy for the predictions [49, 50]. In
 396 addition, the method is open to progress in knowledge: Should an unmode-
 397 lable variable become modelable because of an increased understanding of the
 398 biological mechanisms, this variable can change side and become modelable.

399 Care must be taken with the selection of the metric that should be im-
 400 proved by the BaM³ method. For the *in silico* case, for example, considering
 401 the expected value of the final pdf resulted in better method performance
 402 when compared to selecting the pdf mode (see Fig. 3 and Fig. S4). This was
 403 probably due to the very similar natures of the FK and full models. Indeed,
 404 for lower clinical presentation times the FK model is already 'primed' to-
 405 wards the correct solution (in terms of outcomes of the full model); applying
 406 the BaM³ method might result in adding noise to the FK prediction, degrad-
 407 ing the final prediction. However, for some patients the FK model provides
 408 pdfs with multiple local maxima, sometimes far away from the full model
 409 values. In these instances (see Fig. S2) the BaM³ method is able to correct
 410 for the right peak, shifting the pdf to the correct values. Therefore, a good
 411 practice would be to try multiple pdf metrics and test the BaM³ method on
 412 each of them. This would result in a more thorough picture of the problem,
 413 eventually allowing for better predictions.

414 Another important issue is that the correlations between unmodelables
 415 and clinical outputs should be persistent over time, evolving on a timescale
 416 that is faster than the dynamics of the problem. In our synthetic dataset this
 417 was partially accomplished at later clinical prediction times, especially for
 418 the case of the tumor infiltration width. Indeed, the unmodelables variables
 419 need to provide as much as possible time-invariant information on the clinical
 420 output variables, implying an equilibrated pdf. Such data can be for instance
 421 from genetic origin (such as mutations) or from other variables with slow
 422 characteristic evolution time. We stress that it is the probability distribution
 423 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 BaM³ method makes it extremely versatile, allowing one to change the density estimation step, the modeling part or both of them at the same time to improve the final prediction scores.

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 BaM³ approach. For instance, predicting the fate of renal grafts by using pre- and post transplantation data is a prime application of our proposed methodology.

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

452 cells $c(x, t)$, the concentration of oxygen $n(x, t)$ and the density of functional
453 vasculature $v(x, t)$. For simplicity we consider a one-dimensional computa-
454 tional domain. The variables are non-dimensional and the equations read:

$$455 \quad \frac{\partial c}{\partial t} = D \frac{\partial^2}{\partial x^2} \left(\frac{\alpha}{\alpha_0} c \right) + b \frac{\beta}{\beta_0} c(1 - c), \quad (7)$$

$$456 \quad \frac{\partial n}{\partial t} = D_n \frac{\partial^2 n}{\partial x^2} + h_1 v(n_0 - n) - h_2 c n, \quad (8)$$

$$457 \quad \frac{\partial v}{\partial t} = D_v \frac{\partial^2 v}{\partial x^2} + g_1 \mathcal{H}(n - n_0) v(1 - v) - g_2 v c^\delta. \quad (9)$$

459 Here $\mathcal{H}(\cdot)$ is a sigmoid function allowing for tumor angiogenesis in hypoxic
460 conditions, i.e. for $n < n_0$ where n_0 is the hypoxic oxygen threshold. Then,
461 the functions $\alpha = \alpha(n)$ and $\beta = \beta(n)$ account for the dependence of cellular
462 motility and proliferation on the oxygen level, respectively [16, 18, 20]. They
463 are defined as:

$$464 \quad \alpha = \frac{\lambda_1 - n}{(\lambda_2 - 1)n + \lambda_1}, \quad (10)$$

$$465 \quad \beta = \frac{\lambda_2 n}{(\lambda_2 - 1)n + \lambda_1}. \quad (11)$$

467 When the oxygen level is fixed to the maximum level $n = 1$ in the tissue
468 $\alpha = \alpha_0$ and $\beta = \beta_0$, so that the equation for c reduces to

$$469 \quad \frac{\partial c}{\partial t} = D \frac{\partial^2 c}{\partial x^2} + bc(1 - c), \quad (12)$$

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

474 Eq. (7)-(9) define an extended version of the FK equation, enriched with
475 nonlinear glioma cell diffusion and proliferation terms. The latter terms
476 depend on the oxygen concentration in the tumour microenvironment, which
477 is in turn coupled to cell density through the oxygen consumption term. The
478 functional vascular density controls the supply of oxygen to the tissue. Blood

479 vessel density increases due to tumor angiogenesis and decreases because of
480 vaso-occlusion by high tumor cell density. The values of the parameters used
481 in the simulations and their descriptions is given in Tab. S1.

