A calibration and uncertainty quantification analysis of classical, fractional and multiscale logistic models of tumour growth

Nikolaos M. Dimitriou[a], Ece Demirag[b], Katerina Strati[b], Georgios D. Mitsis[a,*]

[a] *Department of Bioengineering, McGill University, Montreal, H3A 0E9, QC, Canada*
[b] *Department of Biological Sciences, University of Cyprus, Nicosia, 2109, Cyprus*

Abstract

The validation of mathematical models of tumour growth is typically hampered by a lack of sufficient experimental data, resulting in qualitative rather than quantitative studies. Recent approaches to this problem have attempted to extract information about tumour growth by integrating multiscale experimental measurements, such as longitudinal cell counts and gene expression data. In this study, we investigated the fitness of several mathematical models of tumour growth, including classical logistic, fractional and novel multiscale models, with respect to measurements of *in-vitro* tumour growth in the presence and absence of therapy and the expression profiles of genes associated with changes in chemosensitivity. State-of-the-art Bayesian inference, likelihood maximisation and uncertainty quantification techniques allowed a thorough evaluation of the model fitness. The results suggest that the classical single-cell population model (SCPM) was the best fit for the untreated and low-dose treatment conditions, while the multiscale model with a cell death rate symmetric with the expression profile of OCT4 (Sym-SCPM) showed the best fit with the high-dose treatment data. Further identifiability analysis showed that the multiscale model was both structurally and practically identifiable under the condition of known OCT4 expression profiles. Overall, this study demonstrates that model performance can be improved by incorporating multiscale measurements of tumour growth.

Keywords: Bayesian inference, likelihood optimization, uncertainty quantification, multiscale, fractional, logistic models, tumour growth

1. Introduction

Over the past few decades, mathematical models have provided significant insights into the understanding of tumour growth and its underlying mechanisms [1, 2, 3, 4, 5], as well as the optimization of treatment [6, 7, 8]. One of the goals of the emerging field of mathematical oncology is the translation of models from an experimental to a clinical setting, contributing to personalized therapy [6, 9, 10, 11, 12]. However, computational and experimental challenges have shifted from the qualitative and informative use of models to patient-specific predictions.

A major computational challenge is the selection of an appropriate mathematical model. Tumour growth is a complex phenomenon that can be modelled at multiple levels. At the macroscopic level, there are several models that can provide a longitudinal description of tumour growth, including exponential-linear, Gompertz, Verhulst or logistic, and von Bertalanffy models [13] and, more recently, the expansion of these models to other cell populations, such as drug sensitive and resistant cells [7]. Longitudinal measurements of tumour growth have been used to calibrate mathematical models. Depending on the cancer type, case and stage, some models may be more appropriate than others. Recent studies have focused on several aspects

*Corresponding author
Email addresses: nikolaos.dimitriou@mail.mcgill.ca (Nikolaos M. Dimitriou), georgios.mitsis@mcgill.ca (Georgios D. Mitsis)

that make a model suitable for a cancer case. These include the accuracy of the model solution, the degree to which a model can describe the tumour behaviour, and the uncertainties associated with a prediction [14]. Similarly, studies have focused on evaluating of methods for estimating model parameters from experimental data [15, 16]. Although these techniques can provide a robust assessment of model predictions, this is only possible under the use of appropriate experimental data.

The sparsity of experimental measurements and single-level data extraction limit the ability of a model to accurately describe tumour growth. Recent studies have addressed this problem by incorporating multiscale data on tumour growth. Johnson et al. [17] integrated longitudinal bulk cell count measurements and single-cell transcriptomic data to estimate the ratio between drug-sensitive and drug-resistant cell populations at multiple time points, resulting in more accurate model predictions. Similarly, Smalley et al. [8] used single-cell mRNA analysis to define the transcriptional heterogeneity of melanoma and its response to BRAF inhibitors. They combined these data with a mathematical model for maintaining competition between drug-sensitive and drug-resistant states to control resistance.

Promising evidence for the use of gene expression data integrated with bulk cell count measurements comes from studies of bacterial growth [18]. Klumpp et al. [19] demonstrated specific patterns between gene expression levels and growth rates. Studies in *E. coli* [20, 21] have shown that the RNA/protein ratio is linearly correlated with the specific growth rate $\lambda = (\ln 2)/\text{doubling time}$. Although bacteria are much simpler organisms than cancer cells, we hypothesize that a similar phenomenon may occur in cancer, specifically between cell death rates and the expression of genes involved in altering chemosensitivity, such as OCT4 [22, 23, 24].

In this study, we provide a thorough investigation of classical, fractional and multiscale models of tumour growth. We calibrated these models against experimental measurements of cell counts from the C33A cell line. We used two calibration methods: likelihood optimization using the Improved Stochastic Ranking Evolution Strategy (ISRES) algorithm [25] and Bayesian inference using a Transitional Markov Chain Monte Carlo algorithm [26]. The multiscale models utilized information on both cell counts and gene expression data that may have contributed to the change in chemosensitivity. We examined two experimental conditions; no treatment, and treatment with 5FU. We found that the single-cell population model (SCPM) was the most appropriate model for non-treatment and low-treatment conditions, showing high agreement with the experiments and low uncertainty compared to the other models. For higher doses, we investigated four multiscale models derived from the classical SCPM and the fractional model (FM). We integrated these models with death rate parameters that took into account gene expression profiles based on a gene expression model. The multiscale model with a death rate symmetrical to the expression profile of OCT4 (Sym-SCPM) showed a significant improvement in agreement with the data and uncertainty. We also demonstrated the structural and practical identifiability of this model and that the knowledge of gene expression data is important for a practically identifiable model.

2. Methods

2.1. Experiments

Cell culture, 5FU treatment and counting: C33A cells were obtained from ATCC and cultured in MEM medium, 1% Penicillin/Streptomycin, 1% L-Glutamine and 10% FBS. The cell media were regularly monitored using PCR with mycoplasma-specific primers to ensure that the cultures remained mycoplasma-negative throughout the experiment. Fresh 5FU (Sigma Aldrich) was prepared weekly. 5×10^4 C33a cells were plated in 24-well plates for cell counting, and 1×10^6 cells were plated in 6-well plates for RNA extraction. One plate was prepared for each time point. Each biological replicate consisted of three technical replicates per treatment condition. Three days post-seeding, the cells were treated with 100, 10, 1, 0.1 and 0.01 μM 5FU diluted in media and incubated indicated time-points. Cell counts were performed using a hemocytometer, and dead cells were omitted from counting using trypan blue exclusion.

RNA Extraction and qRT-PCR RNA extraction was performed using Qiagen RNeasy Mini Kit (169014761) according to the manufacturer's instructions. RNA yield and quality (260 nm/280 nm ratio) were measured using Nanodrop™. RNA integrity was further validated on 1% agarose gel (28S and 18S rRNA bands). cDNA

was synthesized using iScript cDNA synthesis kit, according to the manufacturer's instructions. Following cDNA synthesis, qRT-PCR was performed using the KAPA SYBR® FAST qPCR Master Mix (2x) kit. PCR was performed on a CFX96 qPCR system (BioRad). The following primer sequences were used (Oct4: Fwd: GAAGGATGTGGTCCGAGTGT, Rev: GTGAAGTGAGGGCTCCCATA, GAPDH: Fwd: TGCACCAC-CAACTGCTTAGC, Rev: GGCATGGACTGTGGTCATGAG). Relative expression levels were determined using the DCT method.

2.2. Classical mathematical models

The *classical single-cell population model* (1) is a logistic growth model with three parameters; growth rate r_g, carrying capacity K, and effective cell death rate u_{eff}. The variable $N = N(t)$ represents the number of cancer cells over time. The effective cell death rate depends on the drug dose u and the death rate λ hence, $u_{eff} = \lambda u$.

$$\frac{\partial N}{\partial t} = r_g N(1 - \frac{N}{K}) - u_{eff} N \tag{1}$$

The *two-cell population model* considers both sensitive and resistant cancer cells. The model in (2)-(4) is similar to the single-cell population model found in (1) with the addition of the parameter α, which represents the spontaneous transition rate of sensitive cells to resistance. An additional constraint is that the effective death rate of sensitive cells, $u_{eff,s}$, should be greater than the effective death rate of resistant cells, $u_{eff,r}$ hence, $u_{eff,s} < u_{eff,r}$. Similarly, the effective death rates depend on the drug dose and the death rate parameter, and they are equal to $u_{eff,s} = \lambda_s u,\ u_{eff,r} = \lambda_r u$. Since the initial numbers of sensitive and resistant cells were unknown, we considered unknown initial conditions for the two populations. The total initial number of cells was determined from the experiments; hence, we chose to estimate the resistant and sensitive cells at $t = 0$ from (4), where T is the total number of cells.

$$\frac{\partial S}{\partial t} = r_s S(1 - \frac{S+R}{K}) - \alpha u S - u_{eff,s} S \tag{2}$$

$$\frac{\partial R}{\partial t} = r_r R(1 - \frac{S+R}{K}) + \alpha u R - u_{eff,r} R \tag{3}$$

$$T = S + R \tag{4}$$

Both models were numerically solved using the 5-order accurate Runge-Kutta Dormand-Prince method coded in C++ [27] and the 5-order accurate Tsitouras method in Julia [28, 29].

2.3. Fractional mathematical model

We examined the generalized fractional version of the single-cell population model as found in (1). The fractional logistic model defined in (5) takes into account the growth rate r_g, carrying capacity K, and the order of the derivative α. The order of the derivative takes the values $0 < \alpha \leq 1$ and is considered as unknown parameter during model calibration. The model was solved analytically using the formulation of Caputo's derivatives, the Mittag-Leffler function, and the Laplace transform (Appendix).

$$\frac{\partial^\alpha N}{\partial t^\alpha} = r_g N(1 - \frac{N}{K}) - u_{eff} N \tag{5}$$

2.4. Multiscale classical and fractional models

Previous studies have shown that gene expression and cell proliferation can be interdependent [18]. In this section, we provide a multiscale formulation of classical and fractional models by combining gene expression levels and cell counts in a single equation.

