Research article

Environmental stress level to model tumor cell growth and survival

Sabrina Schönfeld[1,*], Alican Ozkan[2], Laura Scarabosio[3], Marissa Nichole Rylander[4] and Christina Kuttler[1]

[1] Center of Mathematics, Technical University of Munich, Garching, Germany

[2] Wyss Institute for Biologically Inspired Engineering, Harvard University, Boston, MA, United States

[3] Institute for Mathematics, Astrophysics and Particle Physics, Radboud University, Nijmegen, The Netherlands

[4] Department of Mechanical Engineering, The University of Texas, Austin, TX, United States

* **Correspondence:** Email: sabrina.schoenfeld@tum.de.

Abstract: Survival of living tumor cells underlies many influences such as nutrient saturation, oxygen level, drug concentrations or mechanical forces. Data-supported mathematical modeling can be a powerful tool to get a better understanding of cell behavior in different settings. However, under consideration of numerous environmental factors mathematical modeling can get challenging. We present an approach to model the separate influences of each environmental quantity on the cells in a collective manner by introducing the "environmental stress level". It is an immeasurable auxiliary variable, which quantifies to what extent viable cells would get in a stressed state, if exposed to certain conditions. A high stress level can inhibit cell growth, promote cell death and influence cell movement. As a proof of concept, we compare two systems of ordinary differential equations, which model tumor cell dynamics under various nutrient saturations respectively with and without considering an environmental stress level. Particle-based Bayesian inversion methods are used to quantify uncertainties and calibrate unknown model parameters with time resolved measurements of *in vitro* populations of liver cancer cells. The calibration results of both models are compared and the quality of fit is quantified. While predictions of both models show good agreement with the data, there is indication that the model considering the stress level yields a better fitting. The proposed modeling approach offers a flexible and extendable framework for considering systems with additional environmental factors affecting the cell dynamics.

Keywords: tumor growth; mathematical model; parameter estimation; uncertainty quantification; Bayesian inversion

1. Introduction

As cancerous diseases still cause many deaths in human population, there is a high need of improving the understanding of tumor growth dynamics and treatment strategies. For that purpose and for treatment optimization, mathematical approaches can help with adequate quantitative descriptions and predictions. Numerous mathematical models and methods are already available (see [1, 2]), including differential equations of all types, stochastic models, phenomenological models as well as more mechanistic ones. Despite good new experimental techniques, e.g. measuring tumor growth also *in vivo* [3–5], it is still difficult to quantify growth by determining parameter values in realistic situations without harming a patient or affecting a potential treatment. To get a better quantitative understanding of the underlying processes, investigating tumor cell cultures and even single cells under well-controlled conditions provides a promising way to get at least parts of the necessary information (see e.g. [6]). But even under laboratory conditions, parameters can underlie variations and cells may behave individually to some extent. To consider this, uncertainty quantification can be an adequate tool. Having such *in vitro* experiments, it is much easier to separate and measure different environmental influences on the tumor growth. The probably most relevant environmental influences are the availability of nutrients and oxygen. For modeling purposes, one often focuses on the limiting nutrient, without specifying it in detail. Nutrient deprivation causes "stress" to any cells, but other factors may cause stress to the cells, as well. Thus, it is worth to consider the concept of stress as a relevant factor for modeling growth and death of tumor cells.

To approach this problem, we suggest the introduction of an "environmental stress level" as an immeasurable auxiliary variable, which quantifies to what extent viable cells would get stressed out under certain environmental conditions. These may include quantities like nutrient saturation, oxygen level, drug concentrations or mechanical forces. On the one hand, this environmental stress level can e.g. inhibit cell growth and promote death by inducing necrosis for present cells. On the other hand, it is assumed to have an influence on cell movement, as cells favor areas with a low environmental stress level, which may cause metastasis formation, as well. In this work, this approach is presented in a simplified spatially homogeneous setting, where only the nutrient saturation effects the environmental stress level. Nevertheless, the presented model could be easily extended to more environmental variables, which provides several advantages from a modeling point of view. Even in this very simplified setting, model comparison shows a slight preference for the newly introduced model.

For comparison, we present two systems of ordinary differential equations, which model tumor cell population dynamics respectively with and without considering an environmental stress level. To do so, time-resolved *in vitro* data (from [7]) are used to perform model parameter calibration with Bayesian inversion. The latter approach provides a quantitative estimate of the uncertainty in the estimated parameters and hence brings more information than a deterministic inversion method, at the price of a higher computational cost. In the Bayesian inversion setting several methods are available for sampling. Among the most popular ones are Markov Chain Monte Carlo (MCMC) methods [8, Ch. 6-7], which are very simple to implement, but they need ad hoc tuning when the dimension of the parameter space is high or for complex posterior distributions [9]. In the last decade, particle based methods have gained popularity for their higher efficiency and robustness compared to simple MCMC, especially in the presence of high dimensional parameter spaces and multimodal posterior distributions [10]. Most of these algorithms are population based, which means they deal with a collection of samples in every

iteration. In this category are sequential importance sampling and resampling (annealed importance sampling [11] and Sequential Monte Carlo [12, 13]) and population MCMC [14, 15]. To calibrate the unknown parameters of the presented models, we use Sequential Monte Carlo (SMC). It is worth mentioning that the ensemble Kalman filter [16] can also be used to solve the inverse problem. It can be viewed as an approximate SMC with Gaussian approximation of the posterior [17]. However, the rigorous analysis of this method is confined for the moment to linear forward operators [17, 18]. Eventually, the calibration results are used to compare the models via a validation metric and the Bayes factor. For this purpose, the chosen SMC method is advantageous opposed to other algorithms (e.g. MCMC), since its structure provides easy access to the Bayes factor [10].

The article is structured as follows: The upcoming Section 2 explains the investigated mathematical models with the underlying experimental setup and provides background about the applied methods for uncertainty quantification, parameter calibration and model comparison. The results from the model calibrations are summarized and discussed in Section 3. Finally, Section 4 concludes the presented results and the use of the novel modeling approach. At the end of this article, more details about mathematical analysis of the models and the calibration results are enclosed in a supplementary section.

2. Materials and method

Section 2.1 presents different models for population dynamics of tumor cells and introduces the idea of using a stress level to model environmental influences on the cells collectively. To investigate the effect of nutrient changes on viable cells, we consider a special case of the models. The underlying biological setting and the measurement methods for the corresponding *in vitro* experiments are explained in Section 2.2. With these data we calibrate the unknown parameters of the models using Bayesian inversion and Sequential Monte Carlo methods. Section 2.3 provides the theoretical background behind the algorithms. Finally, we present the validation metric and the Bayes factor in Section 2.4, which are used to investigate and compare the resulting model calibrations.

2.1. Mathematical modeling

The presented ODE models describe the basic growth and death dynamics of tumor cells in a spatially homogeneous, avascular environment under consideration of the present nutrient saturation. The time-dependent variables are the density of viable tumor cells $V = V(t)$, the nutrient saturation $S = S(t)$, and the environmental stress level $\eta = \eta(t)$ at day $t \geq 0$.

We consider two possibilities to model the influence of the nutrient saturation on the cells: Model $\mathcal{M}_S$ assumes that the nutrient saturation directly affects the growth/death rates of the cells, while in model $\mathcal{M}_\eta$ the nutrient supply influences the cell population indirectly by changing the environmental stress level, which itself has an effect on the viability of the cells. The nutrient saturation is bounded and normalized: $S(t) \in [0, 1]$. The bounds represent the complete absence of nutrients ($S = 0$) and a nutrient supply, which creates optimal growth conditions for the cells ($S = 1$). In the latter case, both models can be reduced to a simplified model, which is independent of the nutrient saturation S. We denote this model by $\mathcal{M}_{\text{opt}}$ ("opt" short for "optimal" growth conditions).

A detailed description of each model follows in the consecutive paragraphs. All models are biologically reasonable in a first check, as they have been mathematically analyzed in terms of positivity and boundedness of the solutions as well as steady states and their stability. The last paragraph of this

section provides a summary of the variables, parameters, and mathematical features of the models.

What if the cells find optimal nutrient conditions?   External nutrients serve as an energy source for cell proliferation. Growth can happen with a maximal possible proliferation rate β, which is virtually reached under optimal nutrient conditions ($S = 1$). Under these circumstances, cell death is assumed to occur only for nutrient-independent reasons (e.g. dying of old age) with a constant rate λ. These dynamics can be modeled with a combination of a generalized logistic growth term (first proposed in [19]) and an exponential death term. The population growth is limited by the carrying capacity K of the biological system, taking into account e.g. limited space but not nutrient shortage. In particular, cell growth is inhibited in a cell density dependent manner, which is known as "proliferation contact inhibition" [20]. The strength of this phenomenon is specific for the cell type and is modeled by the shape parameter m of the logistic growth. A small value for m indicates strong contact inhibition, i.e. reduction of proliferation already starts at small cell densities. Therefore, $1/m$ can be interpreted as the strength of contact inhibition. Overall, this results in the following initial value problem:

$$
S = 1 \quad \Rightarrow \quad
\begin{cases}
\dot{V} = \beta\, V \left(1 - \left(\dfrac{V}{K}\right)^m\right) - \lambda\, V, \\[2mm]
V(0) = V_0.
\end{cases}
\qquad (\mathcal{M}_{\mathrm{opt}})
$$

As a feature of the pure logistic term, population growth for very small populations (i.e. $V \ll K$) can be approximated by exponential growth with rate β. It is a reasonable assumption that, independently of the initial population size V_0, the cell population's size actually increases over time under optimal nutrient conditions, which translates to a parameter relation: $\beta > \lambda$. With this constraint, the ODE can be rewritten as a purely logistic growth term with a new rate $\tilde{\beta}$ and capacity $\tilde{K}$:

$$
\dot{V} = \beta\, V \left(1 - \frac{\lambda}{\beta} - \left(\frac{V}{K}\right)^m\right) \overset{\beta > \lambda}{=} \beta\, V \left(1 - \frac{\lambda}{\beta}\right) \cdot \left(1 - \left(\frac{V}{K\left(1 - \frac{\lambda}{\beta}\right)^{1/m}}\right)^m\right) = \underbrace{(\beta - \lambda)}_{=\tilde{\beta}}\, V \left(1 - \left(\underbrace{\frac{V}{K\left(1 - \frac{\lambda}{\beta}\right)^{1/m}}}_{=\tilde{K}}\right)^m\right).
$$

Therefore, under optimal nutrient conditions, the cell population actually grows logistically with

$$
\text{"net" growth rate:} \quad \tilde{\beta} = (\beta - \lambda) \in (0, \beta)
$$

$$
\text{and "net" carrying capacity:} \quad \tilde{K} = K \left(1 - \frac{\lambda}{\beta}\right)^{1/m} = K \left(\frac{\beta - \lambda}{\beta}\right)^{1/m} \in (0, K).
$$

For $t \geq 0$ the known analytical solution of such an initial value problem, i.e. for model $\mathcal{M}_{\mathrm{opt}}$, is

$$
V(t) = \frac{V_0 \tilde{K}}{\left(V_0^m + \left(\tilde{K}^m - V_0^m\right) \cdot e^{-\tilde{\beta} m t}\right)^{1/m}} = \left(\frac{(V_0 K)^m (\beta - \lambda)}{\beta V_0^m + \left((\beta - \lambda)\, K^m - \beta V_0^m\right) \cdot e^{-(\beta - \lambda) m t}}\right)^{1/m}. \qquad (2.1)
$$

What if suboptimal nutrient conditions directly affect the cell dynamics?   Exposing the cells to suboptimal nutrient conditions ($0 \leq S < 1$) can affect the cells' metabolism by decelerating proliferation or even lead to cell death by starvation. To include this, an additional death term is needed, whose rate as well as the proliferation rate are scaled appropriately to the nutrient supply.

A time-dependent nutrient supply can be described by adding an additional differential equation to the system: $\dot{S} = g(S, V, t)$ with $S(0) = S_0$. The reaction term $g(S, V, t)$ can e.g. include terms for nutrient consumption by viable cells or an external nutrient source. In the considered *in vitro* experiments, whose measurements are used to calibrate the model later on, the nutrient concentration is maintained approximately constant. Therefore, all presented models assume

$$\dot{S} = g(S, V, t) = 0 \quad \Rightarrow \quad S(t) = \text{const.} = S_0 \in [0, 1] \quad \forall t \geq 0.$$

The present nutrient saturation S_0 contributes to the cells' reproductivity as well as to their ability to survive. If there are no nutrients available, the cells die from starvation with a maximal possible rate λ_{st}, assumed to happen in an exponential manner. With increasing nutrient supply, the starvation rate λ_{st} decreases and the proliferation rate β is up-regulated. Adapting the previous model $\mathcal{M}_{\text{opt}}$ accordingly yields the more general initial value problem

$$
\begin{cases}
\dot{V} = \underbrace{\delta^+(S_0) \cdot \beta}_{=\beta_S} V \left(1 - \left(\frac{V}{K}\right)^m\right) - \underbrace{(\lambda + \delta^-(S_0) \cdot \lambda_{\text{st}})}_{=\lambda_S} V, \\
V(0) = V_0,
\end{cases}
\tag{$\mathcal{M}_S$}
$$

with appropriate nutrient-dependent functions $\delta^+(S)$ and $\delta^-(S)$, scaling the total growth and death rates:

$$\beta_S = \delta^+(S_0) \cdot \beta \quad \text{and} \quad \lambda_S = (\lambda + \delta^-(S_0) \cdot \lambda_{\text{st}}).$$

The scaling functions $\delta^\pm(S)$ need to fulfill certain criteria to be biologically meaningful in our model setting. First, the cells should not die from starvation but proliferate with the maximal possible rate under optimal nutrient conditions. In contrast, without nutrients the cells should not be able to reproduce but starve with a maximal possible rate. These features translate to the properties

$$
\delta^+(S) \cdot \beta
\begin{cases}
\to \beta & \text{for } S \to 1, \\
\in (0, \beta) & \text{for } 0 < S < 1, \\
\to 0 & \text{for } S \to 0,
\end{cases}
\quad \text{and} \quad
\delta^-(S) \cdot \lambda_{\text{st}}
\begin{cases}
\to 0 & \text{for } S \to 1, \\
\in (0, \lambda_{\text{st}}) & \text{for } 0 < S < 1, \\
\to \lambda_{\text{st}} & \text{for } S \to 0.
\end{cases}
$$

Second, increasing the nutrient saturation should promote growth and reduce starvation: $\delta^+(S_0)$ resp. $\delta^-(S_0)$ need to be monotonically increasing resp. decreasing functions bounded by $[0, 1]$. Therefore, a possible choice are Hill type functions with Hill coefficient $k = 2$, i.e.

$$\delta^+ : S \mapsto \frac{S^2}{S_{\text{thr}}^2 + S^2} \quad \text{and} \quad \delta^- : S \mapsto 1 - \delta^+(S) = 1 - \frac{S^2}{S_{\text{thr}}^2 + S^2}, \tag{2.2}$$

where $0 < S_{\text{thr}} \ll 1$ denotes the nutrient threshold, for which the cells proliferate/starve with half-maximal rate. For simplicity, we do not assume any hysteresis effect, motivating the relation $\delta^-(S) = 1 - \delta^+(S)$. The scaling functions $\delta^\pm(S)$ can be interpreted as influence functions, which describe how tolerant viable cells are to nutrient changes. In this context, the parameter S_{thr} can be seen as a nutrient sensitivity threshold. Figure 1 shows a qualitative plot of the behavior of the scaling functions.
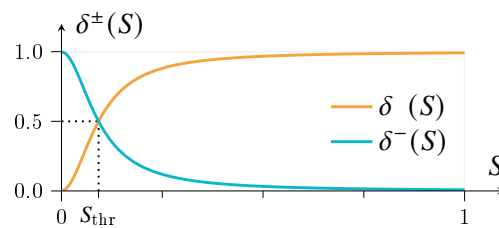
Figure 1. Qualitative behavior of the Hill type scaling functions $\delta^+(S)$ and $\delta^-(S)$

It can be shown (for calculations see Section 4.1 in the supplement) that for $t \geq 0$ a well-defined analytical solution of problem $(\mathcal{M}_S)$ is

$$V(t) = \begin{cases} V_0 K \cdot \left(\dfrac{\beta_S - \lambda_S}{\beta_S V_0^m + ((\beta_S - \lambda_S)K^m - \beta_S V_0^m)e^{-(\beta_S - \lambda_S)mt}} \right)^{1/m} & \text{if } \beta_S \neq \lambda_S \,, \\[4mm] V_0 K \cdot \left(\dfrac{1}{mt\beta_S V_0^m + K^m} \right)^{1/m} & \text{if } \beta_S = \lambda_S \,. \end{cases}$$

We note that in the extreme situation of having no nutrient supply (i.e. $S_0 = 0$), it holds $\delta^+(S_0) = 0$ and $\delta^-(S_0) \approx 1$. This results in $\beta_S = 0$ and $\lambda_S \approx \lambda + \lambda_{st} > 0$. By inserting this special case of $\beta_S \neq \lambda_S$ into the given analytical solution, it simplifies to

$$V(t) \overset{S_0 \to 0}{\longrightarrow} V_0 K \cdot \left(\frac{-(\lambda + \lambda_{st})}{-(\lambda + \lambda_{st})K^m e^{(\lambda + \lambda_{st})mt}} \right)^{1/m} = V_0 e^{-(\lambda + \lambda_{st})t} \,.$$

This is an expected result, being the solution of a simple exponential decay model without growth:

$$\begin{cases} \dot{V} = -(\lambda + \lambda_{st})V \,, \\ V(0) = V_0 \,. \end{cases}$$

Advantages of using the environmental stress level (ESL) instead to directly affect the cells. For an alternative way to include the effect of the nutrient conditions in model $\mathcal{M}_{opt}$, we introduce an auxiliary variable: the environmental stress level $\eta = \eta(t)$. It is designed to be an immeasurable quantity, describing to what extent viable cells would get stressed out, if exposed to certain environmental conditions, like in this case the nutrient saturation. We assume that the stress level is limited by a maximal level, which is set to one: $\eta \in [0, 1]$.