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

$$483 \quad c(x, 0) = c_0 \mathcal{H}(x - \varepsilon) \text{ in } 0 \leq x \leq L, \quad (13)$$

$$484 \quad n(x, 0) = n_0 \text{ in } 0 \leq x \leq L, \quad (14)$$

$$485 \quad v(x, 0) = v_0 \text{ in } 0 \leq x \leq L, \quad (15)$$

486
487 where the positive parameters c_0 , n_0 and v_0 are the initial density of glioma
488 cells spatially distributed in a segment of length ε , the density of functional
489 tumour vasculature and the oxygen concentration, respectively. Then, $L > 0$
490 is the length of the one-dimensional computational domain. In addition, we
491 consider an isolated host tissue in which all system behaviours arise solely
492 due to the interaction terms in Eq. (7)-(9). This assumption results in no-flux
493 boundary conditions of the form:

$$494 \quad \frac{\partial c}{\partial x}(0, t) = \frac{\partial n}{\partial x}(0, t) = \frac{\partial v}{\partial x}(0, t) = 0, \quad (16)$$

$$495 \quad \frac{\partial c}{\partial x}(L, t) = \frac{\partial n}{\partial x}(L, t) = \frac{\partial v}{\partial x}(L, t) = 0. \quad (17)$$

496
497 Both the full and FK models are used to calculate two clinical outputs,
498 namely the tumor Infiltration Width (IW) and Tumor Size (TS). The IW at
499 a specific time is defined by the difference between the points where glioma
500 cell density is 80% and 2% of the maximum cellular density. In turn, the TS
501 is obtained by integrating the spatial profile of tumor density and dividing
502 it for the maximum value of the latter.

503 We run the full model and simulate the growth of the tumor for N pa-
504 tients, each one from a parameter set taken randomly from a uniform dis-
505 tribution over the parameter range. We run simulations for $N = 50, 100, 250$
506 and 500 with 10 repetitions within each N-case. To generate the patients,
507 we vary five parameters in the list in Tab. S1, namely the tumor motility D ,
508 proliferation rate b , oxygen consumption h_1 , vascular formation and occlu-
509 sion 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 BaM³ 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. $\hat{y}_1^* = \text{IW}^*$, $\hat{y}_2^* = \text{TS}^*$) 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 hyperparameter, the bandwidth h , which we assume according to Silverman's rule

538 of thumb

$$539 \quad h_i = \sigma_i \left[\frac{4}{(d+2)n} \right]^{\frac{1}{d+4}}, \quad i = 1, 2, \dots, d, \quad (18)$$

540 where d is the number of dimensions, n is the number of observations, and σ_i
 541 is the standard deviation of the i -th variate [53]. After calculating $p(\mathbf{Y}, \mathbf{X}_u)$,
 542 we specify the realization of a specific patient and calculate the value of
 543 $p(\mathbf{Y} = \hat{\mathbf{y}} | \mathbf{X}_u = \mathbf{x}_u^*)$ over the ($\hat{y}_1 = \text{IW}, \hat{y}_2 = \text{TS}$) space of the estimated
 544 clinical outputs.

545 Scoring glioma growth predictions

546 We calculate the pdfs that appear in equation 1 and calculate for each patient
 547 the relative errors d_m and d_b as described the main text. To assess how the
 548 BaM³ method has changed the prediction of the mathematical model we
 549 compare the latter quantities: if $|d_b - d_m| \leq \varepsilon d_m$, then there was no change;
 550 if $d_b > (1 + \varepsilon)d_m$, then the method deteriorated the prediction of the model;
 551 if $d_b < (1 - \varepsilon)d_m$, then the method improved the prediction of the model.
 552 Here, ε is a tolerance used for the comparison, taken to be equal to $\varepsilon = 0.05$.