2.4.1. Correlation of cell counts with gene expression profiles

The OCT4, KLF4, SOX2, and NANOG genes appear to play an important role in the response of cancer cells to chemotherapy [22, 23, 24]. We used longitudinal data of the expression profiles of these genes obtained by real-time quantitative Polymerase Chain Reaction (RT-qPCR) to investigate possible patterns between the expression profiles and cell counts. To do this, we normalized the cell counts of the treatment data with respect to the control data and then estimated the Pearson correlation coefficient [30] between the normalized cell counts and the gene expression profiles.

2.4.2. Model integration with the effects of gene expression in cell death

For the multiscale expansion of the models presented above, we used the expression of genes that showed a clear common pattern with respect to cell counts. Specifically, we used gene expression profiles that were highly correlated with the corresponding cell counts. These profiles were then fitted to a simple asymptotic model defined as

$$f(t) = \frac{A}{t + 0.1} + B \tag{6}$$

using the least-squares method. The gene expression model was related to the cell death parameter of the corresponding classical single-population model, as well as to the order of the derivative for the fractional model.

Since the increase in the expression of genes such as OCT4 has been associated with chemoresistance, in turn a low cell death rate [22, 23, 24], we examined two possible ways of relating the cell death term with the gene expression model. One of the examined cell death parameter is the *inverted* gene expression model of the form

$$d_1(t) = \frac{1}{f(t)} \tag{7}$$

Second, we considered the cell death parameter to be *symmetric* across the y-axis and the shifted function with respect to the gene expression model, i.e.,

$$d_2(t) = \left| \frac{A}{t - (t_f + 0.1)} \right| + B \tag{8}$$

where t_f denotes the final time point of the experiment. We introduced these time-dependent cell death rates into the cell death term of the classical single cell population model,

$$\frac{\partial N}{\partial t} = r_g N(1 - \frac{N}{K}) - d(t) u_{eff} N \tag{9}$$

where $d(t) = d_1(t)$ or $d(t) = d_2(t)$. Additionally, we experimented coupling the fractional model with the cell death rates in the order of the derivative,

$$\frac{\partial^{\alpha(t)} N}{\partial t^{\alpha(t)}} = r_g N(1 - \frac{N}{K}) - u_{eff} N \tag{10}$$

where $\alpha(t) = d(t)/\max\{d(t)\}$ because of the constraint $0 < \alpha \leq 1$ imposed on the fractional model solution.

2.5. Calibration

In this section, we describe the two methods used to calibrate the model parameters with respect to the cell count data. The first is an optimization method to estimate the maximum likelihood between the experimental and simulated data. The second method is Bayesian inference to estimate the Posterior Distribution Functions (PDFs) of the model parameters. We continue with the calibration strategies that will help us understand the contribution of the model parameters to the simulation results, as well as the behaviour of the mathematical models with respect to various constraints applied in the model parameter space.

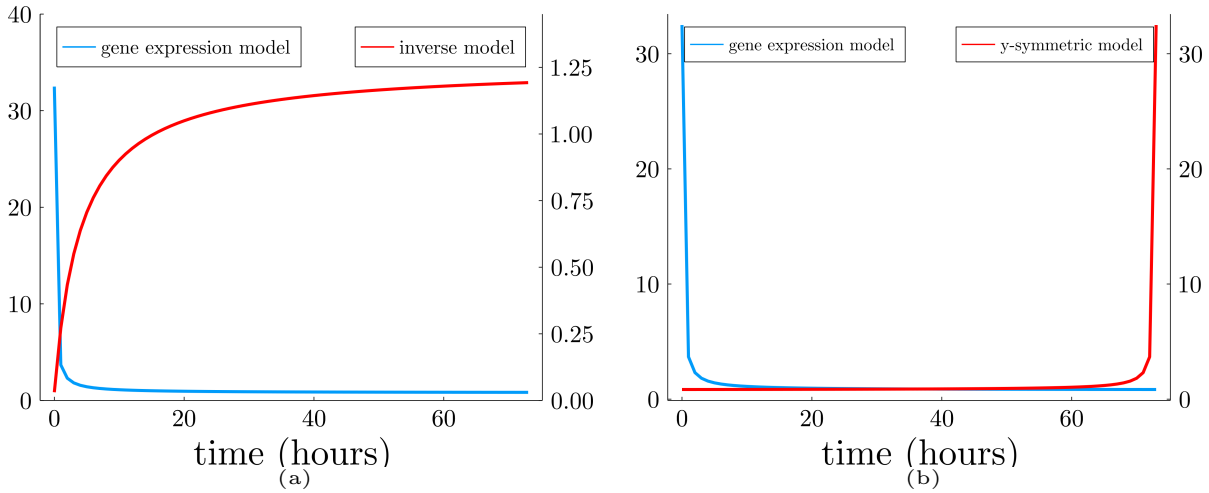


Figure 1: Gene model and two models (a) the inverted, (b) the y-symmetric and shifted model that represent the cell death term for the multiscale classical model, and when normalized, the derivative order of the multiscale fractional model.

Maximum Likelihood estimation. We used the Improved Stochastic Ranking Evolution Strategy (ISRES) algorithm [25] found in the NLopt package [31] in Julia [32]. Like in any Evolutionary Strategy algorithm, ISRES uses an array of candidate solutions that are updated based on *mutation* and *selection* rules. Mutation represents the step size between the parent solution and the offspring. In ISRES, mutation is controlled via a log-normal step size update and exponential smoothing. To avoid biases introduced by spherical symmetry assumptions [33], the ISRES algorithm performs a differential variation step using a Nelder-Mead-like update method [34, 35]. The selection rule is based on fitness ranking, and in this study, the fitness ranking is the log-likelihood. The likelihood was defined as

$$L(\theta; \mathbf{d}) = \sum_{k=1}^{K} \left(\prod_{i=1}^{N} \frac{1}{\sigma_{i,k}\sqrt{2\pi}} \exp\left(-\frac{(d_{i,k} - q_{i,k}(\theta))^2}{2\sigma_{i,k}^2} \right) \right) \tag{11}$$

where N is the total number of time points, K is the number of datasets, $\mathbf{d}$ the cell count, $d_{i,k}$, $q_{i,k}$ the cell count of the experimental sample and simulation result, respectively, at the time point i and dataset k, and $\sigma_{i,k}$ is the variance of the distribution of the likelihood obtained from the experimental error. In this study, $\sigma_{i,k}$ is the standard deviation of the three measurements for each time point for all the datasets. In this study, the number of candidate solutions at each step was $20 \times (n+1)$, where n is the number of model parameters.

Bayesian inference. We used the PDFs of the model parameters and the counts from the cell culture data, $\mathscr{D}$, to estimate the most probable parameter values using Bayes' rule

$$\mathbb{P}(\theta|\mathscr{D}, M) \propto \mathbb{P}(\mathscr{D}|\theta, M)\mathbb{P}(\theta) \tag{12}$$

where $\mathbb{P}(\theta|\mathscr{D}, M)$ is the posterior PDF of the model parameters θ given the observed data $\mathscr{D}$ and the model M, $\mathbb{P}(\mathscr{D}|\theta, M)$ is the likelihood of the observed data $\mathscr{D}$ given the model M and parameters θ, and $\mathbb{P}(\theta)$ is the prior PDF. We assumed uninformative, uniform distributions for the model parameter prior PDFs. The experimental data consisted of three replicates, each containing samples from 10 time points. The datasets were then combined to estimate the likelihood (Eq. (11)).

We used a Transitional Markov Chain Monte Carlo (TMCMC) algorithm implemented in the Π4U package [26]. The TMCMC algorithm iteratively constructs series of intermediate posterior PDFs

$$\mathbb{P}_j(\theta|\mathscr{D}, M) \propto \mathbb{P}(\mathscr{D}|\theta, M)^{\rho_j}\mathbb{P}(\theta) \tag{13}$$

Table 1: Prior distributions of the model parameters. The followed calibration strategy assumes that the growth rates and carrying capacities are not affected by the treatment. Thus, they can be estimated from the control data. The U notation denotes uniform distribution, and the $U[CI_{95}]$ denotes uniform distribution with boundaries the 95% credible interval of parameter posterior distribution found from the control studies.

Model	Priors								
	r_g	r_s	r_r	K	α	u_{eff}	$u_{eff,s}$	$u_{eff,r}$	R_0
SCPM-C	$U[0, 0.15]$	-	-	$U[3 \cdot 10^5, 5 \cdot 10^5]$	-	-	-	-	-
SCPM-T	$U[\mathrm{CI}_{95}]$	-	-	$U[\mathrm{CI}_{95}]$	-	$U[0, 0.15]$	-	-	-
TCPM-C	-	$U[0.05, 0.15]$	$U[0.01, 0.05]$	$U[3 \cdot 10^5, 5 \cdot 10^5]$	-	-	-	-	$U[0, 5 \cdot 10^5]$
TCPM-T	-	$U[\mathrm{CI}_{95}]$	$U[\mathrm{CI}_{95}]$	$U[\mathrm{CI}_{95}]$	$U[0, 0.2]$	-	$U[0.0011, 0.1]$	$U[0, 0.001]$	$U[\mathrm{CI}_{95}]$
FM-C	$U[0, 0.15]$	-	-	$U[3 \cdot 10^5, 5 \cdot 10^5]$	$U(0, 1]$	-	-	-	-
FM-T	$U[\mathrm{CI}_{95}]$	-	-	$U[\mathrm{CI}_{95}]$	$U(0, 1]$	$U[0, 0.15]$	-	-	-
Inv/Sym-SCPM-C	$U[0, 0.15]$	-	-	$U[3 \cdot 10^5, 5 \cdot 10^5]$	-	-	-	-	-
Inv/Sym-SCPM-T	$U[\mathrm{CI}_{95}]$	-	-	$U[\mathrm{CI}_{95}]$	-	$U[0, 0.15]$	-	-	-
Inv/Sym-FM-C	$U[0, 0.15]$	-	-	$U[3 \cdot 10^5, 5 \cdot 10^5]$	$U(0, 1]$	-	-	-	-
Inv/Sym-FM-T	$U[\mathrm{CI}_{95}]$	-	-	$U[\mathrm{CI}_{95}]$	$U(0, 1]$	$U[0, 0.15]$	-	-	-

where $j = 0, ..., m$ is the Monte Carlo time series index (generation index), ρ_j controls the transition between generations, and $0 < \rho_0 < \rho_1 < \cdots < \rho_m = 1$. The TMCMC method can use a large number of parallel chains that are evaluated in each Monte Carlo step to produce a result that is close to the true posterior PDF.