In a general setting, the the ESL can be influenced by multiple environmental factors (e.g nutrient/oxygen/drug concentration), where each can be mathematically represented by a system variable. Let $E_1(t), \ldots, E_n(t)$ be a set of such environmental variables, with

$$\begin{cases} \dot{E}_1 = g_1(E_1, \ldots, E_n, V, t) \,, & E_1(0) = E_{1,0} \,, \\ \quad \vdots \\ \dot{E}_n = g_n(E_1, \ldots, E_n, V, t) \,, & E_n(0) = E_{n,0} \,, \end{cases}$$

being the given reaction equations. Then the ODE for the environmental stress level can be given by

$$\dot{\eta} = \underbrace{\left(\sum_{j=1}^n \alpha_j^- \cdot \delta_j^-(E_j) \right)(1 - \eta)}_{\substack{\text{increasing stress level} \\ \text{(stressful conditions)}}} - \underbrace{\left(\sum_{j=1}^n \alpha_j^+ \cdot \delta_j^+(E_j) \right)\eta}_{\substack{\text{recovery from stress} \\ \text{(beneficial conditions)}}} \qquad \text{with } \eta(0) = \eta_0 \in [0, 1] \,. \qquad (2.3)$$

The first term includes the dynamics raising the ESL. Since $\eta \in [0, 1]$, the increase is bounded by one from above via the factor $(1 - \eta)$. If the environmental factors generate beneficial growth conditions for the cells, the ESL decreases according to the second term. The parameters α_j^- resp. α_j^+ are variable-specific "sensitivity rates", denoting how fast the stress level increases/decreases, if the value of the associated variable E_j enters/leaves a range, which is critical for survival. With distinguishing between α^- and α^+ per variable, it is possible to mathematically consider that cells react e.g. faster to the lack of an environmental factor, which is essential for their survival, than they recover from the deprivation once beneficial conditions are restored. Which values of E_j are considered to be critical, is given by the explicit shape of the corresponding "influence functions" $\delta_j^-(E_j)$ and $\delta_j^+(E_j)$. In particular, large values of $\delta_j^-(E_j)$ promote stress, whereas large values of $\delta_j^+(E_j)$ inhibits stress.

Collecting the influences of the environment in the stress level gives rise to some practical advantages. We can choose an arbitrary number of environmental variables and include them easily into the stress reaction due to the modular structure of (2.3). With the parameters α_j^- and α_j^+ we can consider different sensitivities of the cells to changes in different environmental factors. Although neither these parameters nor the stress level η can be measured directly in experiments, appropriate data of cell viability can provide enough information to estimate the values of the sensitivity rates. For this e.g. statistical methods for parameter calibration can be used, where the only prior assumption is a rough range, in which the parameter values are expected. Then, by grouping these sensitivity rates according to their magnitude, we can distinguish the corresponding environmental variables between having a fast, slow or even no impact on the cells. Under a quasi-steady state assumption, the ODE for η can then be simplified by choosing a time scale of interest and omitting environmental variables with no influence.

In the considered experimental setting, we have the special case of the constant nutrient saturation $S(t)$ being the only environmental variable:

$$E_1(t) = S(t) = S_0 \quad \text{with} \quad g_1(E_1, V, t) = g(S, V, t) = 0 \,.$$

We assume the ESL to increase and decrease with the same "nutrient sensitivity rate" α_S. The corresponding "nutrient influence functions" are then $\delta^-(S)$ and $\delta^+(S)$, as introduced previously in (2.2) for model $\mathcal{M}_S$. Hence, we have

$$\alpha_1^- = \alpha_1^+ = \alpha_S \,, \quad \delta_1^-(E_1) = \delta^-(S_0) \quad \text{and} \quad \delta_1^+(E_1) = \delta^+(S_0) = 1 - \delta^-(S_0) \,,$$

i.e. given the initial condition $\eta(0) = \eta_0 \in [0, 1]$ we consider the ODE

$$\dot{\eta} = \alpha_S \delta^-(S_0) \cdot (1 - \eta) - \alpha_S (1 - \delta^-(S_0)) \cdot \eta = \alpha_S (\delta^-(S_0) - \eta) \,.$$

Since this equation is independent of the variable V, the exact solution can be calculated by separation of variables and for $t \geq 0$ it is given as:

$$\eta(t) = \delta^-(S_0) \cdot (1 - e^{-\alpha_S t}) + \eta_0 e^{-\alpha_S t} \,. \tag{2.4}$$

Instead of the nutrient-dependent scaling functions $\delta^\pm$ in the basic model $\mathcal{M}_S$, now the ESL influences the proliferation and starvation rate. A high stress level causes slower proliferation and faster starvation, whereas a low stress level has the opposite effect. Since by definition $\eta(t) \in [0, 1]$, this results in

the following initial value problem:

$$
\begin{cases}
\dot{V} = \underbrace{(1 - \eta(t)) \cdot \beta}_{=\beta_\eta(t)} V \left(1 - \left(\frac{V}{K}\right)^m\right) - \underbrace{(\lambda + \eta(t) \cdot \lambda_{\text{st}})}_{=\lambda_\eta(t)} V, \\
\dot{\eta} = \alpha_S \delta^-(S_0) \cdot (1 - \eta) - \alpha_S (1 - \delta^-(S_0)) \cdot \eta, \\
V(0) = V_0, \ \eta(0) = \eta_0.
\end{cases}
\qquad (\mathcal{M}_\eta)
$$

By inserting the explicit time-dependent analytical solution (2.4), the system reduces to a non-autonomous ODE for V, which is solved numerically. We observe that model $\mathcal{M}_\eta$ has an interesting relation to the model $\mathcal{M}_S$: under the assumption that critical nutrient changes immediately influence the viable cells, i.e. assuming $\alpha_S \to \infty$, we return to model $\mathcal{M}_S$, since for $t \geq 0$ it holds

$$
\eta(t) \stackrel{(2.4)}{=} \delta^-(S_0) \cdot (1 - e^{-\alpha_S t}) + \eta_0 e^{-\alpha_S t} \stackrel{\alpha_S \to \infty}{\longrightarrow} \delta^-(S_0) = \text{const.} \quad \Rightarrow \quad \beta_\eta(t) \stackrel{\alpha_S \to \infty}{\longrightarrow} \beta_S, \ \lambda_\eta(t) \stackrel{\alpha_S \to \infty}{\longrightarrow} \lambda_S.
$$

The same result can be achieved by a quasi-steady state assumption, stating that changes in the ESL happen at a much faster time scale than changes in the tumor cell number. In this case, the stress level reaches its steady state virtually instantly, which is: $\dot{\eta} = \alpha_S(\delta^-(S_0) - \eta) = 0 \Rightarrow \eta = \delta^-(S_0)$.

We are aware that having multiple environmental factors included into the model would make the advantages of using the concept of the environmental stress level more evident. However, due to the lack of corresponding data, we focus on the special case of having only one environmental variable (nutrient saturation S). We consider the investigations in this manuscript as a proof of principle that the environmental stress level is a feasible alternative way to model the effect of the environment on the cells.

Overview over all models and their mathematical properties. The last model $\mathcal{M}_\eta$ has been constructed from model $\mathcal{M}_S$, which itself uses model $\mathcal{M}_{\text{opt}}$ as a basis. Therefore, the number of variables and parameters increases with adding more complex dynamics. An overview over the variables and strictly positive model parameters of each system as well as abbreviating notations for important parameter terms/functions can be found in Tables 1, 2 and 3 below. The units of the variables and parameters are motivated by the biological setting and the measurement methods, which are explained in more detail in the following Section 2.2.

Table 1. Overview over the model variables and their occurrence in the models.

$\mathcal{M}_{\text{opt}}$	$\mathcal{M}_S$	$\mathcal{M}_\eta$	Variable	Meaning	Unit
✓	✓	✓	$V = V(t)$	Density of viable tumor cells	10^5 cells/ml
	✓	✓	$S = S(t)$	Nutrient saturation	10% FBS
		✓	$\eta = \eta(t)$	Environmental stress level	–

Table 2. Model parameters, their occurrence in the models and their meaning.

$\mathcal{M}_{\text{opt}}$	$\mathcal{M}_S$	$\mathcal{M}_\eta$	Parameter	Meaning	Unit
✓	✓	✓	β	Maximal possible proliferation rate	1/day
✓	✓	✓	K	Carrying capacity of the biological system	10^5 cells/ml
✓	✓	✓	$1/m$	Strength of proliferation contact inhibition	–
✓	✓	✓	λ	Natural death rate	1/day
	✓	✓	λ_{st}	Maximal possible starvation rate	1/day
	✓	✓	S_{thr}	Nutrient sensitivity threshold	10% FBS
		✓	α_S	Sensitivity rate of nutrient changes on stress level	1/day

Table 3. Abbreviating notations for important/useful parameter terms and functions as well as their occurrence in the model equations (none of them occurs in model $\mathcal{M}_{\text{opt}}$).

$\mathcal{M}_S$	$\mathcal{M}_\eta$	Notation		Meaning
✓	✓	$\delta^+(S_0) =$	$\frac{S_0^2}{S_{\text{thr}}^2 + S_0^2}$	Influence function for nutrient-promoted dynamics
✓	✓	$\delta^-(S_0) =$	$1 - \delta^+(S_0)$	Influence function for nutrient-inhibited dynamics
✓		$\beta_S =$	$\delta^+(S_0) \cdot \beta$	Nutrient-dependent net proliferation rate (constant)
✓		$\lambda_S =$	$\lambda + \delta^-(S_0) \cdot \lambda_{\text{st}}$	Nutrient-dependent net death rate (constant)
	✓	$\beta_\eta(t) =$	$(1 - \eta(t)) \cdot \beta$	Stress-dependent net proliferation rate (varying in time)
	✓	$\lambda_\eta(t) =$	$\lambda + \eta(t) \cdot \lambda_{\text{st}}$	Stress-dependent net death rate (varying in time)

A mathematical analysis of the models yields positivity and boundedness of the solutions, which are important features for biological reasonableness. Table 4 summarizes the computed bounds as well as the steady states $\bar{V}$ reps. $(\bar{V}, \bar{\eta})^T$ and their stability. The corresponding calculations providing these results can be found in the supplementary Section 4.2.

Table 4. Bounds of the solutions $V(t)$, $\eta(t)$ and steady states $\bar{V}$, $(\bar{V}, \bar{\eta})^T$ of each ODE model.

Model	Positivity & boundedness of the solutions	Steady states stable	unstable
$\mathcal{M}_{\text{opt}}$	$0 \le V(t) \le \max\left\{V_0, K\sqrt[m]{1 - \frac{\lambda_S}{\beta_S}}\right\}$	$K\sqrt[m]{1 - \frac{\lambda_S}{\beta_S}}$	0
$\mathcal{M}_S$ for $\lambda_S < \beta_S$:	$0 \le V(t) \le \max\left\{V_0, K\sqrt[m]{1 - \frac{\lambda_S}{\beta_S}}\right\}$	$K\sqrt[m]{1 - \frac{\lambda_S}{\beta_S}}$	0
for $\lambda_S \ge \beta_S$:	$0 \le V(t) \le V_0$	0	
$\mathcal{M}_\eta$ for all λ_S, β_S:	$0 \le \eta(t) \le \max\{\eta_0, \delta^-(S_0)\} \le 1$		
for $\lambda_S < \beta_S$:	$0 \le V(t) \le \max\left\{V_0, K\sqrt[m]{1 - \frac{\lambda_S}{\beta_S}}\right\}$	$\left(K\sqrt[m]{1 - \frac{\lambda_S}{\beta_S}}, \delta^-(S_0)\right)^T$	$(0, \delta^-(S_0))^T$
for $\lambda_S \ge \beta_S$:	$0 \le V(t) \le V_0$	$(0, \delta^-(S_0))^T$	

2.2. Experimental data

We use data from the experiments described in [7] to calibrate the unknown parameters from Table 2. A CellTiter-Blue® assay was used to monitor the viability of tumor cells. In particular, viable cells metabolize a provided chemical and they emit measurable light as a result of this process. Hence, the data points are fluorescence intensity measurements. This way, viability was measured once every day and there are four biological replicates of each measurement to estimate statistical significance and repeatability.

Let I^{total} be an intensity measurement of a specific cell line at time point $\tilde{t}$. Excluding the corresponding background intensity I^{BG} of the cell-free medium, the fluorescence intensity produced by viable cells I^V is assumed to be directly proportional to the density of viable tumor cells $V(\tilde{t})$. The experiments were performed with different initial cell densities between 10^4 and 10^5 cells per milliliter, which motivates the unit of V and eventually leads to the relation

$$I^V = I^{\text{total}} - I^{\text{BG}} \propto V(\tilde{t}) \, \frac{10^5 \, \text{cells}}{\text{ml}} \qquad \Rightarrow \qquad n = \frac{I^V}{V(\tilde{t})} \in (0,1) \,,$$

where n denotes the proportionality constant translating fluorescence intensity to cell density. Whenever we refer to "intensity measurements" in the following, we mean the fluorescence produced by the cells I^V and neglect the superscript "V" for better readability.

For nutrition the cells are supplemented with a particular concentration between 0% to 10% of fetal bovine serum (FBS). A supplementation with 0% FBS does not provide the cells with any nutrients, whereas 10% FBS generates optimal growth conditions. The nutrient supply is kept constant throughout the whole duration of each experiment, i.e. the nutrient saturation is assumed to stay at its initial level $S_0 \in [0,1]$ at any time. Overall, this motivates:

$$S(t) = S(0) = S_0 \; \forall \, t \geq 0 \quad \text{with unit} \quad [S] = 10\% \, \text{FBS}.$$

The remaining variable, the environmental stress level $\eta = \eta(t)$, is an auxiliary variable and especially immeasurable. Hence, it has no experimental counterpart and is used as a dimensionless quantity.