553 Calculation of the effective variance

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

$$556 \quad \Sigma_{ij} = \int_{y_1} \int_{y_2} (y_1 - \mu_1)(y_2 - \mu_2) f(y_1, y_2) dy_1 dy_2, \quad (19)$$

557 where y_1 and y_2 are the clinical outputs (IW and TS, respectively) and μ_1, μ_2
 558 the expected values of the corresponding variables. The elements of Σ form a
 559 symmetric two-dimensional matrix, for which we calculate the determinant.
 560 We define the effective variance s as the natural logarithm of the latter deter-
 561 minant. In Eq. (19) we consider $f(y_1, y_2)$ to be the pdf from the mathematical
 562 model or from the BaM³ method depending on whether we are interested in
 563 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

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

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 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

- 639 in biology and medicine: Principles, practice, and opportunities. *Information Fusion*, 50:71–91, 2019.
- 640
- 641 [11] Mark Alber, Adrian Buganza Tepole, William R Cannon, Suvranu
642 De, Salvador Dura-Bernal, Krishna Garikipati, George Karniadakis,
643 William W Lytton, Paris Perdikaris, Linda Petzold, et al. Integrat-
644 ing machine learning and multiscale modeling—perspectives, challenges,
645 and opportunities in the biological, biomedical, and behavioral sciences.
646 *npj Digital Medicine*, 2(1):1–11, 2019.
- 647 [12] Ruth E Baker, Jose-Maria Pena, Jayaratnam Jayamohan, and Antoine
648 Jérusalem. Mechanistic models versus machine learning, a fight worth
649 fighting for the biological community? *Biology letters*, 14(5):20170660,
650 2018.
- 651 [13] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*
652 (*Wiley Series in Telecommunications and Signal Processing*). Wiley-
653 Interscience, USA, 2006.
- 654 [14] JCL Alfonso, A Köhn-Luque, T Stylianopoulos, F Feuerhake,
655 A Deutsch, and H Hatzikirou. Why one-size-fits-all vaso-modulatory
656 interventions fail to control glioma invasion: in silico insights. *Scientific*
657 *reports*, 6:37283, 2016.
- 658 [15] Hiroko Ohgaki and Paul Kleihues. Epidemiology and etiology of gliomas.
659 *Acta neuropathologica*, 109(1):93–108, 2005.
- 660 [16] A Giese, R Bjerkvig, ME Berens, and M Westphal. Cost of migration:
661 invasion of malignant gliomas and implications for treatment. *Journal*
662 *of clinical oncology*, 21(8):1624–1636, 2003.
- 663 [17] JCL Alfonso, K Talkenberger, M Seifert, B Klink, A Hawkins-Daarud,
664 KR Swanson, H Hatzikirou, and A Deutsch. The biology and mathe-
665 matical modelling of glioma invasion: a review. *Journal of the Royal*
666 *Society Interface*, 14(136):20170490, 2017.

- 667 [18] Alf Giese, Melinda A Loo, Nhan Tran, Dorothy Haskett, Stephen W
668 Coons, and Michael E Berens. Dichotomy of astrocytoma migration
669 and proliferation. *International journal of cancer*, 67(2):275–282, 1996.
- 670 [19] John M Heddleston, Zhizhong Li, Roger E McLendon, Anita B Hjelmel-
671 land, and Jeremy N Rich. The hypoxic microenvironment maintains
672 glioblastoma stem cells and promotes reprogramming towards a cancer
673 stem cell phenotype. *Cell cycle*, 8(20):3274–3284, 2009.
- 674 [20] Haralampos Hatzikirou, David Basanta, Matthias Simon, K Schaller,
675 and Andreas Deutsch. ‘go or grow’: the key to the emergence of invasion
676 in tumour progression? *Mathematical medicine and biology: a journal*
677 *of the IMA*, 29(1):49–65, 2012.
- 678 [21] Justin V Joseph, Siobhan Conroy, Kirill Pavlov, Pallavi Sontakke,
679 Tushar Tomar, Ellie Eggens-Meijer, Veerakumar Balasubramaniyan,
680 Michiel Wagemakers, Wilfred FA den Dunnen, and Frank AE Kruijdt.
681 Hypoxia enhances migration and invasion in glioblastoma by promoting
682 a mesenchymal shift mediated by the hif1 α –zeb1 axis. *Cancer letters*,
683 359(1):107–116, 2015.
- 684 [22] Triantafyllos Stylianopoulos, John D Martin, Vikash P Chauhan, Sa-
685 loni R Jain, Benjamin Diop-Frimpong, Nabeel Bardeesy, Barbara L
686 Smith, Cristina R Ferrone, Francis J Hornicek, Yves Boucher, et al.
687 Causes, consequences, and remedies for growth-induced solid stress in
688 murine and human tumors. *Proceedings of the National Academy of*
689 *Sciences*, 109(38):15101–15108, 2012.
- 690 [23] Triantafyllos Stylianopoulos, John D Martin, Matija Snuderl, Fotios
691 Mpekris, Saloni R Jain, and Rakesh K Jain. Coevolution of solid stress
692 and interstitial fluid pressure in tumors during progression: implications
693 for vascular collapse. *Cancer research*, 73(13):3833–3841, 2013.
- 694 [24] Pietro Mascheroni, Juan Carlos López Alfonso, Maria Kalli, Tri-
695 antafyllos Stylianopoulos, Michael Meyer-Hermann, and Haralampos