Calibration strategies. The increasing complexity of the selected models allowed us to follow an incremental calibration strategy. Specifically, the calibration of the single-population model of (1) without treatment, i.e. $u_{eff} = 0$, allowed us to estimate the growth rate, r_g, and the carrying capacity, K. For the treatment datasets, instead of considering these two parameters as completely unknown, we used a uniform prior bounded by the 95% credible intervals (CI), calculated using the high-density interval method [36], of the posterior distributions obtained from the calibration on the control datasets. This process was repeated for all the models examined, as shown in Table 1.

To assess the quality of the fits and the agreement between the simulation and the experimental data, we calculated the Concordance Correlation Coefficient (CCC) [37], defined as

$$\mathrm{CCC} = \frac{2\rho\sigma_d\sigma_q}{(\mu_d - \mu_q)^2 + \sigma_d^2 + \sigma_q^2} \tag{14}$$

where n is the number of time points in each dataset, ρ is the Pearson correlation coefficient [30], and σ_d^2, σ_q^2 are the variances of the *in-vitro* and *in silico* data, respectively.

2.6. Uncertainty analysis

The use of Bayesian inference and the TMCMC algorithm allows the posterior of the model parameters to be estimated. To assess the effect of parameter uncertainty on the resulting cell counts, we performed 5000 simulations of each model with random sets of parameters drawn from their corresponding posterior distributions. The parameters were selected from a typically unknown distribution using the inverse transform sampling technique [38].

2.7. Identifiability of model parameters

Model identification falls into two categories: structural and practical. Structural identifiability examines the identification of model parameters by assuming an infinite number of error-free data points. On the other hand, practical identifiability examines the possibility of uniquely identifying of model parameters given finite and noisy sets of observations. This section examines the structural and practical identifiability of the best candidate model. For structural identifiability, we followed a methodology similar to that of [39, 7, 17], which includes the use of Lie derivatives, and we demonstrate this in the results for the best candidate model.

Simpson et al. [40] compared two methods to test the practical identifiability of a biophysical model using Bayesian inference [41] and the profile likelihood method [42]. The two methods gave similar results. As Bayesian inference allows for the investigation of posterior distributions, we used it to examine the

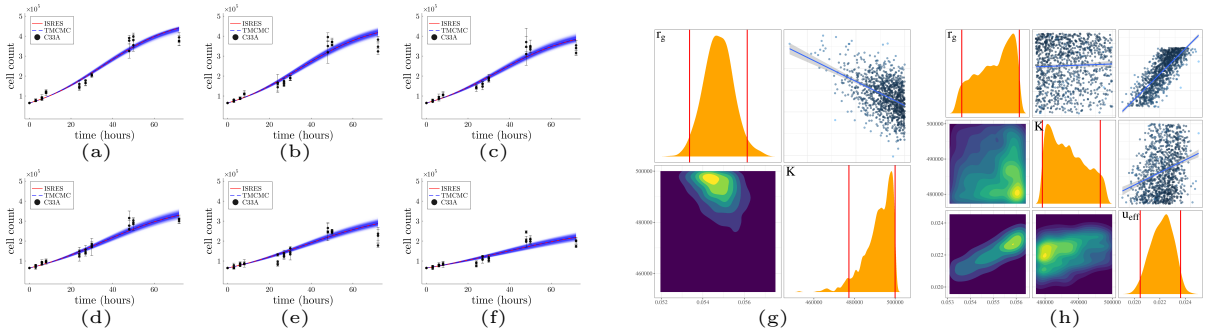


Figure 2: Classical single cell population model calibration against (a) control, (b) 0.01 μM, (c) 0.1 μM, (d) 1 μM, (e) 10 μM, and (f) 100 μM of 5FU datasets. Posterior distributions of the model parameters for (g) control, (h) 100 μM of 5FU datasets. The prior boundaries of the r_g and K parameters of the treatment model are extracted from the 95% CI of the posterior distributions of the corresponding parameters of the control model. The red lines denote the 95% CI of the posterior distributions.

practical identifiability of the best candidate model. Although the calibration has already provided the posterior distributions, we cannot infer the identifiability of the models because of the narrowing of the prior distribution boundaries resulting from the calibration strategy, which defines the prior boundaries from the 95% CI of the posterior distributions of the model calibrated with control data. Therefore, we repeated the inference of the model parameters using the prior boundaries originally used for the calibration with the control data.

3. Results

3.1. Classical single cell population model calibration

The calibration results of the single-cell population model for the control and treatment datasets are shown in Fig. 2. The posterior PDF manifold of the model parameters for the control data in Fig. 2g shows low uncertainty in the growth rate, r_g, and a skewed posterior PDF towards the upper limit of the carrying capacity, K. The simulation results for the control are shown in Fig. 2a and are in good agreement with the experimental data, and the uncertainty in the simulation output remains low. The good agreement between the experimental and simulation data can be confirmed by the CCC results in Table 3. Similar behaviour was observed for calibration with the treatment data (Fig. 2b-2f), where the parameters r_g and K were sampled around the posterior 95% CI obtained from the calibration of the control datasets. The posterior manifold for the highest dose calibration is shown in Fig. 2h and shows uniform distributions within the bounds obtained using the 95% CI of the control posterior PDF. The treatment parameter u_{eff} exhibits low uncertainty compared to the prior boundaries (Table 1).

3.2. Classical two-cell population model calibration

Although the presence of resistance is not clear in these datasets, we examined the fitness of the two-cell population model, which takes into account sensitive and resistant cells. The calibration results for the two-cell population model are shown in Fig. 3. The overall fitness result of the model is comparable to that of the single-cell population model. The posterior PDF manifold of the model parameters for the control data in Fig. 3g shows increased uncertainty compared to the single-population model, particularly in the resistant growth rate parameter and the initial number of resistant cells. This increased uncertainty is reflected in the total number of cancer cells, as shown in Fig. 3a-3f. The uncertainty in the initial number of resistant cells remained high at the highest dose, as shown in Fig. 2h.

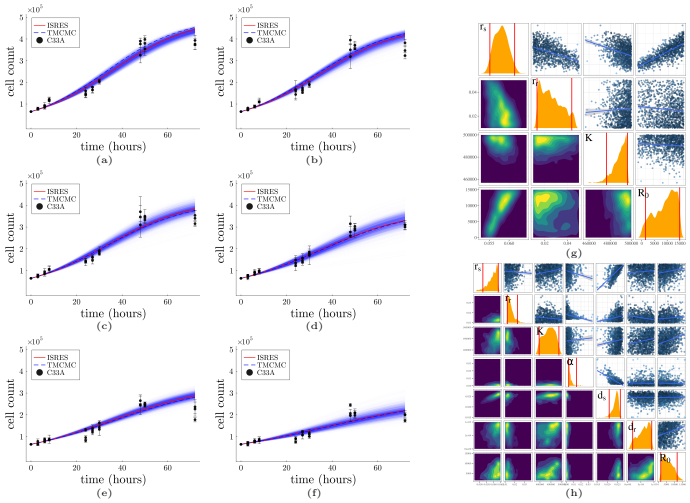


Figure 3: classical two cell population model calibration against (a) control, (b) 0.01 μM, (c) 0.1 μM, (d) 1 μM, (e) 10 μM, and (f) 100 μM of 5FU datasets. Posterior distributions of the model parameters for (g) control, (h) 100 μM of 5FU datasets. The prior boundaries of the r_s, r_r and K parameters of the treatment model are extracted from the 95% CI of the posterior distributions of the corresponding parameters of the control model. The red lines denote the 95% CI of the posterior distributions.

3.3. Fractional model calibration

We generalized the classical single-cell population model to by considering the derivative order as unknown $\alpha \in (0, 1]$. The fractional model calibration had a similar performance to that of the two-cell population model, as shown in Fig. 4. The introduction of the derivative order as a variable resulted in increased uncertainty compared to the classical single-population model. In addition, the derivative order tended to converge to 1, suggesting that the single-cell population ODE model is more appropriate than the fractional model.

3.4. Multiscale model calibration

We performed a multiscale integration of macroscopic single-cell population models with the effects of gene expression. This step involved searching for common patterns between cell death and gene expression changes in OCT4, SOX2, KLF4, and NANOG. To do this, we normalized the cell counts of treatment data with respect to the counts in the non-treatment data and calculated the Pearson correlation coefficient between the normalized cell counts and the gene expression profiles. The correlation of the expression profile of OCT4 with the corresponding cell counts increased with respect to the 5FU dose and showed a strong correlation at 10 and 100 μM doses of 5FU, as shown in Table 2 and Fig. 5a, 5b. We used the expression profiles obtained at 10 and 100 μM to generate two candidate models of gene expression. The expression of OCT4 decreased with time, as shown in Fig. 5c, 5d. Fitting the gene expression model of (6) gave $A = 3.17 \pm 0.49$, $B = 0.795 \pm 0.076$ for 100 μM 5FU, and $A = 1.04 \pm 0.27$, $B = 0.811 \pm 0.042$ for 10 μM 5FU.

Since the decreased in the expression of OCT4 is associated with increased chemosensitivity [22, 23, 24], we examined two functions related to cell death. The first function takes the inverse of the gene expression model, and relates it to the cell death rate (7). The second function takes the y-axis symmetric and time-shifted gene expression model (8), and relates it to the cell death rate. We then combined the resulting time-varying cell death rates with two macroscopic single-cell population models (classical and fractional). For the classical model, time-varying cell death rates were introduced into the cell death term, resulting in two multiscale classical models (9) (namely Inv-SCPM and Sym-SCPM for coupling with the inverse and the y-symmetric gene expression models, respectively). For the fractional model, we experimented with introducing the time-varying cell death rates as a derivative order after normalizing them, resulting in two multiscale fractional models (10) (namely Inv-FM and Sym-FM for coupling with the inverse and the y-symmetric gene expression models, respectively).