How are uncertainties considered?   In reality, the equation $n = I/V(\tilde{t})$ is not rigorously fulfilled. This can be due to e.g. model inadequacy or biological fluctuations of the cells' metabolism, which affect measurement accuracy. To capture this uncertainty, we assume a multiplicative noise for each element of a set of $M \in \mathbb{N}$ measurements $\{I_i\}_{i=1}^{M}$ and a set of model solutions $\{V_i\}_{i=1}^{M}$ considering the corresponding values of t, V_0, and S_0. In the Bayesian framework that we adopt (see Section 2.3) we model this to be a random variable ε_i such that for $i = 1, \ldots, M$:

$$I_i = n \, V_i \cdot \varepsilon_i \,. \tag{2.5}$$

Let ε_i be i.i.d. and have the unimodal and continuous distribution of a random variable ε with probability density function (PDF) f_ε. Then, the following properties should hold:

$$[\text{P1}]: \quad \text{supp}(f_\varepsilon) \subseteq \mathbb{R}^+ \,, \text{i.e. all measurements are positive;}$$
$$[\text{P2}]: \quad \mathbb{E}[\varepsilon] = 1 \,, \text{i.e. measurements are accurate on average;}$$
$$[\text{P3}]: \quad f_\varepsilon(x) \xrightarrow{x \to \infty} 0 \text{ and } f_\varepsilon(x) \xrightarrow{x \to 0} 0 \,, \text{i.e. outliers are possible but not likely.}$$

Different distributions are possible to accomplish these properties. A small number of shape parameters and an easy calculation to ensure property [P2] are desirable. Therefore, we choose a Gamma distribution $\varepsilon \sim \Gamma(a, b)$, $a, b > 0$ with a few restrictions. This is a plausible choice for multiplicative noise and often used in imaging theory (see e.g. [21–23]). Property [P1] and the desired behavior near infinity of [P3] are satisfied by definition. The corresponding PDF is given by $f_\varepsilon^{a,b}(x) = \frac{b^a}{\Gamma(a)} x^{a-1} e^{-bx}$, where $\Gamma(\cdot)$ denotes the Gamma function. To fulfill the remaining properties, we observe

$$\mathbb{E}[\varepsilon] = \frac{a}{b} \overset{[P2]}{=} 1 \text{, if } b = a \qquad \Rightarrow \qquad \lim_{x \to 0} f_\varepsilon^{a,a}(x) = \lim_{x \to 0} \left(\frac{a^a}{\Gamma(a)} x^{a-1} e^{-ax} \right) \overset{[P3]}{=} 0 \text{, if } a > 1 \,.$$

Note that by constraining $b = a > 1$ the shape of the distribution depends only on the parameter a. In fact, a is directly related to the standard deviation σ and hence the variance of the distribution: $\sigma^2 = \mathrm{Var}(\varepsilon) = \frac{a}{a^2} = \frac{1}{a} \Rightarrow a = \frac{1}{\sigma^2}$. Therefore, for $i = 1, \ldots, M$ the uncertainty factor ε_i for a particular measurement I_i can be modeled by

$$\varepsilon_i = \frac{I_i}{n \cdot V_i} \sim \Gamma\left(\frac{1}{\sigma^2}, \frac{1}{\sigma^2} \right) \text{ with } \sigma^2 \in (0, 1) \,. \tag{2.6}$$

We use the percentiles $P_{5\%}$ and $P_{95\%}$ of the Gamma distributed uncertainty factors ε_i to define the

$$\text{"90\% (uncertainty) range" around the solution } V : \quad [V \cdot P_{5\%}, \, V \cdot P_{95\%}] \,. \tag{2.7}$$

In particular, given a specific noise variance σ^2, the model expects 90% of the measurements within this interval, whereas respectively 5% are expected below and above it. The left side of Figure 2 depicts exemplary plots of the PDF f_ε for different values of σ^2. It also shows the positive skewness of the distribution. Under consideration of the measurement method, this is a reasonable feature for the uncertainty factors, assuming the cells might not metabolize the assay to their full potential. On the right side of Figure 2 we see an example of a 90% uncertainty range around a solution.
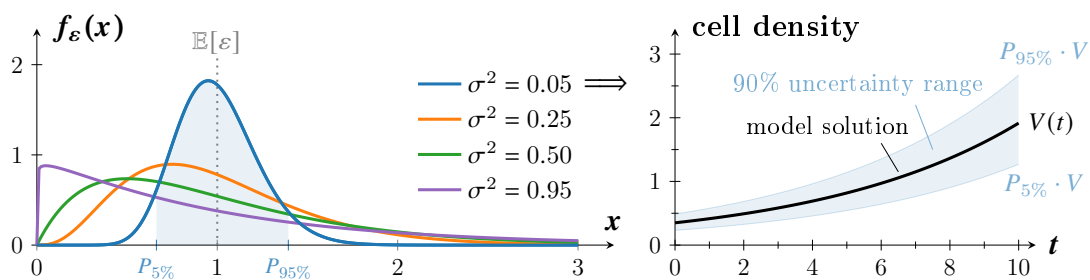

Figure 2. Left: Probability density function f_ε of $\varepsilon \sim \Gamma\left(1/\sigma^2, 1/\sigma^2\right)$ for varying variances σ^2 and the percentiles $P_{5\%}$ and $P_{95\%}$ for $\sigma^2 = 0.05$. Right: Resulting 90% uncertainty range (shaded area) around an exemplary model solution $V(t)$

The reason for considering a multiplicative (and, for instance, not additive) noise term is twofold. From a mathematical perspective, it allows to preserve positivity of the data. A more practical motivation is the reasonable assumption that the fluorescence noise of the intensity measurements is proportional to the density of viable cells. This is also supported by the observation of a larger variance for experiments with larger cell numbers in our data. Furthermore, multiplicative noise has demonstrated to be better suited for fluorescence than additive noise in other experimental settings [24].

How can the measurements be utilized? One of the experiments in [7] monitors the dependence between cell viability and nutrient supply by providing five different nutrient concentrations S_0 over a period of 7 days: this results in five data sets "D1"–"D5". These measurements are employed to calibrate the unknown model parameters for models $\mathcal{M}_S$ and $\mathcal{M}_\eta$ with Bayesian inversion methods (for details see following Section 2.3). Another experiment in [7] was designed to investigate cell behavior under optimal growth conditions, i.e. the cells were provided with 10% FBS over a period of 21 days. We denote the data set generated by this experiment with "D6". It is used additionally to D1–D5 to validate the calibration results with model $\mathcal{M}_{\text{opt}}$. All experiments start with several populations of different initial size V_0. Table 5 shows the experimental setup of all data sets and the resulting initial conditions in the models.

Table 5. Overview over the data sets and the corresponding initial values. D1–D5 are used to calibrate the models $\mathcal{M}_S$ and $\mathcal{M}_\eta$ and D6 for validation of the results with model $\mathcal{M}_{\text{opt}}$.

Data set	Nutrition (in % FBS)	Duration (in days)	Initial values in the models V_0	S_0	η_0
D1	10.0	7	1.00, 0.50, 0.25	1.00	0.00
D2	7.5	7	1.00, 0.50, 0.25	0.75	0.00
D3	5.0	7	1.00, 0.50, 0.25	0.50	0.00
D4	2.5	7	1.00, 0.50, 0.25	0.25	0.00
D5	0.0	7	1.00, 0.50, 0.25	0.00	0.00
D6	10.0	21	1.00, 0.50, 0.25, 0.10, 0.05	(1.00)	(0.00)

2.3. Parameter calibration: Bayesian inversion and Sequential Monte Carlo (SMC)

The task to identify the unknown true parameters from given data is called the "inverse problem". We solve it using a Bayesian approach, which leads, under mild assumptions, to a naturally well-posed inverse problem [25]. Furthermore, this approach allows to quantify the uncertainty in the estimated parameters and hence it brings more information than a deterministic inversion method, at the price of a higher computational cost.

We collect all parameters to be estimated in a vector $\theta \in \Theta \subseteq \mathbb{R}^d$ ($d \in \mathbb{N}$), where Θ denotes the parameter space. Assuming the considered data consist of M intensity measurements (excluding background intensity), we collect them in a vector $\mathcal{I} = (I_1, \ldots, I_M)^T = (I_i)_{i=1}^M \in \mathbb{R}^M$. Defining $\mathcal{G} : \Theta \to \mathbb{R}^M$ with $\theta \mapsto (\mathcal{G}_i(\theta))_{i=1}^M$ as the forward operator mapping parameter values to the corresponding intensities, such a measurement can be rewritten as

$$I_i \overset{(2.5)}{=} n\, V_i \cdot \varepsilon_i = \mathcal{G}_i(\theta) \cdot \varepsilon_i, \tag{2.8}$$

where V_i is the corresponding model solution to $(\mathcal{M}_{\text{opt}})$, $(\mathcal{M}_S)$ or $(\mathcal{M}_\eta)$ using θ as parameters and ε_i is the multiplicative noise. Note that we use parts of the data sets D1–D5 for parameter estimation, hence each measurement I_i can refer to a different nutrient condition, initial cell density, and time point. Therefore, V_i has to be calculated in consideration of the corresponding values for S_0, V_0 and t.

Now, we want to consider all measurements in $\mathcal{I} = (I_i)_{i=1}^M$ collectively. The parameter vector θ and the noise $(\varepsilon_i)_{i=1}^M$ are modeled as multi-dimensional random variables taking values in Θ and $\mathbb{R}^M$ respectively. The Bayesian formulation of the problem is the following: Given a prior (measure) μ_0

on Θ, compute the posterior (measure) μ^I given the data I. The prior in the Bayesian setting is the correspondent of a regularization in deterministic inverse problems [25] and it reflects the knowledge about the parameters before including any information given by the data, whereas the posterior describes the knowledge after seeing the data. Let π_0 and π^I denote the probability densities of μ_0 and μ^I, respectively. By Bayes' formula, we have

$$\pi^I(\theta) = \frac{L(I \mid \theta) \cdot \pi_0(\theta)}{\int_\Theta L(I \mid \theta) \cdot \pi_0(\theta)\, d\theta} \propto L(I \mid \theta)\, \pi_0(\theta)\,, \tag{2.9}$$

where L is the data likelihood [25]. The proportionality constant of relation (2.9) depends only on I. It is called "model evidence" and it can be used to quantitatively compare two models (see Section 2.4). We remind from the previous Section 2.2 that (2.6) states i.i.d. $\varepsilon_i \sim \Gamma\left(\frac{1}{\sigma^2}, \frac{1}{\sigma^2}\right)$ with $\sigma^2 \in (0,1)$ for every measurement I_i, $i = 1, \dots, M$. Using this together with (2.8), the data likelihood of $I = (I_i)_{i=1}^M$ is:

$$L(I \mid \theta) \propto \prod_{i=1}^M \left(\frac{I_i}{G_i(\theta)}\right)^{1/\sigma^2 - 1} \exp\left(-\frac{1}{\sigma^2}\frac{I_i}{G_i(\theta)}\right).$$

How to sample from the posterior? In order to make predictions, we want to sample from the given posterior distribution (2.9). However, this has a complicated, concentrated density, so we cannot sample from it exactly with a random number generator. To approximate the posterior measure, we use therefore the Sequential Monte Carlo (SMC) method, which we explain now based on [26]. In SMC one considers a sequence of intermediate distributions $(\mu_k)_{k=0}^N$, such that μ_0 is the prior and $\mu_N = \mu^I$ coincides with the posterior distribution. The probability density π_k of the intermediate measure μ_k can be defined by

$$\pi_k(\theta) = \frac{1}{Z_k}\prod_{i=1}^k L_i(I_i \mid \theta)\, \pi_0(\theta) \quad \text{or, equivalently,} \quad \pi_k(\theta) = \frac{1}{\tilde{Z}_k} L_k(I_k \mid \theta)\, \pi_{k-1}(\theta)\,, \tag{2.10}$$

where Z_k and $\tilde{Z}_k$ are normalizing constants and

$$L_i(I_i \mid \theta) = \frac{1}{Z_i}\left(\frac{I_i}{G_i(\theta)}\right)^{1/\sigma^2 - 1} \exp\left(-\frac{1}{\sigma^2}\frac{I_i}{G_i(\theta)}\right)$$

is the likelihood associated to observation I_i with normalization constant Z_i, $i = 1, \dots, M$. The intermediate densities could also be constructed with an adaptive approach using tempering [9]. However, since the considered data measures the quantity of interest in a time series, the presented filtering method is computationally more efficient. The SMC algorithm samples sequentially from the intermediate measures μ_k using a weighted swarm of samples, called particles. Let $P \in \mathbb{N}$ be the sample size, i.e. the number of particles. At the k-th iteration ($k = 1, \dots, N$) the algorithm leads to a collection of particles $\{\theta_p\}_{p=1}^P$ with associated weights $\{W_p^k\}_{p=1}^P$, which gives the approximation

$$\pi_k(\theta) \approx \sum_{p=1}^P W_p^k \delta_{\theta_p}(\theta)\,.$$

The SMC algorithm is summarized in Algorithm 1: we achieve appropriately weighted particles to approximate $\mu_N = \mu^I$ by starting with uniformly weighted particles distributed according to the prior μ_0 (line 1) and iteratively move the samples from the previous measure μ_{k-1} to μ_k in a selection (lines 3-7) and a mutation step (line 8) [27, Ch. 5], which are explained in detail in the following paragraphs.

Algorithm 1 Sequential Monte Carlo

1:   $k = 0$: sample $\theta_p \sim \mu_0$ and set $W_p^0 = \frac{1}{P}$, $p = 1, \ldots, P$

2:   **for** $k = 1, \ldots, N$ **do**

3:      $w_p^k = L_k(I_k \,|\, \theta)\, W_p^{k-1}$, $W_p^k = w_p^k / (\sum_{p=1}^{P} w_p^k)$, $p = 1, \ldots, P$ ▷ selection step

4:      **if** $P_{\text{eff}} < P_{\text{thr}}$ **then** resample:

5:          (i): sample indices acc. to distribution $\mathcal{R}$ of particle indices: $(\pi_1, \ldots, \pi_P) \sim \mathcal{R}(W_1^k, \ldots, W_P^k)$

6:          (ii): set $\theta_p = \theta_{\pi_p}$ and $W_p^k = \frac{1}{P}$, $p = 1, \ldots, P$

7:      **end if**

8:      move $\theta_p \sim \kappa_k(\theta_p, \cdot)$, $p = 1, \ldots, P$ ▷ μ_k-invariant mutation step

9:   **end for**

Selection step. We start with a collection of particles $\{\theta_p\}_{p=1}^{P}$, distributed according to μ_{k-1}. Their weights $\{W_p^{k-1}\}_{p=1}^{P}$ are updated to $\{W_p^k\}_{p=1}^{P}$ by importance sampling: for $p = 1, \ldots, P$ we have

$$W_p^k = \frac{w_p^k}{\sum_{p=1}^{P} w_p^k} \quad \text{with} \quad w_p^k \overset{(2.10)}{=} L_k(I_k \,|\, \theta)\, W_p^{k-1}.$$

We see that W_p^k is normalized to ensure that $\sum_{p=1}^{P} W_p^k = 1$, i.e. having a probability distribution. Note that the importance sampling only changes the weights and not the particles. However, if there are many particles with low weights, the estimation is only as accurate as a Monte Carlo approximation with a very small number of particles [28]. In this case, the reweighing step is followed by a resampling step, where the particles are replaced according to their updated weights. Resampling is needed, if the effective sample size

$$P_{\text{eff}} \approx \left(\sum_{p=1}^{P} \left(W_p^k \right)^2 \right)^{-1}$$

is small, which we check by comparison with a threshold $P_{\text{thr}} = \tau P$, $\tau \in (0, 1)$:

$$P_{\text{eff}} \begin{cases} < P_{\text{thr}} & \Rightarrow \text{ resample } \{\theta_p\}_{p=1}^{P} \text{ according to } \{W_p^k\}_{p=1}^{P} \text{ and uniformly weigh new particles}, \\ \geq P_{\text{thr}} & \Rightarrow \text{ do not resample}. \end{cases}$$

This discards particles with low weight and improves the representation of the distribution μ_k.

Mutation step. Performing only selection steps will eventually lead to degeneracy in the diversity of the particle population. In particular, after some resampling steps, few particles will survive and be replicated. Therefore, we introduce diversity in the particles by moving them according to a Markov Chain Monte Carlo (MCMC) kernel $\kappa_k(\cdot, \cdot)$. This kernel is μ_k-invariant, i.e. it does not modify the particle distribution. We adopt the adaptive strategy developed in [9] to construct such a MCMC kernel. A random walk Metropolis-Hastings (MH) proposal is used on each univariate component,

conditionally independently. More precisely, remembering that each particle is a vector $\theta_p \in \mathbb{R}^d$, MH proposes $\{q_p\}_{p=1}^P$ with $q_p \in \mathbb{R}^d$ by computing

$$(q_p)_j = (\theta_p)_j + \epsilon_j^k \cdot \xi_j \quad \text{with} \quad \xi_j \sim \mathcal{N}(0, 1),$$

where $(\cdot)_j$ denotes the j-th component ($j = 1, \ldots, d$) and the scale ϵ_j^k is tuned to the acceptance rate of the previous SMC iteration. This means, we choose ϵ_j^k adaptively as

$$\epsilon_j^k = \rho_k \sqrt{\widehat{\mathrm{Var}}\big((\theta)_j\big)} \quad \text{with recursively defined scaling parameter} \quad \rho_k = \begin{cases} \rho_{k-1} \cdot 2 & \text{if } a_{k-1} > 0.30, \\ \rho_{k-1} / 2 & \text{if } a_{k-1} < 0.15, \\ \rho_{k-1} & \text{otherwise}, \end{cases}$$

where a_{k-1} is the average acceptance rate over the particles at the previous iteration and $\widehat{\mathrm{Var}}\big((\theta)_j\big)$ denotes the empirical marginal variance from the j-th components of all particles. It is also possible to adapt the proposal using empirical covariances instead [29], which would be computationally more expensive in view of possible extensions to parameters, which are random fields and therefore very high dimensional. Eventually, the particles are moved by randomly accepting the proposed ones:

$$\text{set } \theta_p = q_p \quad \text{with probability} \quad \min\left\{\frac{\pi_k(q_p)}{\pi_k(\theta_p)}, 1\right\}.$$

To improve the mixing, it is possible to repeat this process more than once by applying κ_k again on the moved particles. Afterwards, the total ratio of accepted particles determines the a_k, which is used to scale ϵ_j^{k+1} in the MCMC update of the next iteration.

How is the theory applied to the models? Parameter estimation is performed in Python adapting the code provided in [28]. The set of parameters, which need to be calibrated, can be distinguished between model parameters (see Table 2) and hyper parameters. Latter includes the unknown uncertainty variance σ^2 and the proportionality constant n from relation (2.6). All calibrations are performed using the SMC method with sample size $P = 50\,000$ using the resampling threshold $P_{\text{thr}} = 75\% \, P$. In the mutation step, five MCMC updates are performed, i.e. line 8 in Algorithm 1 is performed five times.

