

# A DNA-structured mathematical model of cell-cycle progression in cyclic hypoxia

Giulia L. Celora<sup>a,1</sup>, Samuel B. Bader<sup>c</sup>, Ester M. Hammond<sup>c</sup>, Philip K. Maini<sup>a</sup>, Joe Pitt-Francis<sup>b</sup>, Helen M. Byrne<sup>a</sup>

<sup>a</sup>Mathematical Institute, University of Oxford, Oxford, UK

<sup>b</sup>Department of Computer Science, University of Oxford, Oxford, UK

<sup>c</sup>Oxford Institute for Radiation Oncology, Department of Oncology, University of Oxford, Oxford, UK

## Abstract

New experimental data have shown how the periodic exposure of cells to low oxygen levels (*i.e.*, cyclic hypoxia) impacts their progress through the cell-cycle. Cyclic hypoxia has been detected in tumours and linked to poor prognosis and treatment failure. While fluctuating oxygen environments can be reproduced *in vitro*, the range of oxygen cycles that can be tested is limited. By contrast, mathematical models can be used to predict the response to a wide range of cyclic dynamics. Accordingly, in this paper we develop a mechanistic model of the cell-cycle that can be combined with *in vitro* experiments, to better understand the link between cyclic hypoxia and cell-cycle dysregulation. A distinguishing feature of our model is the inclusion of impaired DNA synthesis and cell-cycle arrest due to periodic exposure to severely low oxygen levels. Our model decomposes the cell population into four compartments and a time-dependent delay accounts for the variability in the duration of the S phase which increases in severe hypoxia due to reduced rates of DNA synthesis. We calibrate our model against experimental data and show that it recapitulates the observed cell-cycle dynamics. We use the calibrated model to investigate the response of cells to oxygen cycles not yet tested experimentally. When the re-oxygenation phase is sufficiently long, our model predicts that cyclic hypoxia simply slows cell proliferation since cells spend more time in the S phase. On the contrary, cycles with short periods of re-oxygenation are predicted to lead to inhibition of proliferation, with cells arresting from the cell-cycle when they exit the S phase. While model predictions on short time scales (about a day) are fairly accurate (*i.e.*, confidence intervals are small), the predictions become more uncertain over longer periods. Hence, we use our model to inform experimental design that can lead to improved model parameter estimates and validate model predictions.

**Keywords:** cell-cycle, cyclic hypoxia, cancer, delay-differential equation, model calibration, model selection

## 1. Introduction

The cell-cycle is one of the most fundamental and energy consuming processes in cell biology. It is divided into four phases: G1 (growth), S (DNA synthesis), G2 (growth and preparation for mitosis) and M (mitosis). Transitions between these phases are regulated by complex interactions between cellular pathways and external stimuli which normally act to maintain tissue homeostasis. These interactions are impaired in transformed cells, leading to uncontrolled proliferation, which is a key hallmark of cancer [33]. Cell-cycle dysregulation is further linked to the hallmarks of cancer because it promotes genetic instability, *i.e.*, increasing mutation frequency [45]. Loss of cell-cycle control also plays a significant role in the failure of standard treatments, such as chemotherapy and radiotherapy, where treatment relapse is driven by the emergence of small subpopulations of resistant cells. In normal cells, the DNA damage response (DDR) maintains genetic stability by promoting cell-cycle arrest to allow time for DNA repair or, when DNA damage is irreparable, by promoting cell death via the induction of apoptosis. The DDR is activated early during tumourigenesis as an anti-cancer barrier to oncogene activity and physiological

stresses [7, 30]. However, continuous activation of the DDR results in selective pressure for the outgrowth of mutated cancer cells, with aberrant cell-cycle progression and apoptotic control [12, 30]. Exposure to insufficient oxygen levels, *i.e.*, *hypoxia*, is a key driver of tumorigenesis. Hypoxic regions are commonly found in solid tumours [6, 29, 39, 40] as a result of uncontrolled cell proliferation and abnormal vascular structures. Exposure to severe levels of hypoxia (< 0.1%  $O_2$ ), which are only observed in pathophysiological conditions, leads to replication stress and consequent activation of DDR and the pro-apoptotic p53 tumour suppressor [27, 42, 47]. Further, conditions of less than 0.1%  $O_2$  are associated with resistance to radiotherapy and are therefore commonly referred to as radio-biological hypoxia (RH) [64].