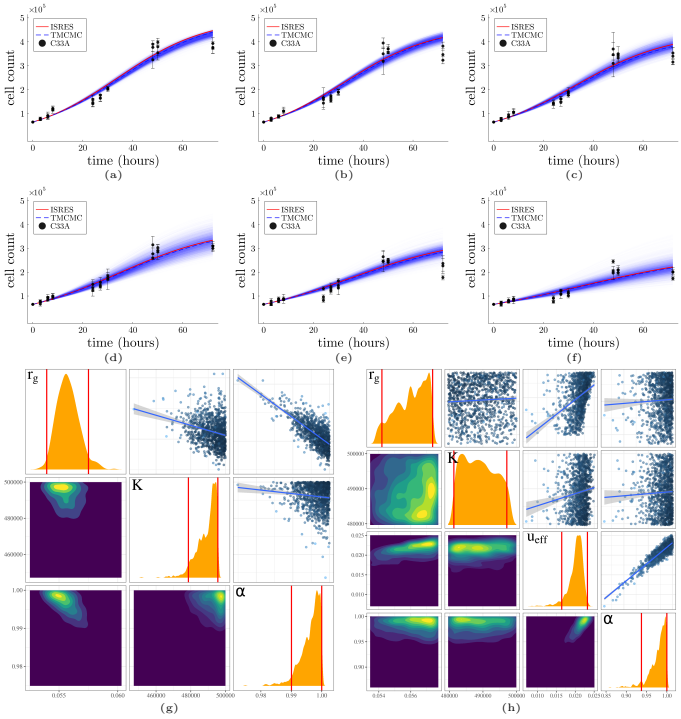


Figure 4: Fractional single cell population model calibration against (a) control, (b) 0.01 μM, (c) 0.1 μM, (d) 1 μM, (e) 10 μM, and (f) 100 μM of 5FU datasets. Posterior distributions of the model parameters for (g) control, (h) 100μM of 5FU datasets. The prior boundaries of the r_g and K parameters of the treatment model are extracted from the 95% CI of the posterior distributions of the corresponding parameters of the control model.

Table 2: Pearson correlation coefficient between the normalized cell counts of the treatment data with respect to the controls, and the expression profiles of the genes OCT4, KLF4, SOX2, NANOG. The correlation of the normalized counts with respect to the expression of OCT4 is increasing with respect to the dose increment, indicating strong correlation on the two highest doses.

5FU (μM)	OCT4	KLF4	SOX2	NANOG
0.01	-0.11	0.20	0.12	-0.26
0.1	0.39	0.19	0.79	0.00
1	0.34	-0.57	0.04	-0.25
10	0.89	-0.07	0.68	0.03
100	0.90	-0.24	0.97	0.68

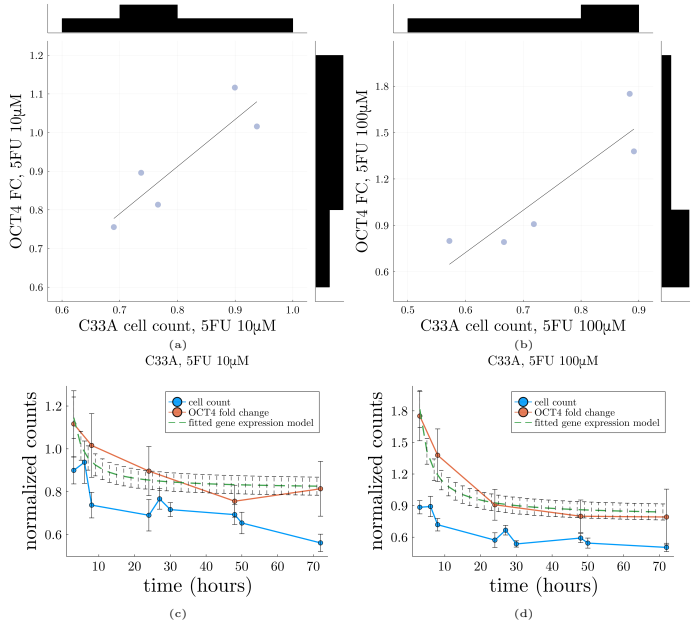


Figure 5: Correlation of OCT4 expression profiles with normalized cell counts for (a) 10μM and (b) 100μM of 5FU. Gene expression model fitting results for (c) 10μM and (d) 100μM of 5FU. The error bars denote the standard deviation of the two model outputs.

The calibration results for the multiscale models are shown in Fig. 6, 7. Overall, the multiscale SCPM performed better than the FM. The Sym-SCPM showed the best performance of all models, achieving a curvature that the single-scale models were unable to capture at the two highest doses during the last days of the experiment Fig. 8. The sym-SCPM trajectories are more plausible than the corresponding classical SCPM because we expect cell elimination after prolonged times of the cell cultures, while the classical SCPM predicts stabilization of the population. The agreement with the experimental data, shown in Table 3, confirms that the Sym-SCPM exhibits the best performance. In contrast, the Sym-FM showed poor agreement with the experiments. Inv-SCPM and Inv-FM have similar performances to macroscopic models. The prior distribution boundaries of the parameters r_g and K were taken from the 95% CI posterior PDFs of the corresponding macroscopic no-treatment models, since the multiscale expansion focuses on cell death. The uncertainty of the multiscale model output remained low and was comparable to that of the macroscopic models. The derivative order scaling parameter α shows increased uncertainty for both Inv-FM and Sym-FM, as shown in Fig. 7e and 7f, suggesting insensitivity to the derivative scaling.

3.5. Identifiability of the Sym-SCP model parameters

The multiscale Sym-SCP exhibited the best agreement with the experimentally measured cell counts. Here we investigate the identifiability of this model.

3.5.1. Structural identifiability

We demonstrate the structural identifiability of the Sym-SCP model. We start by non-dimensionalizing the model, by setting $n = N/K$, the model becomes

$$\frac{\partial n}{\partial t} = n r_g (1-n) - u_{eff} \left(\frac{A}{|t - t'_f|} + B \right) n \tag{15}$$

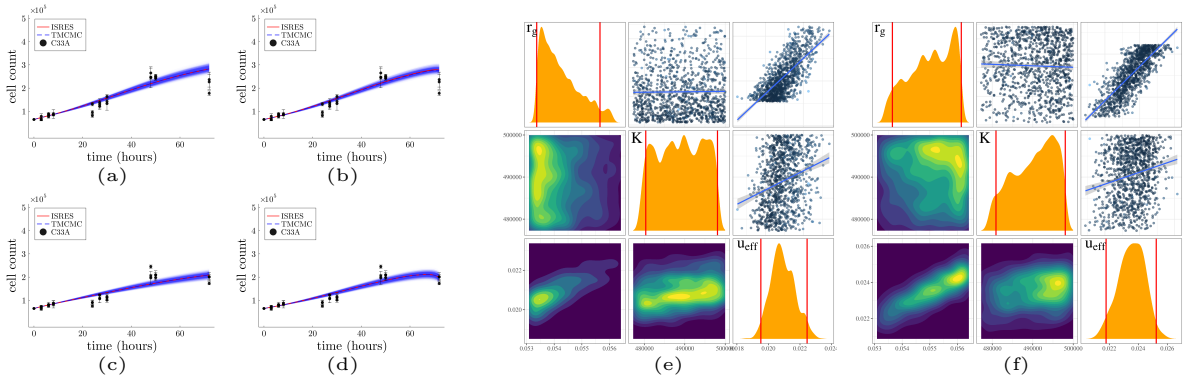


Figure 6: Multiscale classical single cell population model calibration against (a), (b) 10μM, (c), (d) 100μM of 5FU datasets. The model is integrated with (Left column) inverted asymptotic gene expression model, and (Right column) the y-symmetric gene expression model. Parameter posterior distributions of the macroscopic model integrated with (g) the inverted gene expression model, (h) the y-symmetric gene expression model for the 100μM of 5FU datasets.

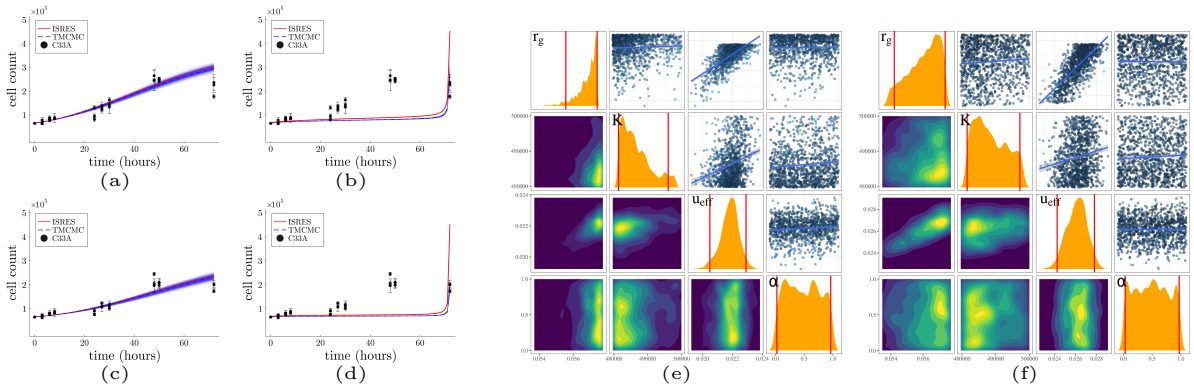


Figure 7: Multiscale fractional single cell population model calibration against (a), (b) 10 μM, (c), (d) 100 μM of 5FU datasets. The model is integrated with (Left column) inverted asymptotic gene expression model, and (Right column) the y-symmetric gene expression model. Parameter posterior distributions of the macroscopic model integrated with (g) the inverted gene expression model, (h) the y-symmetric gene expression model for the 100μM of 5FU datasets.