- 696 Hatzikirou. On the impact of chemo-mechanically induced phenotypic
697 transitions in gliomas. *Cancers*, 11(5):716, 2019.
- 698 [25] JD Murray. *Mathematical biology II: spatial models and biomedical ap-*
699 *plications*. Springer New York, 2001.
- 700 [26] Kristin R Swanson, Russell C Rockne, Jonathan Claridge, Mark A
701 Chaplain, Ellsworth C Alvord, and Alexander RA Anderson. Quan-
702 tifying the role of angiogenesis in malignant progression of gliomas:
703 in silico modeling integrates imaging and histology. *Cancer research*,
704 71(24):7366–7375, 2011.
- 705 [27] Kristin R Swanson, Ellsworth C Alvord, and JD Murray. Virtual brain
706 tumours (gliomas) enhance the reality of medical imaging and highlight
707 inadequacies of current therapy. *British journal of cancer*, 86(1):14–18,
708 2002.
- 709 [28] Hana LP Harpold, Ellsworth C Alvord Jr, and Kristin R Swanson. The
710 evolution of mathematical modeling of glioma proliferation and invasion.
711 *Journal of Neuropathology & Experimental Neurology*, 66(1):1–9, 2007.
- 712 [29] Andrea Hawkins-Daarud, Russell C Rockne, Alexander RA Anderson,
713 and Kristin R Swanson. Modeling tumor-associated edema in gliomas
714 during anti-angiogenic therapy and its impact on imageable tumor.
715 *Frontiers in oncology*, 3:66, 2013.
- 716 [30] Hill Peter D. Kernel estimation of a distribution function. *Communica-*
717 *tions in Statistics-Theory and Methods*, 14(3):605–620, 1985.
- 718 [31] Artur Gramacki. *Nonparametric kernel density estimation and its com-*
719 *putational aspects*. Springer, 2018.
- 720 [32] Yen-Chi Chen. Modal regression using kernel density estimation: A
721 review. *Wiley Interdisciplinary Reviews: Computational Statistics*,
722 10(4):e1431, 2018.

- 723 [33] Thomas J Kipps, Freda K Stevenson, Catherine J Wu, Carlo M Croce,
724 Graham Packham, William G Wierda, Susan O'brien, John Gribben,
725 and Kanti Rai. Chronic lymphocytic leukaemia. *Nature reviews Disease*
726 *primers*, 3(1):1–22, 2017.
- 727 [34] Nicholas Chiorazzi, Kanti R Rai, and Manlio Ferrarini. Chronic lym-
728 phocytic leukemia. *New England Journal of Medicine*, 352(8):804–815,
729 2005.
- 730 [35] Xose S Puente and Carlos López-Otín. The evolutionary biography of
731 chronic lymphocytic leukemia. *Nature Genetics*, 45(3):229–231, 2013.
- 732 [36] Yoshikane Kikushige, Fumihiko Ishikawa, Toshihiro Miyamoto, Takahiro
733 Shima, Shingo Urata, Goichi Yoshimoto, Yasuo Mori, Tadafumi Iino,
734 Takuji Yamauchi, Tetsuya Eto, et al. Self-renewing hematopoietic stem
735 cell is the primary target in pathogenesis of human chronic lymphocytic
736 leukemia. *Cancer cell*, 20(2):246–259, 2011.
- 737 [37] Symeon Savvopoulos, Ruth Misener, Nicki Panoskaltsis, Efstratios N
738 Pistikopoulos, and Athanasios Mantalaris. A personalized framework
739 for dynamic modeling of disease trajectories in chronic lymphocytic
740 leukemia. *IEEE Transactions on Biomedical Engineering*, 63(11):2396–
741 2404, 2016.
- 742 [38] Bradley T Messmer, Davorka Messmer, Steven L Allen, Jonathan E
743 Kolitz, Prasad Kudalkar, Denise Cesar, Elizabeth J Murphy, Prasad
744 Koduru, Manlio Ferrarini, Simona Zupo, et al. In vivo measurements
745 document the dynamic cellular kinetics of chronic lymphocytic leukemia
746 b cells. *The Journal of clinical investigation*, 115(3):755–764, 2005.
- 747 [39] Igor Kononenko. Machine learning for medical diagnosis: history, state
748 of the art and perspective. *Artificial Intelligence in medicine*, 23(1):89–
749 109, 2001.