Our study is motivated by evidence that the tumour micro-environment is characterised by highly dynamic oxygen levels. While chronic hypoxia affects tumour regions at a significant distance from vessels, acute/cycling hypoxia can occur close to, and far from, blood vessels, with periods ranging from seconds to hours/days [51]. While high frequency fluctuations are usually associated with vasomotor activity, processes that occur on longer time scales (*e.g.*, vascular remodelling) can generate cycles with longer periods [44]. Such periodic changes in the environment are known to cause inflammation, which promotes

<sup>1</sup>celora@maths.ox.ac.uk

the survival of more aggressive forms of cancer that are resistant to standard treatments [4, 12, 15, 44, 52, 54].

It is possible to culture cells *in vitro* in controlled oxygen environments that partially mimic the fluctuating oxygen levels experienced by tumours *in vivo*. However, *in vitro* experiments are limited by the range of oxygen cycles that can be tested. By contrast, mathematical models can provide insight into a wider range of experimental conditions. Our aim in this work is therefore to develop a novel mechanistic model of the cell-cycle that can be combined with *in vitro* experiments, to increase our understanding of how cyclic hypoxia impacts the cell-cycle.

When exposed to radio-biological hypoxia (RH) *in vitro*, progress through S phase is inhibited due to a rapid reduction in the rate of DNA synthesis [27, 49]. This hypoxia-induced S phase block has been attributed to impaired functioning of the enzyme ribonucleotide reductase (RNR) [27, 46], which mediates *de novo* production of *deoxynucleotide triphosphates* (dNTPs). Since dNTPs are the building blocks of DNA, the decrease in dNTP levels in severe hypoxia causes DNA synthesis to stall. In contrast, milder levels of hypoxia (1-2%  $O_2$ ) do not impact dNTP levels [27]. Replication stress, which is defined as any condition impacting normal DNA replication, leads to activation of the DDR when local oxygen levels are sufficiently low. While cells can initially survive at such low oxygen levels, if severe conditions are prolonged then cell death occurs [49]. Alternatively, if oxygen levels are restored, cells can re-enter the cell-cycle although they may accumulate DNA damage, likely associated with the accumulation of reactive-oxygen species (ROS) during reoxygenation [44]. Depending on the amount of damage sustained, activation of cell-cycle checkpoints causes cells to accumulate in the G2-phase and prevents damaged cells from entering mitosis [15, 31, 46]. Alternatively, if the reoxygenated cells are unable to repair the damage accumulated, they die. Periodic exposure to RH (cyclic hypoxia) may therefore lead to cell death and strong selective pressure for more aggressive clones with impaired checkpoint activation.

Common techniques for monitoring cell-cycle dynamics include flow cytometry and time-lapse microscopy using the fluorescent ubiquitination-based cell-cycle indicator (FUCCI). Both methods can be used to indirectly estimate how the fraction of cells in different stages of the cell-cycle changes over time, by measuring either DNA content (flow cytometry) or expression of cell-cycle related proteins (FUCCI). Since we have access to flow cytometry data, we build our mathematical model so that it can be calibrated against this type of data. As shown in the schematic in Fig. 1, flow cytometry can distinguish cells in different stages of the cycle sorting them according to their DNA content. While cells in the S phase are synthesising new DNA and, therefore, have a variable amount of DNA (from one to two copies), cells in the G1 and G2/M phases have exactly one and two copies of DNA, respectively. Although this method can fail to distinguish cells with similar DNA content (*i.e.*, cells in G1 and early S phases or, cells in the G2/M and late S phases), bromodeoxyuridine (BrdU) labelling can overcome this limitation. Since cells in the S phase incorporate BrdU into newly synthesised DNA, it is possible to distinguish these cells by measuring BrdU uptake [58]. The bivariate

distribution of cells over DNA content and BrdU labelling can therefore be used to estimate cell-cycle distributions which are usually presented as time series data for the evolution of the fractions of cells in the G1, S or G2/M phases. These estimates can be further refined by measuring the expression of proteins involved with cell-cycle regulation, such as cyclin-dependent kinases [18, 58].