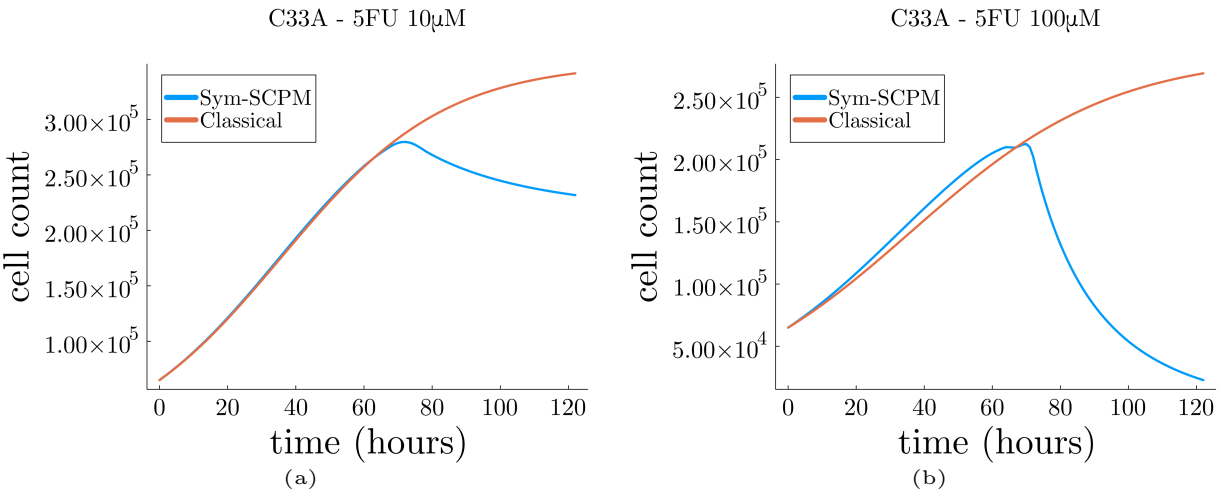


Figure 8: Comparison of dynamics of the classical and the sym-SCPM with estimated parameters for (a) 10 μM and (b) 100 μM dose of 5FU. The sym-SCPM trajectory is more plausible as we observed cell elimination in the experiments after prolonged times. On the other side, the classical SCPM predicts stabilisation of the cell population.

Table 3: Concordance correlation coefficient between experimental data and models The abbreviated names SCPM, TCPM, FM stand for single cell population model, two cell population model, and fractional model, respectively. The Inv and Sym abbreviations denote the integration of the corresponding model with the inverse gene model and y-symmetric model, respectively.

Model	CCC					
	Control	0.01μM	0.1μM	1μM	10μM	100μM
SCPM	0.97 ± 0.01	0.96 ± 0.04	0.96 ± 0.02	0.97 ± 0.02	0.89 ± 0.08	0.88 ± 0.05
TCPM	0.97 ± 0.01	0.96 ± 0.03	0.96 ± 0.02	0.97 ± 0.02	0.89 ± 0.07	0.88 ± 0.05
FM	0.97 ± 0.01	0.96 ± 0.04	0.96 ± 0.02	0.97 ± 0.02	0.89 ± 0.08	0.89 ± 0.05
Inv-SCPM	-	-	-	-	0.89 ± 0.08	0.87 ± 0.05
Sym-SCPM	-	-	-	-	0.90 ± 0.07	0.91 ± 0.04
Inv-FM	-	-	-	-	0.89 ± 0.08	0.88 ± 0.05
Sym-FM	-	-	-	-	0.14 ± 0.03	0.06 ± 0.01

where $t'_f = t_f + 0.1$. For simplicity, we set $r_g = r$, and $u_{eff} = u$. Since, t_f is the final time of the experiment, and $0 \leq t \leq t'_f$, we can write

$$\frac{\partial n}{\partial t} = nr(1-n) - u\left(\frac{A}{t'_f - t} + B\right)n \tag{16}$$

We assume that all parameters are non-negative. At $t = 0$, we have $n(0) = n_0$, and $0 < n_0 < 1$ due to the non-dimensionalization presented above. The model can be written in the input-output form

$$\dot{x}(t) = f(x(t)) + u(t)g(x(t)),\ x(0) = x_0 \tag{17}$$

where f and g are

$$f(x) = r(1-x)x \tag{18}$$

$$g(x) = \left(\frac{A}{t'_f - t} + B\right)x \tag{19}$$

and $x(t) = n(t)$. We can write

$$y(t) = h(x(t)) = x(t) \tag{20}$$

and introduce p to be the vector of the parameters to be identified, i.e.

$$p = (n_0, r, A, B) \tag{21}$$

A system as in (17) is uniquely structurally identifiable if the map

$$(p, u(t)) \mapsto (x(t,p), u(t)) \tag{22}$$

is injective almost everywhere [43, 44].

Assuming *perfect* (continuous and noise-free) input-output data given in the form of y and its derivatives at any time interval, we can write the measurements in the form of

$$y(0) = h(x(0)) \tag{23}$$

$$\dot{y}(0) = \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} h(x(t)) \tag{24}$$

and relate their values to the unknown parameters found in (21). The system is identifiable if there exist inputs, $u(t)$, such that system of (23), (24) may be solved for p. The right hand sides of (23), (24) can be

computed in terms of the Lie derivatives of f (18) and g (19) found in (17). Recall that Lie differentiation of a $\mathcal{C}^\omega$ function H by a $\mathcal{C}^\omega$ vector field X, $L_X H$, is

$$L_X H(x) := \nabla H(x) \cdot X(x) \tag{25}$$

The iterated Lie derivatives are well defined, and can be written as $L_Y L_X H = L_Y(L_X H)$, and $L_X^2 H = L_X(L_X H)$.

We can define the observable quantities as the zero-time derivatives of the output $y = h(x)$ as follows

$$Y(x_0, U) = \left.\frac{\mathrm{d}^k}{\mathrm{d}t^k}\right|_{t=0} h(x(t)) \tag{26}$$

where $U \in \mathbb{R}^k$ is the control value, $u(t)$, and its derivatives at $t = 0$; $U = (u(0), u'(0), ..., u^{(k-1)}(0))$. The observation space is then defined as the span of the $Y(x_0, U)$ elements

$$F_1 := \mathrm{span}_{\mathbb{R}}\{Y(x_0, U) | U \in \mathbb{R}^k, k \geq 0 \tag{27}$$

Conversely, the span of iterated Lie derivatives with respect to the output h and the vector fields f, g is

$$F_2 := \mathrm{span}_{\mathbb{R}}\{L_{i1}...L_{ik}h(x_0) | (i_1, ..., i_k) \in \{g, f\}^k, k \geq 0\} \tag{28}$$

By [45, 7], $F_1 = F_2$, hence identifiability may be formulated as the reconstruction of parameter in p from elements in F_2. Thus, p is identifiable if the map

$$p \mapsto (L_{i1}...L_{ik}h(x_0) | (i_1, ..., i_k) \in \{g, f\}^k, k \geq 0) \tag{29}$$

is "1-1".

Computing the Lie derivatives we can determine the parameters of p

$$n_0 = h(x_0) \tag{30}$$

$$L_f h = r(1 - n_0)n_0 \implies r = \frac{L_f h}{(1 - n_0)n_0} \tag{31}$$

$$L_g h = \left(\frac{A}{t'_f} + B\right) n_0 \implies B = \frac{L_g h}{n_0} - \frac{A}{t'_f} \tag{32}$$

$$L_g L_f h = r(1 - 2n_0)\left(\frac{A}{t'_f} + B\right) n_0 \implies A = \left(\frac{L_g L_f h}{n_0 r(1 - 2n_0)} - B\right) t'_f \tag{33}$$

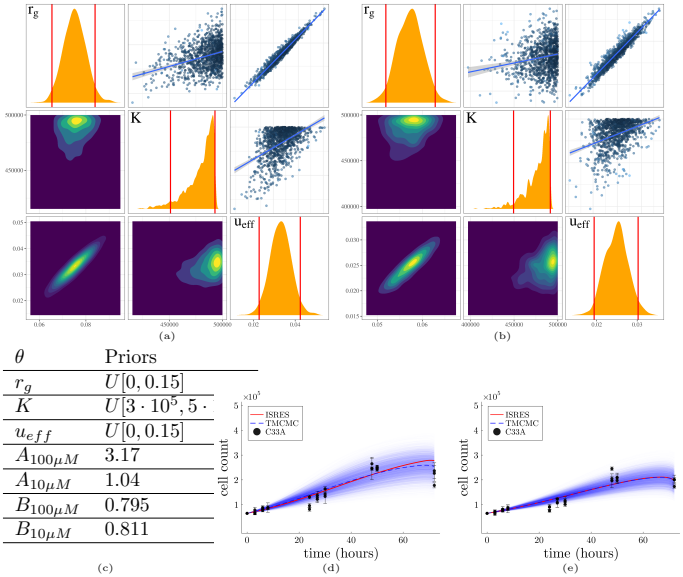
As a result, all parameters of (16) are identifiable.

3.5.2. Practical identifiability

We expanded the prior distribution boundaries of the Sym-SCP model parameters to the values used in the SCP model calibration with the control data. We performed Bayesian inference of the Sym-SCPM using the treatment datasets of 10 and 100 µM FU. The resulting posterior distributions are presented in Fig. 9a and 9b and show unimodal distributions that are significantly narrower than the prior distributions in Fig. 9c. The uncertainty in the output, Fig. 9d and 9e, was increased compared to Fig. 6c and 6d; however, this uncertainty follows the trend of the data and is found within the overall range of the experimental cell counts. Thus, the Sym-SCP model is considered to be practically identifiable for the given dataset.