- 750 [40] Ashnil Kumar, Lei Bi, Jinman Kim, and David Dagan Feng. Machine
751 learning in medical imaging. In *Biomedical Information Technology*,
752 pages 167–196. Elsevier, 2020.
- 753 [41] Cynthia Rudin. Stop explaining black box machine learning models
754 for high stakes decisions and use interpretable models instead. *Nature*
755 *Machine Intelligence*, 1(5):206–215, 2019.
- 756 [42] David J Albers, Matthew E Levine, Andrew Stuart, Lena Mamykina,
757 Bruce Gluckman, and George Hripcsak. Mechanistic machine learning:
758 how data assimilation leverages physiologic knowledge using bayesian
759 inference to forecast the future, infer the present, and phenotype. *Journal of the American Medical Informatics Association*, 25(10):1392–1401,
760 2018.
761
- 762 [43] Chiara Nicolò, Cynthia Périer, Melanie Prague, Carine Bellera, Gaëtan
763 MacGrogan, Olivier Saut, and Sébastien Benzekry. Machine learning
764 and mechanistic modeling for prediction of metastatic relapse in early-
765 stage breast cancer. *JCO Clinical Cancer Informatics*, 4:259–274, 2020.
- 766 [44] Maziar Raissi and George Em Karniadakis. Hidden physics models:
767 Machine learning of nonlinear partial differential equations. *Journal of*
768 *Computational Physics*, 357:125–141, 2018.
- 769 [45] David Poole and Adrian E Raftery. Inference for deterministic simula-
770 tion models: the bayesian melding approach. *Journal of the American*
771 *Statistical Association*, 95(452):1244–1255, 2000.
- 772 [46] Ioannis Andrianakis, Ian R Vernon, Nicky McCreesh, Trevelyan J
773 McKinley, Jeremy E Oakley, Rebecca N Nsubuga, Michael Goldstein,
774 and Richard G White. Bayesian history matching of complex infectious
775 disease models using emulation: a tutorial and a case study on hiv in
776 uganda. *PLoS Comput Biol*, 11(1):e1003968, 2015.

- 777 [47] Marc C Kennedy and Anthony O’Hagan. Bayesian calibration of com-
778 puter models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(3):425–464, 2001.
779
- 780 [48] Katalin Csilléry, Michael GB Blum, Oscar E Gaggiotti, and Olivier
781 François. Approximate bayesian computation (abc) in practice. *Trends*
782 *in ecology & evolution*, 25(7):410–418, 2010.
- 783 [49] Isabelle Guyon and André Elisseeff. An introduction to variable and
784 feature selection. *Journal of machine learning research*, 3(Mar):1157–
785 1182, 2003.
- 786 [50] Jie Cai, Jiawei Luo, Shulin Wang, and Sheng Yang. Feature selection
787 in machine learning: A new perspective. *Neurocomputing*, 300:70–79,
788 2018.
- 789 [51] David W Scott. *Multivariate density estimation: theory, practice, and*
790 *visualization*. John Wiley & Sons, 2015.
- 791 [52] Peter J Rousseeuw and Annick M Leroy. *Robust regression and outlier*
792 *detection*, volume 589. John wiley & sons, 2005.
- 793 [53] H Läuter. Silverman, bw: Density estimation for statistics and data
794 analysis. chapman & hall, london–new york 1986, 175 pp., £ 12.—. *Bio-*
795 *metrical Journal*, 30(7):876–877, 1988.