Several mathematical formalisms have been proposed to investigate cell-cycle evolution in normal ‘healthy’ conditions, as well as perturbed environments, for example in the context of drug development. These formalisms encompass discrete [26, 59], continuous (either deterministic [1, 8, 10, 25, 43, 54, 53] or stochastic [3, 63]) and hybrid approaches [2, 55]. When interested in the population scale, population balance (PB) models are often used [8, 10, 25, 43, 53, 56]. These take the form of age- and/or phase-structured models, where a structure variable is introduced to track progress through the cell-cycle. Several recent reports have focussed on developing compartment models of the cell-cycle, proliferation and migration which exploit time-resolved FUCCI data [16, 60, 61, 62]. When analysing flow cytometry data, instead, continuous structure variables are commonly used. For example, in a series of papers [8, 9, 10], Basse and coauthors developed cell-cycle models based on coupled partial differential equations in which cells are structured according to their DNA content. In these models, cells in the G1 and G2/M phases have constant DNA content ( $x$ ), with  $x = 1$  and  $x = 2$ , respectively, while  $x$  increases at a constant rate for cells in the S phase. Given a cell DNA content  $x$  at time  $t$ , and given its rate of DNA synthesis, it is possible to estimate the amount of time that a cell has spent in the S phase. However, information about the amount of time spent in the other phases of the cycle is lost. This shortcoming has motivated the development of age-structured models [11], in which the age of a cell in a certain phase of the cycle corresponds to the time it has spent in that phase. Models combining both structure variables (*i.e.*, DNA and cell age) have also been proposed [17], but they are complex and difficult to validate against data. Further cell-cycle specific properties such as size [25], or protein expression levels [18], can also be included in this framework. PB models can be extended to account for variability in compartment specific parameters, such as transition and/or death rates [25] and length of cell-cycle phases [8], to capture the effect of different drugs on cell-cycle progression. Several theoretical studies have investigated the effect of chronic exposure to hypoxia on cell-cycle arrest [1, 22] with particular emphasis on hypoxia-induced G1 arrest. At the population level, quiescence is usually represented by introducing an additional phase to the standard cell-cycle: the so-called ‘G0’ (or quiescent) phase [24, 32], which cells enter prior to committing to DNA synthesis (*i.e.*, prior to entering the S-phase) when the local oxygen levels are too low ( $\approx 1\% O_2$ ). While this extension can account for cell responses to chronic hypoxia, recent experimental findings [5, 27] suggest that it is not sufficient to account for cell behaviours in acute/cyclic hypoxia.

In this paper we develop a mathematical model to investigate how periodic (rather than constant) radio-biological hypoxia (RH) influences cell-cycle dynamics. In §2 we introduce