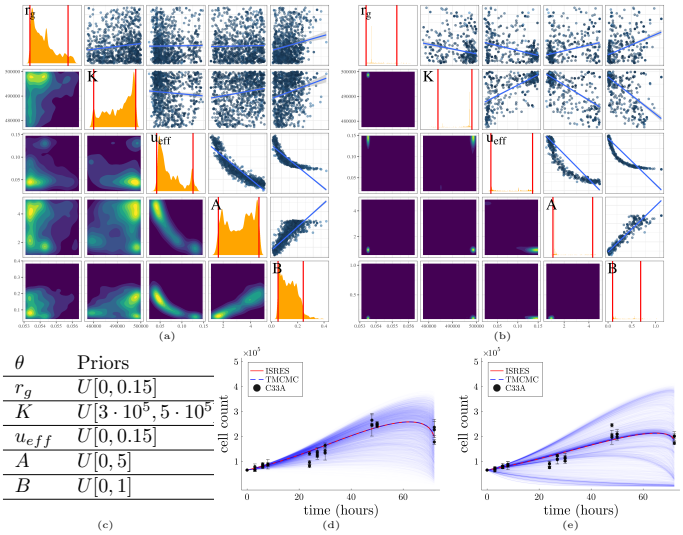
3.5.3. The knowledge of gene expression estimates is necessary for a practically identifiable model

The Sym-SCP model was practically identifiable. As shown in Fig. 9c, the parameters A and B were estimated from the gene expression data and fixed for the identifiability analysis. In this section, we considered the scenario of absence of the gene expression data. Therefore, we considered the values of A and B to be unknown, and repeated the same process as in the previous section with the prior parameter distributions of Fig. 10c. The resulting posterior distributions, found in Fig. 10a, 10b show increased uncertainty, a result that is reflected to the model output, Fig. 10d, 10e, making the model practically unidentifiable. Thus, the information obtained from the gene expression data for the calibration of the parameters A and B is necessary for a practically identifiable model.



θ	Priors
r_g	$U[0, 0.15]$
K	$U[3 \cdot 10^5, 5 \cdot$
u_{eff}	$U[0, 0.15]$
$A_{100\mu M}$	3.17
$A_{10\mu M}$	1.04
$B_{100\mu M}$	0.795
$B_{10\mu M}$	0.811

Figure 9: Practical identifiability analysis of the Sym-SCPM with known parameters A and B. (a), (b) Posterior manifold for the calibration with the 10μM, and 100μM datasets, respectively. (c) Table of the model parameters and their corresponding prior distributions. The A and B are calibrated from the gene expression data and they are fixed. (d), (e) *In-vitro* and *in-silico* cell counts and their corresponding uncertainties.



θ	Priors
r_g	$U[0, 0.15]$
K	$U[3 \cdot 10^5, 5 \cdot 10^5]$
u_{eff}	$U[0, 0.15]$
A	$U[0, 5]$
B	$U[0, 1]$

Figure 10: Practical identifiability analysis of the Sym-SCPM considering parameters A and B unknown. (a), (b) Posterior manifold for the calibration with the 10μM, and 100μM datasets, respectively. (c) Table of the model parameters and their corresponding prior distributions. (d), (e) *In-vitro* and *in silico* cell counts and their corresponding uncertainties. The unknown parameters A and B increase dramatically the uncertainty of the model output. Thus, the integration of the macroscopic cell count measurements with the gene expression data is necessary for a practically identifiable model.

4. Discussion

The analysis yielded important insights into the agreement of the calibrated models with the experimental data and their uncertainty, both in terms of parameters and final output. In this section, we discuss the fitting performance of the examined models, the performance of the two calibration methods, and the identifiability of these models.

4.1. Fitness and Interpretation of the model behaviour

In this study, we investigated the behaviour of 7 mathematical models of tumour growth in relation to longitudinal *in-vitro* measurements of tumour growth and the expression of genes associated with changes in chemosensitivity. The macroscopic SCPM exhibited good agreement with the data and a low uncertainty in the output. The agreement with the experiments decreased for the two highest doses of 5FU, with the model showing a more linear than sigmoid behaviour. This behaviour can be explained by the carrying capacity compared to the maximum number of cells under these conditions. The carrying capacity was significantly higher than the cell counts and therefore it was not possible to create a plateau at later time points. Overall, the SCPM exhibited high agreement with the experiments, especially for the no-treatment and low-dose conditions. To improve accuracy at the highest doses, we expanded our analysis to examine more complex models.

We then increased the complexity, leading to a TCPM that takes into account sensitive and resistant cell populations, although we had no evidence for the presence of resistance in this dataset. The resulting agreement with the experiments did not improve, and the uncertainty of the output increased slightly. In addition, the initial number of resistant cells remained highly uncertain in both the control and treatment data, probably due to the low growth rate of the resistant cells. The results showed that TCPM did not improve the agreement or uncertainty.

Fractional calculus is a relatively new area of mathematics that expands the ordinary integer order of derivatives to non-integer orders. Fractional derivative models can take into account non-local interactions in space and dynamic memory in the time domain [46]. We expanded the SCPM to a non-integer order of the time derivative and considered the derivative order as unknown, leading to the FM. The calibration of the FM yielded a derivative order close to 1, suggesting that static fractional derivatives do not contribute to the improvement of the SCPM. In fact, the FM exhibited similar performance to the SCPM.

Multiscale expansion of the SCPM and FM was achieved by constructing an expression model for OCT4 and relating it to cell death rates. The two cell death rate functions were coupled with the cell death term of the SCPM and the derivative order of the FM, resulting in Inv-SCPM, Sym-SCPM, Inv-FM, and Sym-FM. The Sym-SCPM showed the best agreement with the experimental data (10 μM and 100 μM 5FU) and low uncertainty in the output, resulting in a better trajectory than the other models. The increase in the agreement can be explained by the fact that the cell death rate performed a steep increase during the last days of the experiment, which in turn curved the cell counts towards lower values. This hypothesis suggests that there is a threshold level of OCT4 expression that triggers or suppresses cell death (gain-of-function and loss-of-function, respectively). This suggests that OCT4 activates or suppresses mechanisms related to chemoresistance as a binary function rather than a transient function of time.

In contrast, Sym-FM exhibited the worst performance compared with the other models. This is justified by the fact that the derivative order remains low most of the time, delaying the model's entry into the exponential phase. The sudden increase in the derivative order towards the end of the simulation leads to a rapid entry into the exponential phase, which is inconsistent with the experimental observations. Unexpectedly, the scaling parameter of the derivative order α remains highly insensitive. This result suggests that the longitudinal behaviour of the derivative order is independent of its magnitude. Thus, the dynamic behaviour of the derivative order seems to be more important than its amplitude.

4.2. Evidence for the use of multiscale models and hypotheses

The multiscale expansion of the SCPM to the Sym-SCPM exhibited higher agreement for the two highest doses than the other models. However, the design of this multiscale model is not arbitrary. In fact, it was based on observations of growth curves, gene expression estimates, and recent findings from the literature.

From the growth curves, we observed a sigmoid trend in the presence and absence of therapy, even though the treatment was initiated at $t = 0$ h. Instead, we would expect that the cell counts to stabilize or decrease in the presence of the treatment. This observation led us to hypothesize that the cell death rate was not static and may sow a delayed increase after treatment initiation. At the same time, we observed that the expression of OCT4 – a potential regulator of chemosensitivity in cells [22, 23, 24] – decreased over time. In addition, studies in bacteria [18, 19] have shown that there is a dependence between gene expression rates and growth rates. We therefore hypothesized that the cell death rate is linked to the expression of OCT4 and investigated the two possible transient cell death rates.

4.3. Structural and practical identifiability of the Sym-SCPM

Structural and practical identifiability tests were carried out to further investigate the identifiability of the model parameters. The structural identifiability analysis was performed using the methodology presented in [39]. The Sym-SCP model, which included both the macroscopic parameters and the parameters included in transient cell death, was found to be structurally identifiable. The practical identifiability analysis was performed using Bayesian inference, using the same prior distributions as in the SCP model calibrated with the control data. The model was practically identifiable provided that the parameters of the transient cell death term were known. These two parameters were estimated from OCT4 expression data. The model did not pass the identifiability test for completely unknown parameters, suggesting that information on gene expression levels is crucial.

4.4. Monte Carlo and Global optimization algorithms

We used two calibration methods in this study; the likelihood optimization method using the ISRES algorithm, and Bayesian inference using the TMCMC algorithm. The ISRES algorithm is considered to be a global or semiglobal optimization method. It uses a population size of $20\times(n+1)$ candidate solutions, where n is the number of parameters. In contrast, TMCMC used 1024 candidate solutions for each generation. Overall, the two methods showed comparable performance in terms of finding the maximum likelihood, and TMCMC was able to find parameters with a higher likelihood than ISRES. Although the ISRES algorithm, and the likelihood optimization methods in general, are considerably faster than Bayesian inference and Monte Carlo methods, recent efforts to parallelize of the MCMC algorithm have significantly improved its efficiency. The TMCMC algorithm is inherently parallel, and its function does not require a burn-in period as in traditional MCMC methods. In our case, the TMCMC algorithm took about 10 minutes to run on a 4×Intel Core™ i7-6820HQ CPUs running at 2.70GHz. Overall, Bayesian inference and the TMCMC algorithm provide information not only on the fitness of the model, but also on the uncertainty and identifiability. Together with the improved performance, it can be considered as an efficient all-in-one tool.

5. Conclusions

In this study, we performed extensive calibration and uncertainty quantification analysis of classical, fractional, and multiscale models of tumour growth using data obtained from cell cultures of the C33A cell line. The results suggested that the single-cell population model was the most appropriate model to describe of tumour under no-treatment and low dose 5FU treatment conditions. The multiscale expansion of this model by using data on the expression of the transcription factor OCT4 improved the agreement with the experimental observations for the two highest doses of 5FU, suggesting that Sym-SCPM is the most appropriate model to describe tumour growth under high-dose treatment conditions. Investigation of two potential mechanisms for the transient cell death rate indicated that the cell death process is more likely to function as a binary mechanism triggered after a threshold level of OCT4 expression rather than as a more gradual activation inversely related to the level of OCT4 expression. The two-cell population model and the fractional models did not significantly improve the fitness with the experiment. The multiscale expansion of the fractional model exhibited the worst performance. The sym-SCPM proved to be both structurally and practically identifiable allowing the model parameters to be unambiguously estimated. Overall, we examined

a large collection of mathematical models and provided novel information on the relationship between gene expression and cell death, using powerful calibration and uncertainty quantification techniques.

References

[1] N. M. Dimitriou, S. Flores-Torres, J. M. Kinsella, G. D. Mitsis, Quantifying the Morphology and Mechanisms of Cancer Progression in 3D in-vitro environments: Integrating Experiments and Multiscale Models, IEEE Transactions on Biomedical Engineering (2022). doi:10.1109/TBME.2022.3216231.