The parameters are calibrated over the course of $N = 8 \cdot 3 = 24$ SMC steps: As depicted on the left side of Figure 3, starting with the data corresponding to $V_0 = 1.00$ and $t = 0$, the SMC steps iterate over the eight points in time (inner loop) and over the three separate seeding densities (outer loop). With each step k, another set of $5 \cdot 4 = 20$ data points, containing all measurements of D1–D5 at a specific time point t regarding a particular initial cell density V_0 (see right side of Figure 3), is included into the considered data collection $\mathcal{I}_k$ for the selection step (line 3 in Algorithm 1). This maximizes the information about the effect of varying nutrients on the cells at each calibration step. In particular, in the first step a set of measurements $\mathcal{I}_1 = \{\mathcal{I}_{1,i}\}_{i=1}^{M_1}$ with $M_1 = 5 \cdot 4 = 20$ is considered and with each SMC step 20 more data points are included incrementally: $\mathcal{I}_1 \subset \mathcal{I}_2 \subset \ldots \subset \mathcal{I}_N$ with $|\mathcal{I}_k| = M_k = k \cdot 20$, $k = 1, \ldots, N$. In total, the parameters are calibrated using $M_N = 24 \cdot 20 = 480$ measurements.
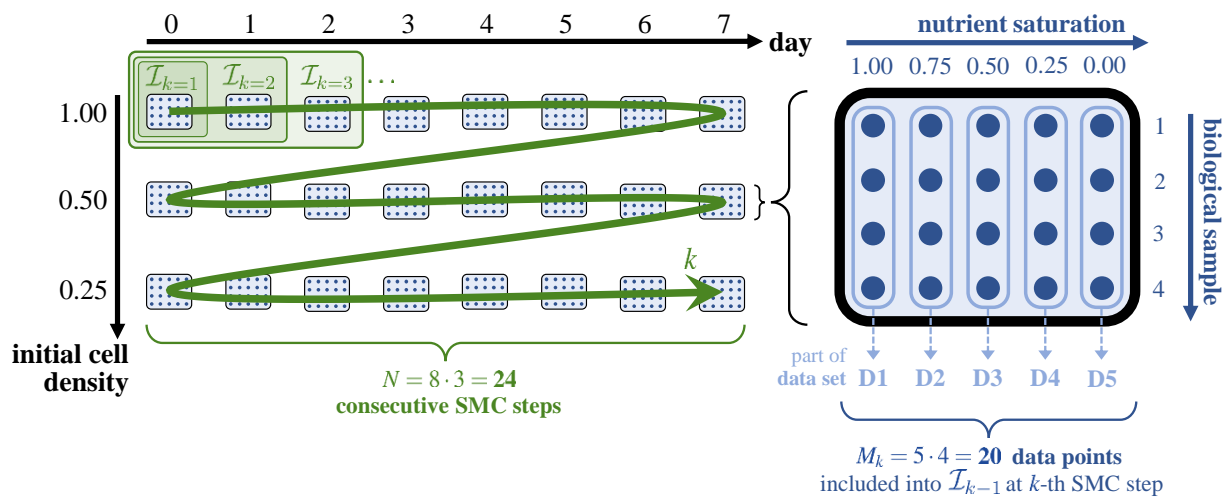
Figure 3. Schematic description of how the data sets D1–D5 are used to calibrate the parameters step by step with SMC

We want to construct informative prior distributions for the parameters. Information about the parameters' magnitude can be described by using an appropriate uniform distribution $\mathcal{U}(a,b)$ on an interval $[a,b]$. If we can additionally assume that the neighborhood of a parameter value h has a high probability, a triangular distribution $\text{Triang}(a,h,b)$ on $[a,b]$ is used, where $h \in [a,b]$ is the mode of the distribution, i.e. the value that is most likely to be sampled. Note that the mode can also lie on the interval bounds, if low/large values within the interval are assumed to have high probability. We set $h = (a+b)/2$ to the center of the interval, if we expect the borders a, b to have a low probability but there is no particular tendency to a value within the interval.

The doubling time T of the used cell type is larger than one day while in exponential growth [30,31], i.e. $T > 1$. We can estimate an upper bound for the growth rate β:

$$2V_0 = V(T) = V_0 e^{\beta T} \quad \Rightarrow \quad 2 = e^{\beta T} \quad \Rightarrow \quad \beta = \frac{\ln(2)}{T} \overset{T>1}{<} \ln(2) < 1 \,.$$

Therefore, it is reasonable to expect $\beta \in (0,1)$. Under optimal nutrient conditions, the cell population size increases. This translates to the parameter relation $\beta > \lambda$, which can also be written as $\lambda = c_1 \cdot \beta$ with $c_1 \in (0,1)$. Since the parameter c_1 is bounded, we calibrate this one instead of λ. A similar reparametrization can be done for the starvation rate λ_{st}: it is reasonable to assume that the cells die faster from starvation than from natural causes in case of a nutrient-free environment, which leads to $\lambda_{\text{st}} > \lambda \Rightarrow \lambda_{\text{st}} = \frac{\lambda}{c_2} = \frac{c_1}{c_2}\beta$ with $c_2 \in (0,1)$. Furthermore, the underlying experimental setting motivates the assumption that the population size does not exceed 3×10^5 cells per milliliter. All cell lines are seeded in initial densities below the carrying capacity and $V_0 = 1$ is the largest seeding density, i.e. $K \in (1,3)$ is plausible. The restrictions $m > 1$ and $S_{\text{thr}} \in (0,1)$ with $S_{\text{thr}} \ll 1$ are motivated by the modeling framework. Utilizing all this information, we adopt the following prior distributions for the model parameters of $(\mathcal{M}_S)$:

$$\beta \sim \mathcal{U}(0,1)\,, \qquad c_1 \sim \text{Triang}(0,1/2,1)\,, \qquad c_2 \sim \text{Triang}(0,1/2,1)\,,$$
$$K \sim \mathcal{U}(1,3)\,, \qquad m \sim \mathcal{U}(1,12)\,, \qquad \text{and} \qquad S_{\text{thr}} \sim \text{Triang}(0,0,1)\,.$$

The same prior distributions are used for model $\mathcal{M}_\eta$. Its equations only have one additional model parameter α_S, for which we do not have any particular information – we set its prior distribution to

$$\alpha_S \sim \mathcal{U}(0, 12) \,.$$

To consider a certain degree of confidence in the measurements, we assume a small uncertainty variance $\sigma^2 \in (0, 1)$ is more likely than a large one. Therefore, we use $\mathrm{Triang}(0, 0, 1/2)$ as its prior. All experiments are started with cells from a batch with optimal nutrient conditions. If they are put into a nutrient-free environment without going through a weaning process beforehand, they can undergo a starvation shock. This might disturb or decrease the cells' ability to metabolize the chemical for the fluorescence measurements. To consider this in the hyper parameters, we allow the data sets D1–D4 ($S_0 > 0$) and D5 ($S_0 = 0$) to have different uncertainty variances $\sigma^2_{\mathrm{D1:4}}$ resp. σ^2_{D5} and proportionality constants $n_{\mathrm{D1:4}}$ resp. n_{D5}, where potentially $n_{\mathrm{D1:4}} \geq n_{\mathrm{D5}}$. Using the reparametrization $n_{\mathrm{D5}} = c_n \cdot n_{\mathrm{D1:4}}$ with $c_n \in (0, 1)$, leads to the prior distributions

$$\sigma^2_{\mathrm{D1:4}} \,, \ \sigma^2_{\mathrm{D5}} \sim \mathrm{Triang}(0, 0, 1/2) \,, \quad n_{\mathrm{D1:4}} \sim \mathcal{U}(0, 1/2) \,, \quad \text{and} \quad c_n \sim \mathrm{Triang}(0, 1, 1) \,.$$

A large value for the uncertainty variance σ^2 allows a larger deviation of the model solution from the data. For the purpose of model comparison, using the same uncertainty variance for both models is desirable to increase comparability. Hence, we first calibrate each model separately to get an estimate for their variances $\sigma^2_{\mathrm{D1:4}}$ and σ^2_{D5}. Then, we take the average $\bar{\sigma}^2$ of the means respectively over both models, i.e.

$$\bar{\sigma}^2_{\mathrm{D1:4}} = \frac{1}{2} \left(\mathbb{E}_S \left[\sigma^2_{\mathrm{D1:4}} \right] + \mathbb{E}_\eta \left[\sigma^2_{\mathrm{D1:4}} \right] \right) \qquad \text{and} \qquad \bar{\sigma}^2_{\mathrm{D5}} = \frac{1}{2} \left(\mathbb{E}_S \left[\sigma^2_{\mathrm{D5}} \right] + \mathbb{E}_\eta \left[\sigma^2_{\mathrm{D5}} \right] \right) , \tag{2.11}$$

where "S" resp. "η" in the subscripts of the expected value $\mathbb{E}$ indicate the underlying model $\mathcal{M}_S$ resp. $\mathcal{M}_\eta$, which is used for the calibration. These average values $\bar{\sigma}^2_{\mathrm{D1:4}}$ and $\bar{\sigma}^2_{\mathrm{D5}}$ are then used deterministically and are especially not estimated anymore in further calibrations. For better comparability, we start each SMC algorithm from the same prior particle sample. In particular, the algorithm is performed for model $\mathcal{M}_\eta$ and the generated initial sample but without the component regarding to parameter α_S, is used to start the calibration of model $\mathcal{M}_S$.

2.4. Model comparison

The calibration results of each model are compared in a quantitative manner. For this, we calculate different comparison measures and validate the model solutions with the data sets D1–D6. An overview over the applied methods is given in the following paragraphs of this section.

How can the quality of fit be quantified? We use the validation metric proposed in [32] to compare the model prediction at a given point in time with the corresponding set of measurements. Their mismatch is measured as the area between the data distribution F^{data} and the prediction distribution $F^{\mathrm{sol}}_{\mathcal{M}}$ using the calibration results of model $\mathcal{M}$, mathematically defined by the metric

$$d(F^{\mathrm{data}}, F^{\mathrm{sol}}_{\mathcal{M}}) = \int_0^\infty \left| F^{\mathrm{data}}(I) - F^{\mathrm{sol}}_{\mathcal{M}}(I) \right| \, \mathrm{d}I \,. \tag{2.12}$$

For a set $\{I_i\}_{i=1,\ldots,M}$ of M intensity measurements, the data distribution function is given by

$$F^{\text{data}}(I) = \sum_{i=1}^{M} \mathbb{I}(I_i, I) \quad \text{with} \quad \mathbb{I}(I_i, I) = \begin{cases} 1/M & \text{for } I_i \le I, \\ 0 & \text{for } I_i > I. \end{cases}$$

The prediction distribution function $F_{\mathcal{M}}^{\text{sol}}$ is determined approximately: for each weighted particle θ_p of the posterior, we calculate the empirical cumulative distribution function from the solution of model $\mathcal{M}$. We get a set $\{\tilde{I}_p\}_{p=1,\ldots,P}$ with $\tilde{I}_p = n_p \cdot V_p$, where V_p is the model solution using the parameter sample θ_p scaled with the corresponding sample n_p of the proportionality constant. The prediction distribution is then

$$F_{\mathcal{M}}^{\text{sol}}(I) \approx \sum_{p=1}^{P} \mathbb{I}_W(\tilde{I}_p, I) \quad \text{with} \quad \mathbb{I}_W(\tilde{I}_p, I) = \begin{cases} W_p^N & \text{for } \tilde{I}_p \le I, \\ 0 & \text{for } \tilde{I}_p > I, \end{cases}$$

where W_p^N is the final weight of the p-th particle after the SMC algorithm is finished. Figure 4 depicts an exemplary plot of the above distribution functions and the resulting validation metric for a set of four measurements.
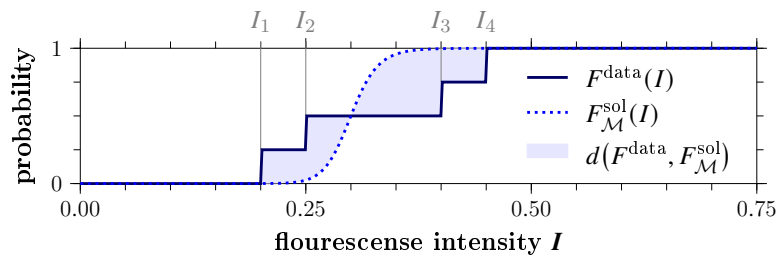

Figure 4. Plots of the data distribution function F^{data} given by four exemplary intensity measurements $I_1, \ldots, I_4$, an approximate prediction distribution function $F_{\mathcal{M}}^{\text{sol}}$ and the corresponding validation metric $d(F^{\text{data}}, F_{\mathcal{M}}^{\text{sol}})$ given by the enclosed area between the two functions

How can the different model approaches be compared? We use the Bayes factor [33] for a quantitative model comparison. It gives the ratio between the respective model evidences. Latter describe the posterior probability of the data given the model type in consideration of the parameter distribution. For a model $\mathcal{M}$, its evidence is defined by the marginal likelihood of a set of measurements $\mathcal{I}$:

$$p(\mathcal{I} \mid \mathcal{M}) = \int_{\Theta} \pi_0(\theta \mid \mathcal{M}) \cdot L(\mathcal{I} \mid \theta, \mathcal{M}) \, d\theta,$$

where $\pi_0(\theta \mid \mathcal{M})$ denotes the prior on θ using model $\mathcal{M}$ and $L(\mathcal{I} \mid \mathcal{M}, \theta)$ the likelihood of $\mathcal{I}$ given $\mathcal{M}$ and its parameters θ. Following [10], the evidence Z_k at the k-th SMC step can be approximated by

$$Z_k \approx Z_0 \cdot \prod_{j=1}^{k} \sum_{p=1}^{P} \frac{L_j(\mathcal{I}_j \mid \theta_p, \mathcal{M})}{L_{j-1}(\mathcal{I}_{j-1} \mid \theta_p, \mathcal{M})} \cdot W_p^{j-1},$$

where P is the number of particles, $L_j(\mathcal{I}_j \mid \theta_p, \mathcal{M})$ is the likelihood of observing all data sets included until calibration step j given the parameter sample θ_p with model $\mathcal{M}$, and W_p^{j-1} denotes the normalized

weight of θ_p at step $j - 1$. Since we start the calibrations with priors, from which we can sample directly, the initial evidence is $Z_0 = 1$. The Bayes factor of two models $\mathcal{M}_1$ and $\mathcal{M}_2$ regarding the same set of measurements $\mathcal{I}$ is then given as

$$\mathcal{Z}_{1:2} = \frac{p(\mathcal{I} \mid \mathcal{M}_1)}{p(\mathcal{I} \mid \mathcal{M}_2)}.$$

For $p(\mathcal{I} \mid \mathcal{M}_1) > p(\mathcal{I} \mid \mathcal{M}_2)$, i.e. $\mathcal{Z}_{1:2} > 1 = 10^0$, the strength of evidence can be described by the following scale [10]:

$$\log_{10}(\mathcal{Z}_{1:2}) \in \begin{cases} (0, \frac{1}{2}] & \text{barely worth mentioning}, \\ (\frac{1}{2}, 1] & \text{substantial support for } \mathcal{M}_1, \\ (1, 2] & \text{strong support for } \mathcal{M}_1, \\ (2, \infty) & \text{decisive support for } \mathcal{M}_1. \end{cases} \tag{2.13}$$

3. Results and Discussion

This chapter investigates the results of the model calibrations. We compare the corresponding estimated parameters of both models $\mathcal{M}_S$ and $\mathcal{M}_\eta$ and quantify the quality of fit to the data (Section 3.1). The calibration results can be used to identify parameter correlations (Section 3.2) and to study the reaction of the cells to different nutrient concentrations (Section 3.3). Additionally to the calibration data sets, the models are validated with further data (Section 3.4).

To investigate dispersion of the results, we perform 12 runs of the SMC algorithm. In the following, the numerical deviations are given in terms of the mean and the 95% prediction/confidence interval of a normal distribution $\mathcal{N}(\mu, \sigma^2)$, i.e. denoted by $\mu \pm 1.96\sigma$. If not indicated otherwise, this formula is also used for error bars in visual representations of the deviations.

According to the equations in (2.11), a precalibration of the models yields the uncertainty variances

$$\bar{\sigma}^2_{D1:4} = 0.0355 \qquad \text{and} \qquad \bar{\sigma}^2_{D5} = 0.2410,$$

which are used deterministically for further calibrations. We see that for the data sets D1–D4 ($S_0 > 0$) the variance is considerably smaller than for data set D5 ($S_0 = 0$). On the one hand, this could be a consequence of the previously described starvation shock, which might disturb the cells' metabolism. On the other hand, data set D5 contains mainly measurements of low cell density, for which the measurement accuracy might be decreased because of weak fluorescence.

3.1. Comparison of the model calibration results

Regarding the estimated posteriors, the ESS is larger than P_{thr} after performing the last SMC step for all model calibrations. This implies that the final approximated posterior distribution is nearly as accurate as sampling directly from the correct probability measure. We calculate the expected values and variances (see Tables 6 and 7) of the model parameters' marginal posteriors.

Table 6. Expected values $\mathbb{E}(\cdot)$ of the posterior distributions for each model parameter.