[2] E. A. B. F. Lima, J. T. Oden, D. A. Hormuth, T. E. Yankeelov, R. C. Almeida, Selection, calibration, and validation of models of tumor growth, Mathematical Models and Methods in Applied Sciences 26 (12) (2016) 2341–2368. doi:10.1142/S021820251650055X.
URL https://doi.org/10.1142/S021820251650055X

[3] E. A. Lima, D. Faghihi, R. Philley, J. Yang, J. Virostko, C. M. Phillips, T. E. Yankeelov, Bayesian calibration of a stochastic, multiscale agent-based model for predicting in vitro tumor growth, PLoS Computational Biology 17 (11) (2021). doi:10.1371/journal.pcbi.1008845.

[4] K. A. Rejniak, A. R. A. Anderson, Hybrid models of tumor growth, Wiley Interdisciplinary Reviews: Systems Biology and Medicine 3 (1) (2011) 115–125. doi:10.1002/wsbm.102.
URL https://onlinelibrary.wiley.com/doi/abs/10.1002/wsbm.102

[5] N. M. Dimitriou, S. Flores-Torres, M. Kyriakidou, J. M. Kinsella, G. D. Mitsis, Cancer cell sedimentation in 3D cultures reveals active migration regulated by self-generated gradients and adhesion sites, bioRxiv (2023). doi:10.1101/2023.02.15.528731.
URL https://www.biorxiv.org/content/early/2023/02/16/2023.02.15.528731

[6] M. M. Hadjiandreou, G. D. Mitsis, Mathematical Modeling of Tumor Growth, Drug-Resistance, Toxicity, and Optimal Therapy Design, IEEE Transactions on Biomedical Engineering 61 (2) (2014) 415–425. doi:10.1109/TBME.2013.2280189.

[7] J. M. Greene, C. Sanchez-Tapia, E. D. Sontag, Mathematical details on a cancer resistance model, Frontiers in Bioengineering and Biotechnology 8 (2020) 501.

[8] I. Smalley, E. Kim, J. Li, P. Spence, C. J. Wyatt, Z. Eroglu, V. K. Sondak, J. L. Messina, N. A. Babacan, S. S. Maria-Engler, L. D. Armas, S. L. Williams, R. A. Gatenby, Y. A. Chen, A. R. A. Anderson, K. S. M. Smalley, Leveraging transcriptional dynamics to improve BRAF inhibitor responses in melanoma, EBioMedicine 48 (2019) 178–190. doi:https://doi.org/10.1016/j.ebiom.2019.09.023.
URL http://www.sciencedirect.com/science/article/pii/S2352396419306279

[9] M. A. R. Strobl, J. West, Y. Viossat, M. Damaghi, M. Robertson-Tessi, J. S. Brown, R. A. Gatenby, P. K. Maini, A. R. A. Anderson, Turnover Modulates the Need for a Cost of Resistance in Adaptive Therapy, Cancer Research 81 (4) (2021) 1135–1147. doi:10.1158/0008-5472.CAN-20-0806.
URL https://doi.org/10.1158/0008-5472.CAN-20-0806

[10] M. Strobl, A. L. Martin, J. West, J. Gallaher, M. Robertson-Tessi, R. Gatenby, R. Wenham, P. Maini, M. Damaghi, A. Anderson, Adaptive therapy for ovarian cancer: An integrated approach to PARP inhibitor scheduling, bioRxiv (2023). doi:10.1101/2023.03.22.533721.
URL https://www.biorxiv.org/content/early/2023/03/24/2023.03.22.533721

[11] P. Bayer, J. West, Games and the treatment convexity of cancer, bioRxiv (2023). doi:10.1101/2023.02.27.530257.
URL https://www.biorxiv.org/content/early/2023/02/28/2023.02.27.530257

[12] J. West, B. Desai, M. Strobl, L. Pierik, R. V. Velde, C. Armagost, R. Miles, M. Robertson-Tessi, A. Marusyk, A. R. A. Anderson, Antifragile therapy, bioRxiv (2021). doi:10.1101/2020.10.08.331678.
URL https://www.biorxiv.org/content/early/2021/11/30/2020.10.08.331678

[13] S. Benzekry, C. Lamont, A. Beheshti, A. Tracz, J. M. L. Ebos, L. Hlatky, P. Hahnfeldt, Classical Mathematical Models for Description and Prediction of Experimental Tumor Growth, PLOS Computational Biology 10 (8) (2014) 1–19. doi:10.1371/journal.pcbi.1003800.
URL https://doi.org/10.1371/journal.pcbi.1003800

[14] J. Collis, A. J. Connor, M. Paczkowski, P. Kannan, J. Pitt-Francis, H. M. Byrne, M. E. Hubbard, Bayesian Calibration, Validation and Uncertainty Quantification for Predictive Modelling of Tumour Growth: A Tutorial, Bulletin of Mathematical Biology 79 (4) (2017) 939–974. doi:10.1007/s11538-017-0258-5.
URL https://doi.org/10.1007/s11538-017-0258-5

[15] S. Patmanidis, A. C. Charalampidis, I. Kordonis, K. Strati, G. D. Mitsis, G. P. Papavassilopoulos, Individualized growth prediction of mice skin tumors with maximum likelihood estimators, Computer Methods and Programs in Biomedicine 185 (2020) 105165. doi:https://doi.org/10.1016/j.cmpb.2019.105165.
URL https://www.sciencedirect.com/science/article/pii/S0169260719302184

[16] S. Patmanidis, A. C. Charalampidis, I. Kordonis, G. D. Mitsis, G. P. Papavassilopoulos, Tumor growth modeling: Parameter estimation with Maximum Likelihood methods, Computer Methods and Programs in Biomedicine 160 (2018) 1–10. doi:https://doi.org/10.1016/j.cmpb.2018.03.014.
URL https://www.sciencedirect.com/science/article/pii/S0169260717314505

[17] K. E. Johnson, G. R. Howard, D. Morgan, E. A. Brenner, A. L. Gardner, R. E. Durrett, W. Mo, A. Al'Khafaji, E. D. Sontag, A. M. Jarrett, T. E. Yankeelov, A. Brock, Integrating transcriptomics and bulk time course data into a mathematical framework to describe and predict therapeutic resistance in cancer, Physical Biology 18 (1) (2020) 16001. doi:10.1088/1478-

3975/abb09c.
URL https://doi.org/10.1088/1478-3975/abb09c
[18] M. Scott, C. W. Gunderson, E. M. Mateescu, Z. Zhang, T. Hwa, Interdependence of Cell Growth and Gene Expression: Origins and Consequences, Science 330 (6007) (2010) 1099–1102. doi:10.1126/science.1192588.
[19] S. Klumpp, Z. Zhang, T. Hwa, Growth Rate-Dependent Global Effects on Gene Expression in Bacteria, Cell 139 (7) (2009) 1366–1375. doi:https://doi.org/10.1016/j.cell.2009.12.001.
URL https://www.sciencedirect.com/science/article/pii/S0092867409015050
[20] D. S. Kompala, 10. H. Bremer, PP Dennis, in Escherichia coli and Salmonella, FC Neidhardt, Ed.(ASM Press, Washington, DC, 1996), pp. 1553–1569. 11. O. Maaløe, in Biological Regulation and Development, RF Goldberger, Ed.(Plenum, New York, 1979), J. Gen. Microbiol 19 (1958) 592.
[21] J. Forchhammer, L. Lindahl, Growth rate of polypeptide chains as a function of the cell growth rate in a mutant of Escherichia coli 15, Journal of molecular biology 55 (3) (1971) 563–568.
[22] I. S. Mohiuddin, S.-J. Wei, M. H. Kang, Role of OCT4 in cancer stem-like cells and chemotherapy resistance, Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease 1866 (4) (2020) 165432. doi:https://doi.org/10.1016/j.bbadis.2019.03.005.
URL https://www.sciencedirect.com/science/article/pii/S0925443919300791
[23] X. Q. Wang, W. M. Ongkeko, L. Chen, Z. F. Yang, P. Lu, K. K. Chen, J. P. Lopez, R. T. P. Poon, S. T. Fan, Octamer 4 (Oct4) mediates chemotherapeutic drug resistance in liver cancer cells through a potential Oct4–AKT–ATP-binding cassette G2 pathway, Hepatology 52 (2) (2010) 528–539. doi:https://doi.org/10.1002/hep.23692.
URL https://aasldpubs.onlinelibrary.wiley.com/doi/abs/10.1002/hep.23692
[24] B. S. Koo, S. H. Lee, J. M. Kim, S. Huang, S. H. Kim, Y. S. Rho, W. J. Bae, H. J. Kang, Y. S. Kim, J. H. Moon, Y. C. Lim, Oct4 is a critical regulator of stemness in head and neck squamous carcinoma cells, Oncogene 34 (18) (2015) 2317–2324. doi:10.1038/onc.2014.174.
URL https://doi.org/10.1038/onc.2014.174
[25] T. P. Runarsson, X. Yao, Search biases in constrained evolutionary optimization, IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews) 35 (2) (2005) 233–243. doi:10.1109/TSMCC.2004.841906.
[26] P. E. Hadjidoukas, P. Angelikopoulos, C. Papadimitriou, P. Koumoutsakos, Π4U: A high performance computing framework for Bayesian uncertainty quantification of complex models, Journal of Computational Physics 284 (2015) 1–21. doi:https://doi.org/10.1016/j.jcp.2014.12.006.
URL http://www.sciencedirect.com/science/article/pii/S0021999114008134
[27] K. Ahnert, M. Mulansky, Odeint–solving ordinary differential equations in C++, in: AIP Conference Proceedings, Vol. 1389, American Institute of Physics, 2011, pp. 1586–1589.
[28] C. Tsitouras, Runge–Kutta pairs of order 5(4) satisfying only the first column simplifying assumption, Computers & Mathematics with Applications 62 (2) (2011) 770–775. doi:https://doi.org/10.1016/j.camwa.2011.06.002.
URL https://www.sciencedirect.com/science/article/pii/S0898122111004706
[29] C. Rackauckas, Q. Nie, Differentialequations.jl–a performant and feature-rich ecosystem for solving differential equations in julia, Journal of Open Research Software 5 (1) (2017).
[30] J. Benesty, J. Chen, Y. Huang, I. Cohen, Pearson correlation coefficient, in: Noise reduction in speech processing, Springer, 2009, pp. 1–4.
[31] S. G. Johnson, The NLopt nonlinear-optimization package (2014).
[32] J. Bezanson, S. Karpinski, V. B. Shah, A. Edelman, Julia: A fast dynamic language for technical computing, arXiv preprint arXiv:1209.5145 (2012).
[33] R. Salomon, Some comments on evolutionary algorithm theory, Evolutionary Computation 4 (4) (1996) 405–415.
[34] J. A. Nelder, R. Mead, A simplex method for function minimization, The computer journal 7 (4) (1965) 308–313. doi:10.1093/COMJNL/7.4.308.
URL https://academic.oup.com/comjnl/article/7/4/308/354237
[35] R. Storn, K. Price, Differential Evolution – A Simple and Efficient Heuristic for global Optimization over Continuous Spaces, Journal of Global Optimization 11 (4) (1997) 341–359. doi:10.1023/A:1008202821328.
URL https://doi.org/10.1023/A:1008202821328
[36] D. Makowski, M. S. Ben-Shachar, D. Lüdecke, bayestestR: Describing Effects and their Uncertainty, Existence and Significance within the Bayesian Framework, Journal of Open Source Software 4 (40) (2019) 1541. doi:10.21105/joss.01541.
URL https://doi.org/10.21105/joss.01541
[37] I. Lawrence, K. Lin, A concordance correlation coefficient to evaluate reproducibility, Biometrics (1989) 255–268.
[38] L. Devroye, Discrete Random Variates, Springer New York, New York, NY, 1986, pp. 83–117. doi:10.1007/978-1-4613-8643-8_3.
[39] E. D. Sontag, Dynamic compensation, parameter identifiability, and equivariances, PLOS Computational Biology 13 (4) (2017) 1–17. doi:10.1371/journal.pcbi.1005447.
URL https://doi.org/10.1371/journal.pcbi.1005447
[40] M. J. Simpson, R. E. Baker, S. T. Vittadello, O. J. Maclaren, Practical parameter identifiability for spatio-temporal models of cell invasion, Journal of the Royal Society Interface 17 (164) (2020) 20200055. doi:10.1098/rsif.2020.0055.
URL https://doi.org/10.1098/rsif.2020.0055
[41] K. E. Hines, T. R. Middendorf, R. W. Aldrich, Determination of parameter identifiability in nonlinear biophysical models: A Bayesian approach, Journal of General Physiology 143 (3) (2014) 401–416. doi:10.1085/jgp.201311116.
URL https://doi.org/10.1085/jgp.201311116
[42] A. Raue, C. Kreutz, T. Maiwald, J. Bachmann, M. Schilling, U. Klingmüller, J. Timmer, Structural and practical identi-