	Calibrated model parameters			
Model	β	λ	λ_{st}	α_S
$\mathcal{M}_S$	0.435 ± 0.043	0.103 ± 0.042	0.186 ± 0.031	
$\mathcal{M}_\eta$	0.437 ± 0.023	0.106 ± 0.025	0.196 ± 0.018	6.930 ± 2.561
	K	m	S_{thr}	
$\mathcal{M}_S$	1.740 ± 0.133	4.731 ± 2.371	0.104 ± 0.011	
$\mathcal{M}_\eta$	1.731 ± 0.098	5.315 ± 2.964	0.106 ± 0.007	

Table 7. Variances Var$(\cdot)$ of the posterior distributions for each model parameter.

	Calibrated model parameters			
Model	β	λ	λ_{st}	α_S
$\mathcal{M}_S$	0.007 ± 0.017	0.006 ± 0.014	0.010 ± 0.020	
$\mathcal{M}_\eta$	0.010 ± 0.022	0.009 ± 0.020	0.012 ± 0.024	3.211 ± 3.014
	K	m	S_{thr}	
$\mathcal{M}_S$	0.028 ± 0.061	0.751 ± 0.777	0.004 ± 0.009	
$\mathcal{M}_\eta$	0.033 ± 0.067	1.071 ± 1.782	0.004 ± 0.008	

The tables show very similar values for both models and the differences between them lie within the range of the numerical variations of the SMC algorithm. The resulting average parameter values can be concluded as biologically reasonable, since we chose a prior distribution ensuring this. Up to some adaptions regarding units and reparametrization, the reasonableness of the parameters is also supported by comparing their values to the calibration results of [7], where a set of similar models is used together with parts of the same data.

Taking a closer look at the nutrient sensitivity rate α_S, we observe a large average variance of its marginal posterior (Var$(\alpha_S) \approx 3.211$) as well as large numerical deviations of its expected value (± 2.561) and variance (± 3.014). A possible reason for these large values is that only a minority of the measurements contain meaningful information to estimate α_S. On the one hand, this parameter influences the cell growth mainly for early measurements, when the cells react to the nutrient change from the batch colony ($S = 1$) to the provided nutrient supply of the respective experiment ($S = S_0 \in [0, 1]$). On the other hand, the experiments with $S_0 \gg S_{\text{thr}}$ show little measurable reaction of the cells to nutrient changes due to the small value of the nutrient threshold S_{thr}, and therefore lack information to estimate α_S. Overall, approximately less than 10% of the measurements contain meaningful information to estimate α_S, which leads to a higher uncertainty in its estimation. Nevertheless, we can argue in Section 3.3 that this does not have a big influence on the solution of model $\mathcal{M}_\eta$.

We investigate the ratio of the resulting values of the validation metric (2.12) of each model. This does not show any preference of a particular model (see Table 10 in supplementary Section 4.3). To get a quantitative statement for the models' quality of fit to the data, we calculate the logarithm of the Bayes factor: $\log_{10}(\mathcal{Z}_{\eta:S}) = \log_{10}\left(\frac{p(\mathcal{I}\,|\,\mathcal{M}_\eta)}{p(\mathcal{I}\,|\,\mathcal{M}_S)}\right)$. In Figure 5 the average trend of this value over the course of model calibration steps is shown and interpreted according to the scale given in (2.13).

Additionally to the averaged trend, the plot shows the evolution of the Bayes factor for one particular run of the algorithm (dashed line). For this run, the resulting expected values of the parameters' posteriors match approximately with the ones in Table 6. We will refer to the corresponding posteriors later in Section 3.2, when we investigate the correlation between the model parameters.
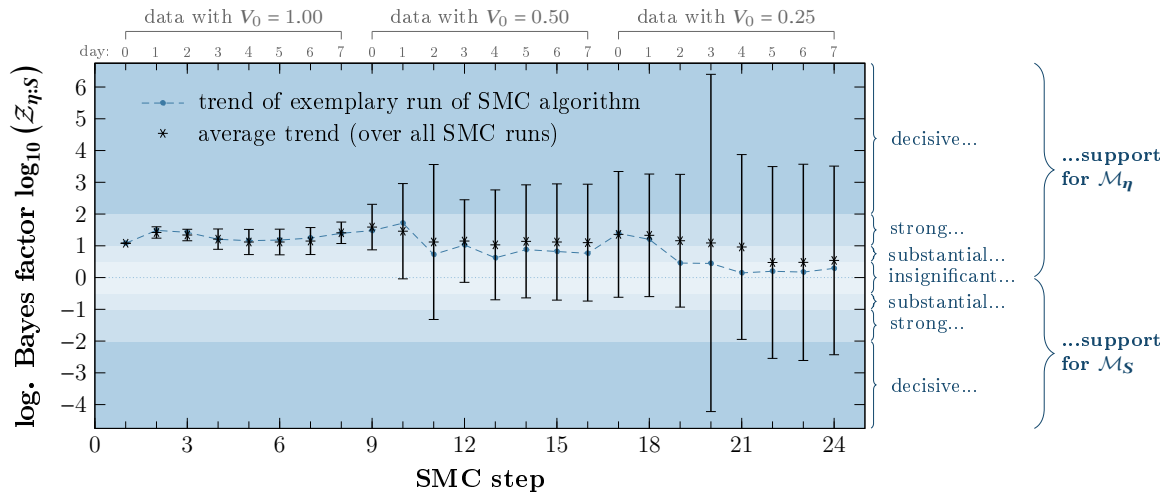

Figure 5. Logarithm of the Bayes factor $\mathcal{Z}_{\eta:S}$ over the course of the SMC steps, where the error bars indicate the 95% confidence interval resulting from the various runs of the SMC algorithm. The dashed trendline shows the evolution for a single run, which estimated the model parameters close to the average values of all runs. The gray labels on top of the plot give the included data points at each SMC step

For the first eight calibration steps, the Bayes factor indicates strong evidence to prefer model $\mathcal{M}_\eta$ over $\mathcal{M}_S$. Until this point only data regarding $V_0 = 1.00$ are considered. Incremental inclusion of the next data set ($V_0 = 0.50$) weakens the weight of the support for model $\mathcal{M}_\eta$, but still allows to conclude a tendency of the evidence towards this model. This can be observed further until SMC step 19. At step 20 (i.e. with inclusion of data with $t \approx 3$ and $V_0 = 0.25$), the uncertainty suddenly increases drastically, which does not allow a clear interpretation of the Bayes factor anymore. The 95% confidence interval reaches from areas of decisive support for model $\mathcal{M}_\eta$ to decisive support for model $\mathcal{M}_S$. After this step the uncertainty decreases again but stays on a high level, still not allowing a clear interpretation of the Bayes factor. Overall, the average trend of the Bayes factor indicates support for model $\mathcal{M}_\eta$, despite needing an additional variable and parameter compared to model $\mathcal{M}_S$, which is already implicitly penalized by the Bayes factor [33]. The modeling advantages of using the environmental stress level presented in Section 2.1 still hold true, even though the present experimental data are not best-suited for showing quantitatively in a prominent way the benefits of this approach. Further investigations with better-suited data are needed to demonstrate the full potential of the environmental stress level.

To understand the large deviations in the Bayes factor, especially the one at SMC step 20, we take a closer look at the model solution and the data. We use the average expected values from Table 6 (model parameters) and supplementary Table 9 (scaling factors $n_{D1:4}$ and n_{D5}) to calculate the corresponding average model solution $V(t)$ and scale the measurements I^V. The resulting time evolution of $V(t)$ in comparison with the calibration data sets D1–D5 can be seen in Figure 6.
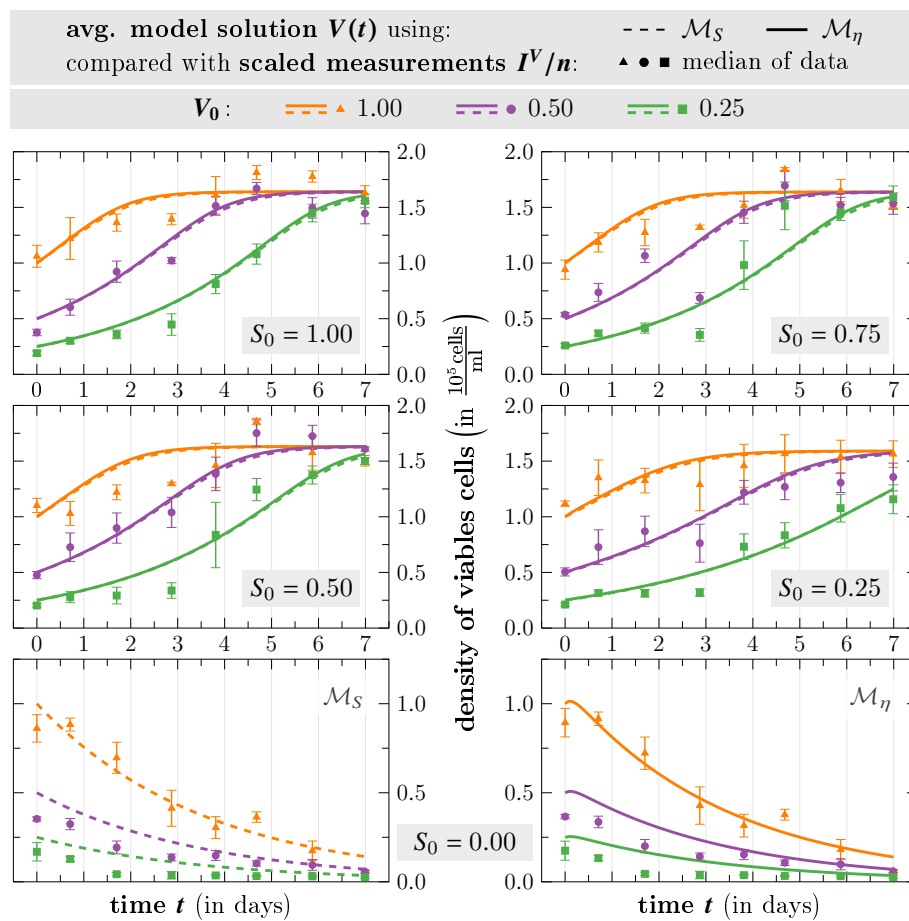
Figure 6. Time evolution of the average model solution V of model $\mathcal{M}_S$ resp. $\mathcal{M}_\eta$ for different V_0 compared to the corresponding median of the scaled measurements I^V/n. The error bars show the median absolute deviation considering the four biological replicates

In general, we observe a good fit for both models. As expected, the similar model parameter values in Table 6 result in nearly the same solution for each model – at least for all cases of $S_0 > 0$ (first two rows of plots). In these cases the estimated proportionality constant is $n_{\mathrm{D1:4}} \approx 0.24$ for both models, whereas for $S_0 = 0$ they differ: $n_{\mathrm{D5}} \approx 0.19$ resp. $n_{\mathrm{D5}} \approx 0.18$ ($\mathcal{M}_S$ resp. $\mathcal{M}_\eta$). This results in different scaling of data set D5 (bottom row of Figure 6). However, the difference of the value of n_{D5} between the models lies in the scale of numerical deviations of the SMC algorithm (see supplemented Table 9).

Taking a closer look into the measurements in Figure 6, the measured population size drops for all S_0 at the third day of the experiment ($t \approx 3$). This measurement might be an outlier and it corresponds to the data included in SMC step 20, where we observe a large deviation in the Bayes factor. To investigate if this drop influences the quality of the fit, we consider the underlying measurement uncertainty. Dependent on the uncertainty variance σ^2, we can calculate the percentiles $P_{5\%}$ and $P_{95\%}$ of the Gamma distributed uncertainty factor ε:

$$\bar{\sigma}^2_{\mathrm{D1:4}} = 0.0355 \quad \Rightarrow \quad P_{5\%} = 0.712 \,, \ P_{95\%} = 1.329 \,,$$
$$\bar{\sigma}^2_{\mathrm{D5}} = 0.2410 \quad \Rightarrow \quad P_{5\%} = 0.350 \,, \ P_{95\%} = 1.920 \,.$$

These determine the 90% uncertainty range (2.7) for the calibration data D1–D4 resp. D5, i.e. give an interval around the solution $V(t)$, where the model expects 90% of the measurements.

We want to know how many measurements are actually situated where they are expected. For each data set, we count the scaled measurements I^V/n, which are situated below/within/above the 90% uncertainty range of the corresponding average solution V. A graphic overview over the calculated percentages can be found in the supplementary Section 4.3 (Figure 10). We see that averaging over the data sets, for both models roughly 90% of the scaled data is actually situated within the 90% uncertainty range. Regarding the remaining data points, we observe that the model solutions tend to be larger than the measurements, since about 8.3–8.5% of the data points lie below the 90% range, whereas only about 2.1% are above. To investigate this observation, we take a closer look at the data below the 90% range by checking if the discrepancy focuses on specific measurements, see Table 8.

Table 8. Percentage of data points in D1–D5, which are below the 90% uncertainty range for different initial cell densities V_0 (rows) and days t (columns). If there are numerical deviations between the runs, they are given in terms of the 95% confidence interval of a normal distribution.

Model	V_0	Day of measurement (i.e. t)							
		0	1	2	3	4	5	6	7
$\mathcal{M}_S$	1.00	0	0.8 ± 3.8	0	0	0	0	0	0
	0.50	5.8 ± 3.8	0	5.8 ± 3.8	35.0	0	0	0	0
	0.25	13.8 ± 6.1	5.0	30.8 ± 3.8	73.8 ± 4.4	5.4 ± 2.8	0	1.3 ± 4.4	4.2 ± 3.8
$\mathcal{M}_\eta$	1.00	0	0	0.8 ± 3.8	5.0	3.3 ± 4.8	0	5.0	5.0
	0.50	5.0	0	7.9 ± 5.0	35.0	0	0	0	0
	0.25	14.2 ± 3.8	5.0	32.1 ± 5.0	72.9 ± 5.0	5.4 ± 2.8	0	1.3 ± 4.4	5.0

On the third day, the measurements show an extraordinarily large amount of data below the 90% uncertainty range, where only 5% would be expected. This supports the hypothesis that the measurements on that day might be outliers. The percentages are higher the smaller V_0 is: approx. 35% resp. 68–78% of the measurements are below the 90% range for $V_0 = 0.50$ resp. $V_0 = 0.25$. This negative correlation is expectable, since the width of the range $[V \cdot P_{5\%}, V \cdot P_{95\%}]$ decreases with smaller V (see right side of Figure 2). This also explains another observation: within the measurements of a particular day, more data points tend to be below the 90% range for smaller V_0. Overall, these discrepancies between measurements and expectation can be a reason for the high deviations of the Bayes factor when including data with $V_0 \leq 0.50$ in the SMC calibration (see Figure 5). In particular, the sudden increase in deviation at SMC step 20 can be linked to the potential outliers for $t \approx 3$ and $V_0 = 0.25$.

The decrease of population size on the third day can be observed for almost all seeding densities V_0 and nutrient saturations $S_0 > 0$, although there is no biological reason for this as long as there are nutrients available. Hence, it is reasonable to assume that the discrepancies result from an unknown experimental bias rather than modeling inaccuracy or biological variation.

3.2. Correlations between the model parameters

To identify parameter correlations, we investigate the estimated posteriors from a single run of the SMC algorithm. We choose a run whose resulting expected values of the parameters' posteriors match approximately the ones in Table 6. Figure 7 shows the corresponding marginal posteriors of the model parameters for each model (first row/column) and samples drawn from the 2D posterior distributions of pairwise parameter combinations. The shape of the scatter plots can give an impression regarding the correlation of parameters.
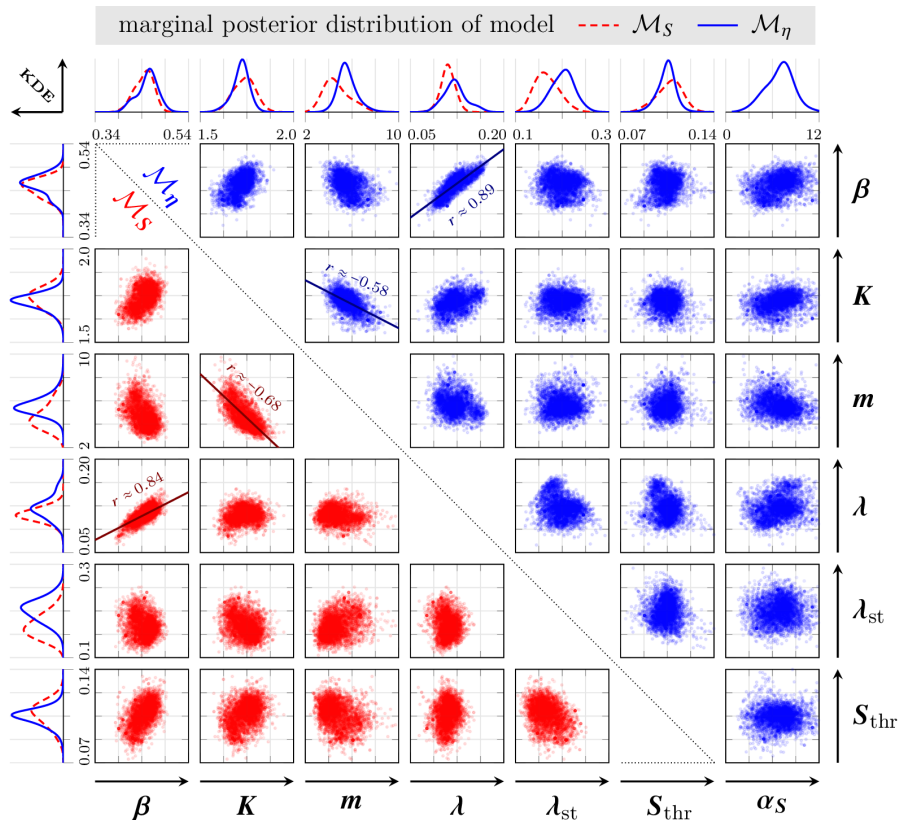
Figure 7. The horizontal/vertical axis labels are given on the bottom/right side. Plots in first row/column: Marginal posterior distributions of the model parameters using model $\mathcal{M}_S$ resp. $\mathcal{M}_\eta$ (note: magnitude of KDE axis is neglected and not the same for the different parameters). Remaining plots: Scatter plots for 5000 samples drawn from the 2D distributions of corresponding pairwise parameter combinations (red, below diagonal: $\mathcal{M}_S$; blue, above diagonal: $\mathcal{M}_\eta$). A regression line and correlation coefficient r is depicted if a linear correlation is observable

It is observable that the SMC algorithm leads to unimodal distributions. The scatter plots indicate no correlation for most of the parameters, as the points are not concentrated along a curve. In particular, the stress-related parameter α_S does not correlate with other parameters (last column of scatter plots). For both models, a linear correlation can be observed between β and λ ($\mathcal{M}_S$: $r \approx 0.84$; $\mathcal{M}_\eta$: $r \approx 0.89$) as well as between K and m ($\mathcal{M}_S$: $r \approx -0.68$; $\mathcal{M}_\eta$: $r \approx -0.58$). These correlations are statistically significant, since their p-values are sufficiently small ($\ll 0.01$).

We have seen in Section 2.1 that the ratio between the growth rate β and the natural death rate λ is crucial for the behavior of the population: it determines the "net" growth rate $\tilde{\beta} = \beta - \lambda$ and the "net" capacity $\tilde{K} = K(1 - \lambda/\beta)^{1/m}$. Since $\tilde{\beta}, \tilde{K} > 0$ should hold, those quantities require the relation $\beta > \lambda$. This motivates the observed strong positive correlation ($r > 0.7$). For the carrying capacity K and the parameter m, we see a moderate negative correlation ($-0.7 < r < -0.5$). These parameters regulate the proliferation contact inhibition in the ODEs of both models $\mathcal{M}_S$ and $\mathcal{M}_\eta$ with the term $(V/K)^m$. This term is smaller, the larger K or m are. Hence, these parameters need to be negatively correlated to mathematically describe a certain level of contact inhibition.

3.3. Influence of different nutrient concentrations on the cell behavior

To affect cell growth and death, both models use the same influence functions $\delta^{\pm}(S)$. In model $\mathcal{M}_S$ these functions scale the rates β and λ_{st} directly, whereas in model $\mathcal{M}_\eta$ they scale them indirectly via the stress level $\eta(t)$. Therefore, the value of the nutrient sensitivity threshold S_{thr} should not depend on the choice of the model. This behavior is actually observable in the results: The expected value of the estimated nutrient sensitivity threshold S_{thr} appears to be nearly identical for both models with $\mathbb{E}[S_{\text{thr}}] \approx 0.1$ (see Table 6), resulting in the same shape of the influence function $\delta^-(S)$.

We use the corresponding expected values from Table 6 to calculate the stress level in model $\mathcal{M}_\eta$. The resulting time evolution of the stress level $\eta(t)$ is depicted in the first two plots of Figure 8. The left plot shows $\eta(t)$ for fixed S_{thr} and varying α_S, while the middle one also fixes α_S to its calibrated mean. Fixing S_{thr} is feasible, as we concluded in previous Section 3.2 that it is uncorrelated to α_S. The third plot shows the corresponding influence functions $\delta^{\pm}(S)$.
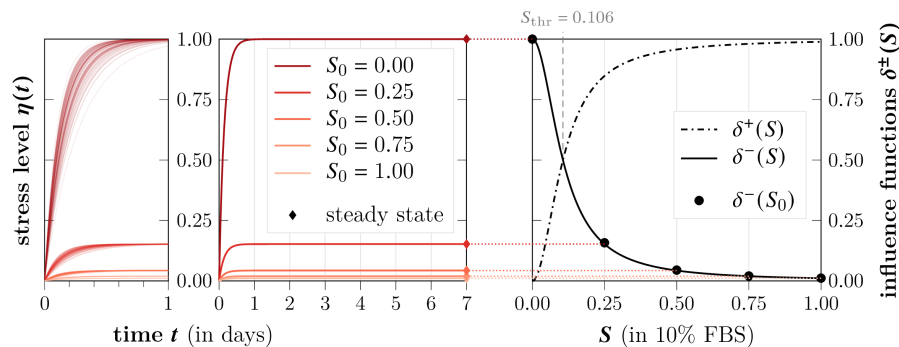

Figure 8. Time evolution of the environmental stress level $\eta(t)$ with $S_{\text{thr}} = 0.106$ for different nutrient conditions S_0 (left plot: $\alpha_S \sim \mathcal{N}(6.930, 2.567/1.96)$; middle plot: $\alpha_S = 6.930$), where its steady states coincide with the corresponding values $\delta^-(S_0)$ of the influence function (right plot)

Since the estimated value of the sensitivity parameter α_S is large, the steady state is already reached after approximately one day. This holds despite the relatively large deviations of the calibrated value for α_S (see left plot). Therefore, variation of α_S within the calibrated scale does not change the solution of model $\mathcal{M}_\eta$ significantly. As calculated in Section 2.1, the steady state of $\eta(t)$ coincides with the function value $\delta^-(S_0)$ of the influence function (see right plot). Hence, the high nutrient sensitivity explains why the calibrated model parameters and the resulting model solutions turn out to be very similar.

3.4. Model validation with additional data and "limit model" $\mathcal{M}_{opt}$

For additional validation, we calculate the average solution of the "limit model" $\mathcal{M}_{opt}$ using the posteriors (determined with data sets D1–D5) to compare it with data set D6. In this experiment tumor cells were seeded in five different initial densities and supplied with 10% FBS for the duration of 21 days (see Table 5). Hence, it holds $S_0 = 1 > 0$ (like in data set D1) and we use $\mathbb{E}[n_{D1:4}] = 0.24$ to scale the measurements. The number of viable cells is calculated by inserting the estimated expected values of the model parameters β, K, m and λ from Table 6 into the analytical solution (2.1). Figure 9 compares the corresponding scaled data with the calculated time course of $V(t)$.

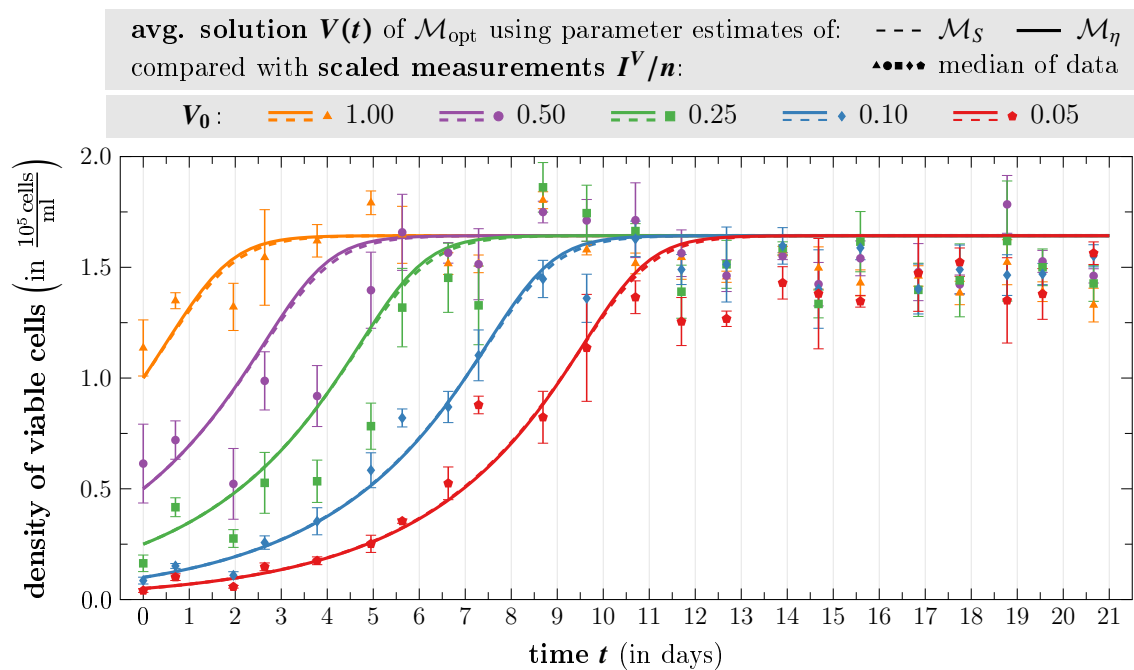**Figure 9.** Time evolution of the average model solution V of model $\mathcal{M}_{opt}$ for different V_0 using the parameter estimates from the calibration of model $\mathcal{M}_S$ resp. $\mathcal{M}_\eta$ compared to the corresponding median of the scaled measurements $I^V/0.24$. The error bars show the median absolute deviation considering the four biological replicates

As expected, both solutions are very similar, as for model $\mathcal{M}_{opt}$ only the parameters β, K, m, and λ are relevant, which have similar estimates for both $\mathcal{M}_S$ and $\mathcal{M}_\eta$. Furthermore, we see a good fit to the data (see also ratios of validation metrics in Table 10 in the supplement). It is observable that the model solutions tend to overestimate the data after the steady state is reached. In reality, population growth often follows the concept of "overshoot": this describes the observation that a population might exceed its carrying capacity and drop back down afterwards (see e.g. [34]). Reasons can be a lag in sensing the available resources in the environment or the time interval between birth and death. The measurements starting with $V_0 \in \{1.00, 0.50, 0.25\}$ (orange, purple, and green markers in Figure 9) suggest an observable overshoot effect, but the model solutions cannot reproduce this phenomenon due to the limited structure of model $\mathcal{M}_{opt}$. This could explain the discrepancy between solution and data close to the carrying capacity.

4. Conclusion

In summary, the presented results show that model $\mathcal{M}_\eta$ is an appropriate alternative way to describe the influences of nutrient changes on viable cells. The model preserves positivity and boundedness of the solutions and calibration yields biologically reasonable parameter values. Keeping in mind that large deviations from the data sets D1–D5 focus on a particular measurement ($t \approx 3$), we altogether see a good agreement to the data for different nutrient scenarios. For the simple setting with the nutrients as the only environmental variable, there is no strong need from a quantitative perspective to model its influence by using the additional variable η. But even in this special case, the Bayes factor indicates a preference for including the environmental stress level η into modeling. A mathematical model should be kept as simple as possible, but as additional influencing factors may become relevant, a suitable and flexible modeling interface is important. From a qualitative perspective, the general modeling approach (2.3) can provide such a setup.

The introduction of an artificial stress level to describe different environmental factors in a collective manner has several advantages. Due to the modular structure of the general ODE for the stress level η (equation 2.3), this model is easily extendable to several environmental variables. This enables the consideration of different sensitivities of the cells to changes in specific environmental factors. If the values of the corresponding sensitivity rates $\alpha_j^\pm$ can be recovered from data, the magnitude of them can give insight into time scales of the cells' reaction to different environmental changes. This information can be used on the one hand to simplify the model by quasi-steady state assumptions for environmental variables regarding large sensitivity rates. On the other hand, if some sensitivity rates are estimated to be close to zero, the model can be reduced easily by setting them to zero and neglecting the ODEs of the corresponding environmental variables. The resulting quality of fit of the simplified model can then be investigated e.g. by the Bayes factor.

Overall, the model allows to describe a survival strategy of the cells assuming they adjust their behavior to avoid a stressful environment by prioritizing according to their sensitivities. If spatial inhomogeneity is considered (i.e. if the model is extended to a system of partial differential equations), such an avoidance strategy can be modeled by influencing the cells' motility. This can be mathematically described similarly to chemotaxis: cells might try to move preferably to areas with a low stress level to improve their growth/survival conditions. Future work aims to investigate the application of the proposed general model in such a setting, especially considering mechanical environmental features like pressure or stiffness of the surrounding tissue, as considered experimentally e.g. in [35]. New experimental approaches like "organ-on-a-chip" will allow for even more realistic and nevertheless (as *in vitro* technique) well-controlled settings and corresponding parameter calibration [36, 37].

Acknowledgments

Conflict of interest

All authors declare no conflicts of interest in this paper.

References

1. L. Preziosi (ed.), *Cancer Modelling and Simulation*, 1st edition, CRC Press, 2003.

2. H. Byrne, T. Alarcon, M. Owen, S. Webb, P. Maini, Modelling aspects of cancer dynamics: A review, *Philosoph. Transact. Ser. A Math. Phys. Eng. Sci.*, **364** (2006), 1563–1578. https://doi.org/10.1098/rsta.2006.1786

3. K. Brindle, New approaches for imaging tumour responses to treatment, *Nat. Rev. Cancer*, **8** (2008), 94–107. https://doi.org/10.1038/nrc2289

4. D. A. Hormuth, A. M. Jarrett, E. A. B. F. Lima, M. T. McKenna, D. T. Fuentes, T. E. Yankeelov Mechanism-based modeling of tumor growth and treatment response constrained by multiparametric imaging data, *JCO Clin. Cancer Inform.*, **3** (2019), 1–10. https://doi.org/10.1200/CCI.18.00055

5. T. E. Yankeelov, R. G. Abramson, C. C. Quarles, Quantitative multimodality imaging in cancer research and therapy, *Nat. Rev. Clin. Oncol.*, **11** (2014), 670–680. https://doi.org/10.1038/nrclinonc.2014.134

6. C. S. Szot, C. F. Buchanan, J. W. Freeman, M. N. Rylander, 3D *in vitro* bio-engineered tumors based on collagen I hydrogels, *Biomaterials*, **32** (2011), 7905–7912. https://doi.org/10.1016/j.biomaterials.2011.07.001

7. E. A. B. F. Lima, N. Ghousifam, A. Ozkan, J. T. Oden, A. Shahmoradi, M. N. Rylander, et al., Calibration of multi-parameter models of avascular tumor growth using time resolved microscopy data, *Sci. Rep.*, **8** (2018), 1–17. https://doi.org/10.1038/s41598-018-32347-9

8. C. P. Robert, G. Casella, G. Casella, *Monte Carlo statistical methods*, 2nd edition, Springer-Verlag, New York, 2004. https://doi.org/10.1007/978-1-4757-4145-2

9. A. Beskos, A. Jasra, E. A. Muzaffer, A. M. Stuart, Sequential Monte Carlo methods for Bayesian elliptic inverse problems, *Stat. Comput.*, **25** (2015), 727–737. https://doi.org/10.1007/s11222-015-9556-7

10. Y. Zhou, A. M. Johansen, J. A. Aston, Toward automatic model comparison: An adaptive sequential Monte Carlo approach, *J. Comput. Graph. Statist.*, **25** (2016), 701–726. https://doi.org/10.1080/10618600.2015.1060885

11. R. M. Neal, Annealed importance sampling, *Stat. Comput.*, **11** (2001), 125–139. https://doi.org/10.1023/A:1008923215028

12. N. Chopin, A sequential particle filter method for static models, *Biometrika*, **89** (2002), 539–552. https://doi.org/10.1093/biomet/89.3.539

13. P. Del Moral, A. Doucet, A.Jasra, Sequential monte carlo samplers, *J. Royal Stat. Soc. Ser. B Stat. Methodol.*, **68** (2006), 411–436. https://doi.org/10.1111/j.1467-9868.2006.00553.x

14. F. Liang, W. H. Wong, Real-parameter evolutionary Monte Carlo with applications to Bayesian mixture models, *J. Am. Stat. Assoc.*, **96** (2001), 653–666. https://doi.org/10.1198/016214501753168325

15. A. Jasra, D. A. Stephens, C. C. Holmes, On population-based simulation for static inference, *Stat. Comput.*, **17** (2007), 263–279. https://doi.org/10.1007/s11222-007-9028-9

16. G. Evensen, The ensemble Kalman filter: Theoretical formulation and practical implementation, *Ocean Dynam.*, **53** (2003), 343–367. https://doi.org/10.1007/s10236-003-0036-9

17. C. Schillings, A. M. Stuart, Analysis of the ensemble Kalman filter for inverse problems, *SIAM J. Numer. Analys.*, **55** (2017), 1264–1290. https://doi.org/10.1137/16M105959X

18. D. Blömker, C. Schillings, P. Wacker, S. Weissmann, Well posedness and convergence analysis of the ensemble Kalman inversion, *Inverse Probl.*, **35** (2019), 085007. https://doi.org/10.1088/1361-6420/ab149c

19. F. J. Richards, A flexible growth function for empirical use, *J. Exp. Bot.*, **10** (1959), 290–301. https://doi.org/10.1093/jxb/10.2.290

20. M. Stoker, H. Rubin, Density dependent inhibition of cell growth in culture, *Nature*, **215** (1967), 171-–172. https://doi.org/10.1038/215171a0

21. G. Aubert, J. F. Aujol, A variational approach to removing multiplicative noise, *SIAM J. Appl. Math.*, **68** (2008), 925–946. https://doi.org/10.1137/060671814

22. Y. M. Huang, Yu-Mei, M. K. Ng, and Y. W. Wen, A new total variation method for multiplicative noise removal, *SIAM J. Imag. Sci.*, **2** (2009), 20–40. https://doi.org/10.1137/080712593

23. G. Steidl, T. Teuber, Removing multiplicative noise by Douglas-Rachford splitting methods, *J. Math. Imag. Vis.*, **36** (2010), 168–184. https://doi.org/10.1007/s10851-009-0179-5

24. R. Šášik, E. Calvo, J. Corbeil, Statistical analysis of high-density oligonucleotide arrays: a multiplicative noise model, *Bioinformatics*, **18** (2002), 1633–1640. https://doi.org/10.1093/bioinformatics/18.12.1633

25. A. M. Stuart, Inverse problems: A Bayesian perspective, *Acta Numer.*, **19** (2010), 451–559. https://doi.org/10.1017/S0962492910000061

26. N. Kantas, A. Beskos, A. Jasra, Sequential Monte Carlo methods for high-dimensional inverse problems: A case study for the Navier–Stokes equations, *SIAM/ASA J. Uncertain. Quan.*, **2** (2014), 464–489. https://doi.org/10.1137/130930364

27. P. Del Moral, *Feynman-Kac Formulae*, Springer-Verlag, New York, 2004. https://doi.org/10.1007/978-1-4684-9393-1

28. M. Bulté, J. Latz, E. Ullmann, A practical example for the non-linear Bayesian filtering of model parameters, in *Quantification of Uncertainty: Improving Efficiency and Technology*, Springer, Cham, (2020), 241–272. https://doi.org/10.1007/978-3-030-48721-8_11

29. H. Haario, E. Saksman, J. Tamminen, An adaptive Metropolis algorithm, *Bernoulli*, **7** (2001), 223–242.

30. *Build models that drive breakthroughs*, Technology Networks, 2018. Available from: `https://www.technologynetworks.com/cancer-research/ebooks/build-models-that-drive-breakthroughs-311843`.