fiability analysis of partially observed dynamical models by exploiting the profile likelihood, Bioinformatics 25 (15) (2009) 1923–1929. doi:10.1093/bioinformatics/btp358.
URL https://doi.org/10.1093/bioinformatics/btp358

[43] A. F. Brouwer, R. Meza, M. C. Eisenberg, A Systematic Approach to Determining the Identifiability of Multistage Carcinogenesis Models, Risk Analysis 37 (7) (2017) 1375–1387. doi:https://doi.org/10.1111/risa.12684.
URL https://onlinelibrary.wiley.com/doi/abs/10.1111/risa.12684

[44] M. C. Eisenberg, Input-output equivalence and identifiability: some simple generalizations of the differential algebra approach, arXiv: Quantitative Methods (2013).

[45] Y. Wang, E. D. Sontag, On two definitions of observation spaces, Systems & control letters 13 (4) (1989) 279–289.

[46] R. Almeida, D. Tavares, D. F. M. Torres, The variable-order fractional calculus of variations, Springer, 2019.

[47] N. Varalta, A. V. Gomes, R. d. F. Camargo, A prelude to the fractional calculus applied to tumor dynamic, TEMA (São Carlos) 15 (2014) 211–221.

Appendix A. Analytical solution of the fractional model

To solve the fractional model, first we need to provide some useful definitions obtained from [46, 47].

Definition 1. Let $n \in \mathbb{N}$ and ν a non-integer number with $\mathrm{Re}(\nu) > 0$, the Gel'fand-Shilov function is defined as

$$\phi_n(t) = \begin{cases} \frac{t^{n-1}}{(n-1)!}, \ t \geq 0 \\ 0, \ t < 0 \end{cases} \qquad \text{and} \quad \phi_\nu(t) = \begin{cases} \frac{t^{\nu-1}}{\Gamma(\nu)}, \ t \geq 0 \\ 0, \ t < 0 \end{cases} \tag{A.1}$$

Definition 2. Let $\alpha > 0$, the Mittag-Leffler function is defined by

$$E_\alpha(t) = \sum_{k=0}^{\infty} \frac{t^k}{\Gamma(\alpha k + 1)}, \quad t \in \mathbb{R} \tag{A.2}$$

The two-parameter Mittag-Leffler function with $\alpha, \beta > 0$ is defined by

$$E_{\alpha,\beta}(t) = \sum_{k=0}^{\infty} \frac{t^k}{\Gamma(\alpha k + \beta)}, \ t \in \mathbb{R}$$

Definition 3. Let $n \in \mathbb{N}$ and $f(t) : \mathbb{R} \to \mathbb{R}$ an integrable function. The Integral Operator I of order 1 and n is defined by

$$If(t) = \int_0^t f(t_1)\mathrm{d}t_1 \quad \textit{and} \quad I^n f(t) = \int_0^t \int_0^{t_1} \cdots \int_0^{t_{n-1}} f(t_n)\mathrm{d}t_n \mathrm{d}t_{n-1}\mathrm{d}t_1 \tag{A.3}$$

Theorem 1. *For $f(t) : \mathbb{R} \to$, the n-order integral of f will be*

$$I^n f(t) = \phi_n * f(t) = \int_0^t \phi(t-\tau) f(\tau)\mathrm{d}\tau = \int_0^t \frac{(t-\tau)^{n-1}}{(n-1)!} f(\tau)\mathrm{d}\tau \tag{A.4}$$

where $*$ denotes the Laplace convolution product, and $\phi_n(t)$ is the Gel'fand-Shilov function.

Definition 4. Let $f(t)$ be an integrable function and $\nu \in \mathbb{C}$, such that $\mathrm{Re}(\nu) > 0$, the Riemann-Liouville fractional integral of order ν of $f(t)$, denoted by $I^\nu f(t)$, is defined as

$$I^\nu f(t) = \phi_\nu(t) * f(t) = \int_0^t \frac{(t-\tau)^{\nu-1}}{\Gamma(\nu)} f(\tau)\mathrm{d}\tau$$

Definition 5. Let $f(t)$ be a differentiable function, $m \in \mathbb{N}$ and $\alpha \notin \mathbb{N}$ such that $m - 1 < \mathrm{Re}(\alpha) < m$. Caputo's fractional derivative is defined by

$$D^\alpha f(t) = I^{m-\alpha} D^m f(t) = \phi_{m-a}(t) * D^m f(t)$$

For the solution of the fractional logistic model we need the laplace transform of Caputo's derivative that is

$$\mathfrak{L}\left[D^{\alpha}f(t)\right] = \mathfrak{L}\left[\Phi_{m-a}\right]\mathfrak{L}\left[D^{m}f(t)\right] = s^{\alpha-m}\mathfrak{L}\left[D^{\alpha}f(t)\right], \tag{A.5}$$

and of the following expression involving the Mittag-Leffler function

$$\mathfrak{L}\left[t^{\beta-1}E_{\alpha,\beta}(\pm rt^{\alpha})\right] = \frac{s^{\alpha-\beta}}{s^{\alpha}\mp r} \tag{A.6}$$

Additionally, an important relation between the two-parameter and one-parameter Mittag-Leffler functions is given by

$$E_{\alpha,\alpha+1}(z) = \frac{1}{z}\left[E_{\alpha}(z)-1\right] \tag{A.7}$$

Taking the classical logistic model, and setting $u = 1/N$ we get

$$\frac{\mathrm{d}u}{\mathrm{d}t} = r\left(\frac{1}{K}-u\right) \tag{A.8}$$

Thus, the fractional version of the model can be written as

$$\frac{\mathrm{d}^{\alpha}u}{\mathrm{d}t^{\alpha}} = r\left(\frac{1}{K}-u\right) \tag{A.9}$$

where the derivative order is $0 < \alpha \leq 1$. Applying the Laplace transform to (A.9) we get

$$\begin{aligned} s^{\alpha}F(s) - s^{\alpha-1}u(0) &= \frac{r}{K}\frac{1}{s} - rF(s) \\ \implies F(s) &= u(0)\frac{s^{\alpha-1}}{s^{\alpha}+r} + \frac{r}{K}\frac{s^{-1}}{s^{\alpha}+r} \\ \implies u(t) &= u(0)E_{\alpha}(-rt^{\alpha}) + \frac{r}{K}t^{\alpha}E_{\alpha,\alpha+1}(-rt^{\alpha}) \end{aligned} \tag{A.10}$$

Using the relation found in (A.7), we get

$$\implies u(t) = E_{\alpha}(-rt^{\alpha})\left[u(0)-\frac{1}{K}\right]+\frac{1}{K} \tag{A.11}$$

Setting $N(t) = 1/u$ results

$$N(t) = \frac{1}{\frac{1}{K}+\left[\frac{1}{N(0)}-\frac{1}{K}\right]E_{\alpha}(-rt^{\alpha})} \tag{A.12}$$

Similarly, by incorporating the therapeutic term in (A.9) we get,

$$\frac{\mathrm{d}^{\alpha}N}{\mathrm{d}t^{\alpha}} = rN\left(1-\frac{N}{K}\right) - u_{eff}N,\ 0 < \alpha \leq 1 \tag{A.13}$$

$$\implies N(t) = \frac{r-u_{eff}}{\frac{r}{K}-\left[\frac{r}{K}-\frac{r-u_{eff}}{N(0)}\right]E_{\alpha}\left(-(r-u_{eff})t^{\alpha}\right)} \tag{A.14}$$