31. K. Wrzesinski, A. Rogowska-Wrzesinska, R. Kanlaya, K. Borkowski, V. Schwämmle, J. Dai, et. al., The cultural divide: Exponential growth in classical 2D and metabolic equilibrium in 3D environments, *PloS One*, **9** (2014), e106973. https://doi.org/10.1371/journal.pone.0118050

32. S. Ferson, W. L. Oberkampf, L. Ginzburg, Model validation and predictive capability for the thermal challenge problem, *Computer Methods in Applied Mechanics and Engineering*, **197** (2008), 2408–2430. https://doi.org/10.1016/j.cma.2007.07.030

33. R. E. Kass, A. E. Raftery, Bayes factors, *J. Am. Statist. Assoc.*, **90** (1995), 773–795. https://doi.org/10.1080/01621459.1995.10476572

34. O. J. Schmitz, *Ecology and ecosystem conservation.* Island Press, 2013.

35. A. Özkan, D. L. Stolley, E. N. Cressman, M. McMillin, S. DeMorrow, T. E. Yankeelov, et. al., Tumor microenvironment alters chemoresistance of hepatocellular carcinoma through CYP3A4 metabolic activity, *Front. Oncol.*, **11** (2021). https://doi.org/10.3389/fonc.2021.662135

36. A. Ozkan, N. Ghousifam, P. J. Hoopes, T. E. Yankeelov, M. N. Rylander, In vitro vascularized liver and tumor tissue microenvironments on a chip for dynamic determination of nanoparticle transport and toxicity, *Biotechnol. Bioeng.*, **116** (2019), 1201–1219. https://doi.org/10.1002/bit.26919

37. A. Özkan, D. L. Stolley, E. N. Cressman, M. McMillin, T. E. Yankeelov, M. N. Rylander, CYP3A4 mediates chemoresistance controlled by cirrhosis and inflammation captured in vascularized hepatocellular carcinoma-on-a-chip, Submitted to *Small*, (2022).

38. P. Hartman, A lemma in the theory of structural stability of differential equations, *Proceed. Am. Math. Soc.*, **11** (1960), 610–620. https://doi.org/10.2307/2034720

39. D. M. Grobman, Homeomorphism of systems of differential equations (in Russian), *Doklady Akademii Nauk SSSR*, **128** (1959), 880–881.

Supplementary

4.1. Analytical solution of model $\mathcal{M}_S$

The following equations (4.1) and (4.3) are the case-dependent analytical solutions of system ($\mathcal{M}_S$) as given in Section 2.1. The corresponding calculations are done in the following paragraphs. It can be shown by differentiation that these derived terms for $V(t)$ actually solve the given initial value problem.

Case 1: $\beta_S > \lambda_S$. Model $\mathcal{M}_S$ has the same form as model $\mathcal{M}_{\text{opt}}$ just with nutrient-scaled growth rate $\beta_S = \delta^+(S_0) \cdot \beta$ and death rate $\lambda_S = \lambda + \delta^-(S_0) \cdot \lambda_{\text{st}}$, each depending on the constant nutrient supply S_0. Hence, for $\beta_S > \lambda_S$ the ODE of ($\mathcal{M}_S$) can again be written in the form of a logistic growth term and the analytical solution can be given analogously to equation (2.1) with scaled rates:

$$V(t) = \left(\frac{(V_0 K)^m (\beta_S - \lambda_S)}{\beta_S V_0^m + \left((\beta_S - \lambda_S) K^m - \beta_S V_0^m \right) \cdot e^{-(\beta_S - \lambda_S)mt}} \right)^{1/m}. \tag{4.1}$$

Case 2: $\beta_S < \lambda_S$. In case of $\beta_S \leq \lambda_S$ the ODE cannot be rewritten in pure logistic form, since the root's argument of the resulting "net" carrying capacity $K(1 - \lambda_S/\beta_S)^{1/m}$ might cause problems. However, it can be shown that for $\beta_S < \lambda_S$ the analytical solution above is still well-defined, since the root's

argument is non-negative:

$$V(t) \overset{(4.1)}{=} \left(\frac{(-1)}{(-1)} \cdot \frac{\overbrace{(V_0 K)^m}^{\geq 0} \overbrace{(\lambda_S - \beta_S)}^{>0}}{\underbrace{\beta_S V_0^m}_{\geq 0} \cdot (-1 + \underbrace{e^{(\lambda_S - \beta_S)mt}}_{>1}) + \underbrace{(\lambda_S - \beta_S) K^m}_{>0} \cdot \underbrace{e^{(\lambda_S - \beta_S)mt}}_{>1}} \right)^{1/m}. \tag{4.2}$$

Case 3: $\beta_S = \lambda_S$. For $\lambda_S \to \beta_S$ the fraction inside the root of formula (4.2) degenerates, as numerator and denominator both tend to zero. Since $V(t)$ is a differentiable function in the parameter space, we can derive the analytical solution in this case using l'Hospital's rule (l'H). It holds

$$\lim_{\lambda_S \to \beta_S} V(t)^m = \lim_{\lambda_S \to \beta_S} \frac{(V_0 K)^m \overbrace{(\lambda_S - \beta_S)}^{\to 0}}{\beta_S V_0^m \cdot \underbrace{(e^{(\lambda_S - \beta_S)mt} - 1)}_{\to 0} + \underbrace{(\lambda_S - \beta_S)}_{\to 0} K^m \cdot \underbrace{e^{(\lambda_S - \beta_S)mt}}_{\to 1}}$$

$$\overset{\text{l'H}}{=} \lim_{\lambda_S \to \beta_S} \frac{(V_0 K)^m}{\beta_S V_0^m \cdot mt \cdot e^{(\lambda_S - \beta_S)mt} + K^m \cdot e^{(\lambda_S - \beta_S)mt} + (\lambda_S - \beta_S) K^m \cdot mt \cdot e^{(\lambda_S - \beta_S)mt}}$$

$$= \lim_{\lambda_S \to \beta_S} \frac{(V_0 K)^m}{\left(mt \cdot (\beta_S V_0^m + \underbrace{(\lambda_S - \beta_S)}_{\to 0} K^m) + K^m \right) \cdot \underbrace{e^{(\lambda_S - \beta_S)mt}}_{\to 1}} = \frac{(V_0 K)^m}{mt\beta_S V_0^m + K^m},$$

yielding $\quad \lim_{\lambda_S \to \beta_S} V(t) = V_0 K \cdot (mt\beta_S V_0^m + K^m)^{-1/m}. \tag{4.3}$

4.2. Mathematical analysis of the ODE models

In the following paragraphs, the solutions of all models are analyzed in terms of positivity and boundedness. Furthermore, we give the respective steady states and check for their stability.

Positivity and boundedness of the solutions. In a reasonable biological context, the size of a cell population cannot grow arbitrarily because of environmental limitations. Hence, for each of the presented models to be feasible, the corresponding solution $V(t)$ has to stay bounded and especially non-negative. The same properties are desired for the environmental stress level η.

Environmental stress level. Before analyzing the variable V in all models, we start with the stress level η from system $(\mathcal{M}_\eta)$. Given $\eta(0) = \eta_0 \in [0, 1]$, its ODE yields

$$\dot{\eta} = \alpha_S(\delta^-(S_0) - \eta) \begin{cases} > 0 & \text{for } \eta < \delta^-(S_0) \\ = 0 & \text{for } \eta = \delta^-(S_0) \ , \\ < 0 & \text{for } \eta > \delta^-(S_0) \end{cases} \tag{4.4}$$

leading to both positivity and boundedness of η: $0 \leq \eta \leq \max\{\eta_0, \delta^-(S_0)\} \leq 1$. Analogous bounds can be calculated for the general ODE (2.3) for the environmental stress level, since $\alpha_i^\pm, \delta_j^\pm \geq 0$.

Density of viable cells. Regarding the remaining variable V, positivity and boundedness in model $\mathcal{M}_{\text{opt}}$ follow directly from the fact that its ODE can be rewritten as a logistic growth model with positive growth rate $\tilde{\beta}$ and capacity $\tilde{K}$, preserving positivity and boundedness by definition. To

conclude these properties also for models $\mathcal{M}_S$ and $\mathcal{M}_\eta$, we analyze a more general, non-autonomous model of the form

$$
\begin{cases}
\dot{V} = \underbrace{h(t) \cdot \beta}_{= \beta_h(t)}\, V\left(1 - \left(\frac{V}{K}\right)^m\right) - \underbrace{\left(\lambda + (1 - h(t)) \cdot \lambda_{\text{st}}\right)}_{= \lambda_h(t)} V\,, \\[2mm]
V(0) = V_0\,,
\end{cases}
\tag{4.5}
$$

where h is a potentially time-dependent and bounded function: $h(t) \in [0, 1]\ \forall t \geq 0$. In fact, problem (4.5) is a generalization of both systems $(\mathcal{M}_S)$ and $(\mathcal{M}_\eta)$, where respectively $h(t) = \delta^+(S_0)$ and $h(t) = 1 - \eta(t) \overset{(2.4)}{=} 1 - \delta^-(S_0) \cdot (1 - e^{-\alpha_S t}) - \eta_0 e^{-\alpha_S t}$. For $h(t) = \delta^+(S_0)$ the required boundedness of h is given by definition of $\delta^+(S)$ and for $h(t) = 1 - \eta(t)$ it is given by the previously derived bounds: $\eta \in [0, 1]$. Hence, we can analyze the general problem (4.5) for positivity and boundedness instead and transfer the results to models $\mathcal{M}_S$ and $\mathcal{M}_\eta$.

With $h(t) \in [0, 1]\ \forall t \geq 0$ and $V_0 \geq 0$, positivity follows directly by observing $\dot{V} = 0 \geq 0$ for $V = 0$. To analyze boundedness, we check the sign of the derivative of V. For $\lambda_h(t) \geq \beta_h(t)$ we estimate:

$$
\dot{V} \overset{(4.5)}{=} \beta_h(t)\, V\left(1 - \left(\frac{V}{K}\right)^m\right) - \lambda_h(t) V = \beta_h(t)\, V\left(\underbrace{1 - \frac{\lambda_h(t)}{\beta_h(t)}}_{\leq 0} - \left(\frac{V}{K}\right)^m\right) \leq 0\,.
$$

In case of $\lambda_h(t) < \beta_h(t)$, we can rewrite the latter term further by placing $\left(1 - \frac{\lambda_h(t)}{\beta_h(t)}\right)$ outside the bracket:

$$
\dot{V} = \underbrace{(\beta_h(t) - \lambda_h(t))}_{>0}\, V\left(1 - \left(\frac{V}{\underbrace{K\,(1 - \lambda_h(t)/\beta_h(t))^{1/m}}_{>0}}\right)^m\right)
\begin{cases}
> 0 & \text{for } V < K\left(1 - \frac{\lambda_h(t)}{\beta_h(t)}\right)^{1/m}, \\[1mm]
= 0 & \text{for } V = K\left(1 - \frac{\lambda_h(t)}{\beta_h(t)}\right)^{1/m}, \\[1mm]
< 0 & \text{for } V > K\left(1 - \frac{\lambda_h(t)}{\beta_h(t)}\right)^{1/m}.
\end{cases}
\tag{4.6}
$$

Together with $\beta_h(t) = h(t) \cdot \beta \leq \beta$ and $\lambda_h(t) = \lambda + (1 - h(t)) \cdot \lambda_{\text{st}} \geq \lambda$, this eventually leads to a time-independent upper bound for the population size:

$$
V \leq \max\left\{V_0\,,\ K\left(1 - \frac{\lambda_h(t)}{\beta_h(t)}\right)^{1/m}\right\} \leq \max\left\{V_0\,,\ K\left(1 - \frac{\lambda}{\beta}\right)^{1/m}\right\}.
$$

Note that relations (4.4) and (4.6) also fit to the non-trivial steady states of the corresponding models (see next paragraph).

Steady states of the models. To calculate the steady states of each problem, we introduce the following notations for the right hand sides of the corresponding autonomous ODEs in each model:

$$
(\mathcal{M}_{\text{opt}}): \ \dot{V} = f_{\text{opt}}(V)\,, \qquad (\mathcal{M}_S): \ \dot{V} = f_S(V)\,, \qquad (\mathcal{M}_\eta): \ \dot{V} = f_\eta(V, \eta)\,, \ \dot{\eta} = g_\eta(V, \eta)\,. \tag{4.7}
$$

The steady states are derived by setting the right hand sides to zero respectively. From a biological point of view, we would expect at least two steady states $\bar{V}$ for the cell density. The cells can either go extinct or the population size tends to the "net" carrying capacity, which depends on the relation

between growth and death rate. For a growing population, we use more general notations during the analysis: Let b be the total growth rate, which is larger than the total death rate a and define the function

$$\mathcal{K} : \left\{ (a,b) \in \mathbb{R}_+^2 \mid a < b \right\} \to (0, K) \quad \text{with} \quad (a,b) \mapsto K(1 - a/b)^{1/m}$$

to be the "net" carrying capacity depending on the rates. Then, the mentioned expected steady states translate to $\bar{V} = 0$ or $\bar{V} = \mathcal{K}(a,b)$.

System $(\mathcal{M}_{\text{opt}})$. For model $\mathcal{M}_{\text{opt}}$ these are actually the only two steady states:

$$\dot{V}|_{V=\bar{V}} \overset{(\mathcal{M}_{\text{opt}})}{=} \beta \bar{V} \left(1 - \left(\frac{\bar{V}}{K} \right)^m \right) - \lambda \bar{V} = \bar{V} \cdot \left(\beta \left(1 - \left(\frac{\bar{V}}{K} \right)^m \right) - \lambda \right) = 0$$

$$\Rightarrow \quad \bar{V} = 0 \quad \text{or} \quad \beta \left(1 - \left(\frac{\bar{V}}{K} \right)^m \right) - \lambda = 0 \overset{\beta > \lambda}{\Longrightarrow} \bar{V} = K \left(1 - \frac{\lambda}{\beta} \right)^{1/m} = \mathcal{K}(\lambda, \beta).$$

Note that the same non-trivial steady state can also be calculated by considering the limit $t \to \infty$ of the corresponding analytical solution (2.1).

System $(\mathcal{M}_S)$. Since model $\mathcal{M}_S$ only differs from model $\mathcal{M}_{\text{opt}}$ by time-independent scaling of the growth and death rate, analogous calculations lead to a similar result:

$$\dot{V}|_{V=\bar{V}} \overset{(\mathcal{M}_S)}{=} \beta_S \bar{V} \left(1 - \left(\frac{\bar{V}}{K} \right)^m \right) - \lambda_S \bar{V} = 0 \quad \Rightarrow \quad \bar{V} = \begin{cases} 0 & \text{for } \lambda_S \geq \beta_S, \\ 0 \text{ or } K \left(1 - \frac{\lambda_S}{\beta_S} \right)^{1/m} = \mathcal{K}(\lambda_S, \beta_S) & \text{for } \lambda_S < \beta_S. \end{cases}$$

Hence, we have the two expected steady states in case of $\lambda_S < \beta_S$, but only the trivial one if $\lambda_S \geq \beta_S$.

System $(\mathcal{M}_\eta)$. For the last model $\mathcal{M}_\eta$, we have to consider a system of two ODEs:

$$(\mathcal{M}_\eta) : \begin{cases} \dot{V}|_{V=\bar{V}, \eta=\bar{\eta}} = (1 - \bar{\eta}) \cdot \beta \bar{V} \left(1 - \left(\frac{\bar{V}}{K} \right)^m \right) - (\lambda + \bar{\eta} \cdot \lambda_{\text{st}}) \bar{V} = 0, \\ \dot{\eta}|_{V=\bar{V}, \eta=\bar{\eta}} = \alpha_S \delta^-(S_0)(1 - \bar{\eta}) - \alpha_S \bar{\eta} \delta^+(S_0) = \alpha_S (\delta^-(S_0) - \bar{\eta}) = 0. \end{cases}$$

Latter equation is independent of V, which directly results in the steady state for the stress level: $\bar{\eta} = \delta^-(S_0)$. Inserting this into the first equation makes it identical to the equation from the previous paragraph for system $(\mathcal{M}_S)$. Therefore, the steady states of problem $(\mathcal{M}_\eta)$ are

$$\begin{pmatrix} \bar{V} \\ \bar{\eta} \end{pmatrix} = \begin{pmatrix} 0 \\ \delta^-(S_0) \end{pmatrix} \text{ for all } \lambda_S, \beta_S \quad \text{and} \quad \begin{pmatrix} \bar{V} \\ \bar{\eta} \end{pmatrix} = \begin{pmatrix} K \left(1 - \frac{\lambda_S}{\beta_S} \right)^{1/m} \\ \delta^-(S_0) \end{pmatrix} = \begin{pmatrix} \mathcal{K}(\lambda_S, \beta_S) \\ \delta^-(S_0) \end{pmatrix}, \text{ if } \lambda_S < \beta_S.$$

Stability of the steady states. The derived steady states can be analyzed for stability. We use the previously introduced notations (4.7) for the right hand sides of the ODEs. Following the theorem of Hartman and Grobman [38, 39], local stability of a steady state can be evaluated by checking the sign of the derivatives of the right hand sides in these points.

Systems $(\mathcal{M}_{\text{opt}})$ *and* $(\mathcal{M}_S)$. For the first two models we have to check the sign of $f'_{\text{opt}}(\bar{V})$ and $f'_S(\bar{V})$ respectively. Keeping the constraint $\beta > \lambda$ in mind, inserting the steady states of model $\mathcal{M}_{\text{opt}}$ yields

$$f_{\text{opt}}(V) = \beta V \left(1 - \left(\frac{V}{K} \right)^m \right) - \lambda V \quad \Rightarrow \quad f'_{\text{opt}}(V) = \beta - \lambda - \frac{(m+1)\beta}{K^m} V^m \quad \text{resulting in}$$

$$f'_{\mathrm{opt}}(0) = \beta - \lambda > 0 \quad \text{and} \quad f'_{\mathrm{opt}}\left(K\left(1 - \frac{\lambda}{\beta}\right)^{1/m}\right) = \beta - \lambda - (m+1)(\beta - \lambda) = -m(\beta - \lambda) < 0 .$$

Hence, the non-trivial steady state is stable, whereas the trivial one is not. Since f_S has the same form as f_{opt} only with scaled growth and death rate, analogous calculations result in

$$f'_S(0) = \beta_S - \lambda_S \begin{cases} > 0 & \text{for } \lambda_S < \beta_S , \\ = 0 & \text{for } \lambda_S = \beta_S , \\ < 0 & \text{for } \lambda_S > \beta_S . \end{cases} \quad \text{and, in case of } \lambda_S < \beta_S : \ f'_S\left(K\left(1 - \frac{\lambda_S}{\beta_S}\right)^{1/m}\right) = -m(\beta_S - \lambda_S) < 0 .$$

The stability of the steady states now depends on the relation between the scaled growth rate β_S and death rate λ_S. For the first case $\lambda_S < \beta_S$, the trivial steady state is locally unstable, whereas the non-trivial one is stable. For $\lambda_S > \beta_S$ only the trivial steady state exists, which is stable.

In case of equality $\lambda_S = \beta_S$, the theorem of Hartman and Grobman cannot be used to determine stability for $\bar{V} = 0$ and we take another approach by investigating the change of a time-dependent perturbation from the steady state $\bar{V} = 0$. Observing that $f_S(V)$ is a polynomial in V of degree $m + 1$ without an absolute term, all but one derivatives vanish in $\bar{V} = 0$:

$$f_S^{(j)}(0) = \begin{cases} -\frac{\beta_S}{K^m} \cdot \frac{(m+1)!}{(m+1-j)!} \cdot 0^{m+1-j} = 0 & \text{for } j < m + 1 , \\ -\frac{\beta_S}{K^m}(m + 1)! = \text{const.} \neq 0 & \text{for } j = m + 1 , \\ 0 & \text{for } j > m + 1 . \end{cases} \tag{4.8}$$

Let $v = v(t) > 0$ be the time-dependent perturbation from the steady state $\bar{V} = 0$ with initial value $v(0) = v_0$. The change of the perturbation in time can be derived by insertion into the ODE, i.e. $f_S(\bar{V} + v(t)) \overset{\mathrm{ODE}}{=} \frac{\mathrm{d}}{\mathrm{d}t}(\bar{V} + v(t)) = \frac{\mathrm{d}}{\mathrm{d}t}v$, and Taylor series expansion of the right hand side:

$$\frac{\mathrm{d}}{\mathrm{d}t}v = f_S(\bar{V} + v(t)) \overset{(4.8)}{=} \left(\sum_{j=0}^{m} \underbrace{f_S^{(j)}(0)}_{=0} \frac{(v-0)^j}{j!}\right) + \left(\frac{f_S^{(m+1)}(0)}{(m+1)!}(v-0)^{m+1}\right) + \left(\sum_{j=m+2}^{\infty} \underbrace{f_S^{(j)}(0)}_{=0} \frac{(v-0)^j}{j!}\right) = -\frac{\beta_S}{K^m}v^{m+1} .$$

The derived ODE in $v(t)$ can be analytically solved by separation of variables, resulting in

$$v(t) = v_0 K \cdot \left(mt\beta_S v_0^m + K^m\right)^{-1/m} .$$

Taking the limit $t \to \infty$ shows that the perturbation vanishes over time, i.e. local stability of the steady state $\bar{V} = 0$ can be concluded.

System $(\mathcal{M}_\eta)$. For the last model $\mathcal{M}_\eta$ we again use Hartman-Grobman theorem to investigate the corresponding locally linearized model. Reminding that the system is given by

$$(\mathcal{M}_\eta) : \begin{cases} \dot{V} = f_\eta(V, \eta) = (1 - \eta) \cdot \beta V(1 - (V/K)^m) - (\lambda + \eta \cdot \lambda_{\mathrm{st}})V , \\ \dot{\eta} = g_\eta(V, \eta) = \alpha_S \delta^-(S_0)(1 - \eta) - \alpha_S \delta^+(S_0)\eta = \alpha_S(\delta^-(S_0) - \eta) , \end{cases}$$

linearization in the steady state $(\bar{V}, \bar{\eta})$ with the Jacobian J yields

$$\begin{pmatrix} \dot{x} \\ \dot{y} \end{pmatrix} = J \begin{pmatrix} x \\ y \end{pmatrix} = \left. \begin{pmatrix} \frac{\partial}{\partial V} f_\eta & \frac{\partial}{\partial \eta} f_\eta \\ \frac{\partial}{\partial V} g_\eta & \frac{\partial}{\partial \eta} g_\eta \end{pmatrix} \right|_{(V,\eta)=(\bar{V},\bar{\eta})} \begin{pmatrix} x \\ y \end{pmatrix}$$

$$= \begin{pmatrix} (1 - \bar{\eta}) \cdot \beta \left(1 - (m+1)(\bar{V}/K)^m\right) - (\lambda + \bar{\eta} \cdot \lambda_{\mathrm{st}}) & -\beta \bar{V} \left(1 - (\bar{V}/K)^m\right) - \lambda_{\mathrm{st}} \bar{V} \\ 0 & -\alpha_S \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}.$$

The eigenvalues $\chi_{1,2}$ of the triangular Jacobian J in dependence on the steady state $(\bar{V}, \bar{\eta})$ are given by

$$\chi_1 = -\alpha_S < 0 \quad \text{and} \quad \chi_2 = \chi_2(\bar{V}, \bar{\eta}) = (1 - \bar{\eta}) \cdot \beta \left(1 - (m+1)(\bar{V}/K)^m\right) - (\lambda + \bar{\eta} \cdot \lambda_{\mathrm{st}}).$$

The first eigenvalue χ_1 is always negative, whereas the sign of χ_2 depends on the relation between β_S and λ_S. In particular, the non-trivial steady state only exists for $\lambda_S < \beta_S$, leading to

$$\chi_2 \left(K \left(1 - \lambda_S/\beta_S\right)^{1/m}, \delta^-(S_0)\right) = \beta_S \left(1 - (m+1)(1 - \lambda_S/\beta_S)\right) - \lambda_S = -m(\beta_S - \lambda_S) < 0,$$

i.e. this steady state is a stable node. For the trivial steady state it holds

$$\chi_2(0, \delta^-(S_0)) = \beta_S - \lambda_S \begin{cases} > 0 & \text{for } \lambda_S < \beta_S, \\ = 0 & \text{for } \lambda_S = \beta_S, \\ < 0 & \text{for } \lambda_S > \beta_S, \end{cases}$$

and, therefore, this steady state is a stable node for $\lambda_S > \beta_S$ and an unstable one for $\lambda_S < \beta_S$. In case of equality $\lambda_S = \beta_S$ the second eigenvalue gets zero and the theorem is again not applicable anymore. To show stability in this case, we use the fact that the ODE of η is independent of V, i.e. $g_\eta(V, \eta) = g_\eta(\eta)$, and thus the equations of model $\mathcal{M}_\eta$ can be analyzed consecutively. On the one hand, it holds

$$g_\eta(\eta) = \alpha_S \left(\delta^-(S_0) - \eta\right) \quad \Rightarrow \quad g'_\eta(\eta) = -\alpha_S < 0$$

and on the other hand, inserting $\bar{\eta} = \delta^-(S_0)$ into the first ODE of the system yields

$$f_\eta(V, \bar{\eta}) = \beta_S V(1 - (V/K)^m) - \lambda_S V = f_S(V).$$

For this equation stability was already shown for model $\mathcal{M}_S$ in case of $\lambda_S = \beta_S$. Eventually, the trivial state of model $\mathcal{M}_\eta$ is stable for this case as well.

4.3. Further model calibration results

Calibrated parameters. For the proportionality constants and reparametrization parameters, the expected values and variances of the posteriors are listed in Table 9.

Table 9. Mean values and variances of the posterior distributions for some calibrated parameters. The deviations between the runs of the SMC algorithm are given in terms of the 95% prediction interval of a normal distribution $\mathcal{N}(\mu, \sigma^2)$: $\mu \pm 1.96\sigma$.

Model		**Calibrated parameters**				
		$n_{\mathrm{D1:4}}$	n_{D5}	c_n	c_1	c_2
$\mathcal{M}_S$	$\mathbb{E}(\cdot)$	0.244 ± 0.006	0.190 ± 0.020	0.780 ± 0.084	0.236 ± 0.076	0.576 ± 0.293
$\mathcal{M}_\eta$		0.243 ± 0.004	0.182 ± 0.022	0.752 ± 0.094	0.240 ± 0.046	0.554 ± 0.159
$\mathcal{M}_S$	$\mathbf{Var}(\cdot)$	0.002 ± 0.004	0.008 ± 0.017	0.034 ± 0.069	0.011 ± 0.024	0.053 ± 0.100
$\mathcal{M}_\eta$		0.002 ± 0.004	0.007 ± 0.014	0.032 ± 0.055	0.016 ± 0.036	0.069 ± 0.122

Uncertainties in the model. In Figure 10 we consider the 90% uncertainty range (2.7) with the percentiles $P_{5\%}$ and $P_{95\%}$ calculated in Section 3.1, using either the calibrated parameters of $\mathcal{M}_S$ or $\mathcal{M}_\eta$ (left/right plot). There are three different markers per data set (vertical axis), showing how many data points (horizontal axis) are situated below/within/above the 90% range. For instance regarding D1 (topmost on vertical axis), in both plots the black bullet marker shows that approx. 87% of the measurements actually lie within the 90% range, whereas the dark/light blue triangle marker shows that about 4%/9% of them are above/below the 90% range.
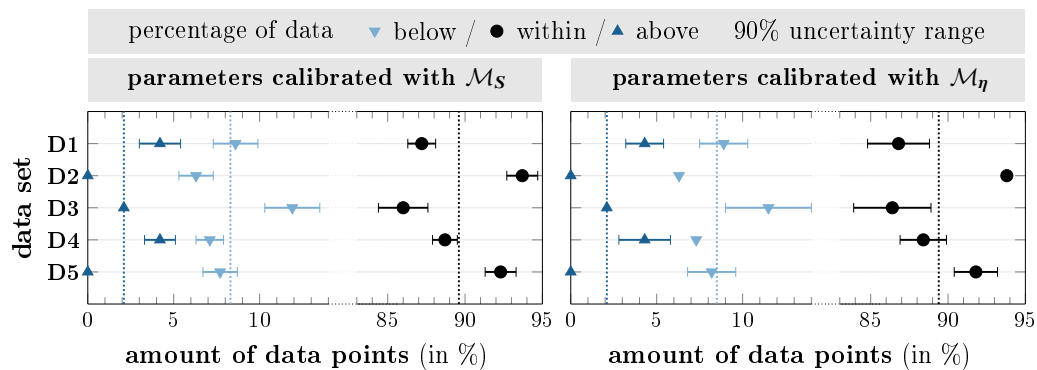

Figure 10. Percentage of data points which are situated below/within/above the 90% range $[V \cdot P_{5\%} , V \cdot P_{95\%}]$ around the solution V. We remind that the error bars indicate the numerical deviation between the SMC runs. The dotted lines in the corresponding color give the average over all data sets D1–D5

Quality of fit to the data. The validation metric (2.12) measures the mismatch between data and model prediction, therefore

$$\frac{d_\eta}{d_S} = \frac{d\left(F^{\text{data}}, F^{\text{sol}}_{\mathcal{M}_\eta}\right)}{d\left(F^{\text{data}}, F^{\text{sol}}_{\mathcal{M}_S}\right)} \begin{cases} \ll 1 & \text{indicates: better fit for } \mathcal{M}_\eta, \\ \cong 1 & \text{indicates: } \mathcal{M}_S \text{ and } \mathcal{M}_\eta \text{ fit equally well}, \\ \gg 1 & \text{indicates: better fit for } \mathcal{M}_S. \end{cases}$$

Calculations for each data set D1–D6 and initial cell number V_0 result in the values given in Table 10. For data set D6 the solution of model $\mathcal{M}_{\text{opt}}$ was used by inserting the average estimated parameters resulting from the calibration of $\mathcal{M}_S$ resp. $\mathcal{M}_\eta$.

Table 10. Ratio d_η/d_S of the validation metrics using model $\mathcal{M}_\eta$ resp. $\mathcal{M}_S$ averaged over all time points considering measurements regarding various V_0 (columns) and data sets, i.e. different S_0 (rows). The values in the last row/column are averaged over all data sets/initial cell densities before taking the mean over the SMC runs. The numerical deviations between the runs are given in terms of the 95% confidence interval of a normal distribution.

| Data set | Initial cell density V_0 | | | | | |
	1.00	**0.50**	**0.25**	**0.10**	**0.05**	**all (avg.)**
D6	1.05 ± 0.54	1.04 ± 0.33	1.07 ± 0.17	1.01 ± 0.24	1.03 ± 0.13	1.04 ± 0.26
D1	1.04 ± 0.14	1.01 ± 0.12	1.00 ± 0.09	-	-	1.02 ± 0.04
D2	1.01 ± 0.07	0.99 ± 0.17	0.99 ± 0.15	-	-	1.00 ± 0.12
D3	1.02 ± 0.04	1.02 ± 0.21	1.03 ± 0.13	-	-	1.02 ± 0.07
D4	1.00 ± 0.06	1.01 ± 0.13	1.01 ± 0.04	-	-	1.01 ± 0.07
D5	0.94 ± 0.15	1.06 ± 0.18	1.03 ± 0.08	-	-	1.01 ± 0.11
all (avg.)	1.01 ± 0.09	1.02 ± 0.12	1.02 ± 0.04	1.01 ± 0.24	1.03 ± 0.13	1.02 ± 0.09

All values are close to one and the differences between them are in the scale of numerical variations of the SMC algorithm. Hence, the investigation of the validation metric does not show any preference of a particular model.

AIMS Press