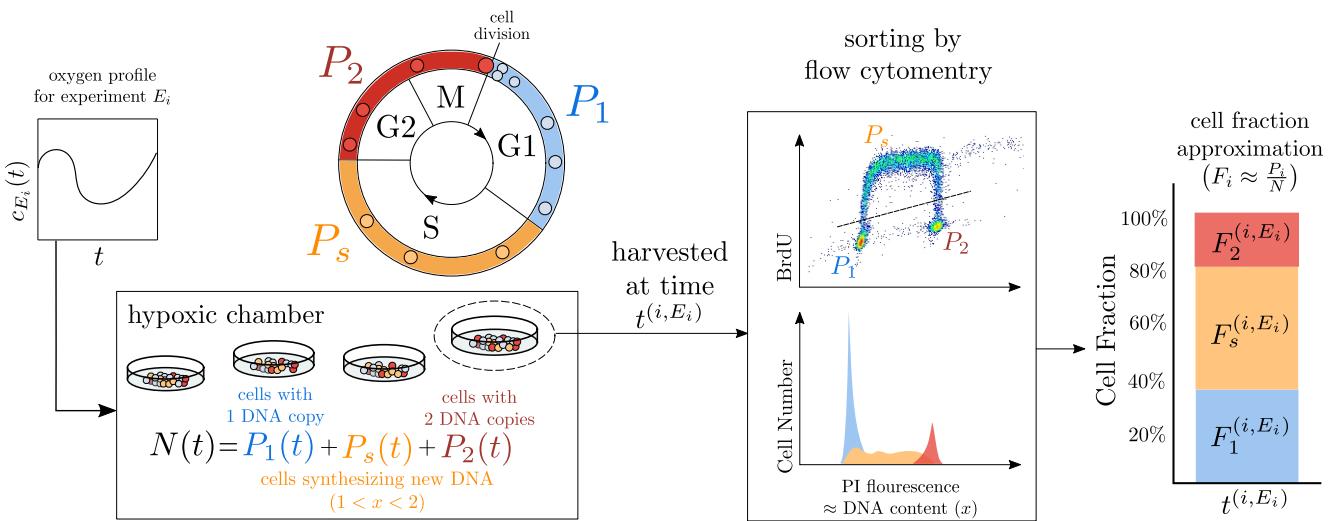


Figure 1: Schematic representation of flow cytometric data: at different time points ( $t^{(i,E_i)}$ ) during experiment  $E_i$  a subpopulation of  $N$  (the population of cells in the chamber) is harvested and sorted via flow cytometry according to their DNA content (PI fluorescence intensity), and BrdU uptake. This allows us to extract the fraction  $F_1$ ,  $F_s$  and  $F_2$  of cells in the subpopulation with one, intermediate (i.e. in between one and two), and two copies of DNA, respectively. This is an approximation of the frequency of cells with different DNA content in the total population in the chamber  $N$ .

the experimental data from [5] and summarise the biological mechanisms on which our model is based. In §3 we present our 5-compartment, DNA-structured, model which describes how the number of cells in each phase of the cycle evolves over time. Novel features of the model include a time and oxygen-dependent rate of DNA synthesis (here denoted by  $v(t)$ ) and the introduction of two “checkpoint” compartments ( $C_1$  and  $C_2$ ), where cells arrest due to unfavourable environmental conditions or due to the accumulation of damage and stress. We further reduce our model to 4 ordinary differential equations coupled to a delay-differential equation, where the delay ( $\tau_s$ ) represents the duration of the S phase and is not necessarily constant; rather it depends on the environment that the cells have encountered. In §4, we first consider model predictions in well-oxygenated conditions. Here we recover the well-known result of cells converging to a regime of balanced exponential growth. In §5, we explore model predictions in constant and cyclic RH, highlighting the distinct effects that these two types of hypoxia have on the cell-cycle dynamics. We first show that the model can replicate experimental data from [5] and then use it to investigate the cell-cycle dynamics for modes of oxygen fluctuations not yet tested experimentally. In §5.2, we introduce a class of models ( $\mathcal{M}$ ) of decreasing complexity by neglecting some of the mechanisms (and complexity) included in the full model (such as arrest in  $C_1$  or  $C_2$ ). In §6 we compare these models using *Bayesian model selection*: we first calibrate our class of models against experimental data by using *Bayesian inference* methods and then identify the “best” model structure based on the *deviance information criteria*. This analysis reveals that the  $C_1$  and  $C_2$  checkpoint compartments are both necessary to describe the experimentally observed dynamics. In §7, we use the selected and calibrated model to predict cell responses to different oxygen environments when considering uncertainty in parameter estimates. We also explain how our results could inform the design of new experiments to validate the model and/or improve

the accuracy of the parameter estimates. In §8 we conclude by summarising our results and outlining possible directions for future work.

## 2. Experimental motivation

This work is inspired by experimental data showing how the cell-cycle of RKO (colorectal cancer) cells changes when they are cultured *in vitro* as 2D monolayers and exposed to fluctuating oxygen levels [5]. As shown in Fig. 1, cells are cultured in chambers where the oxygen levels  $c = c(t)$  (where  $t$  is time) are carefully controlled and assumed to be spatially homogeneous. At prescribed time points, a subset of the cells is analysed using flow cytometry to estimate the fractions of cells in the G1, S and G2/M phases of the cycle, which we denote, respectively, by  $F_1$ ,  $F_s$  and  $F_2$ . Since each measurement requires cells to be harvested, measurement errors can be taken to be independent. In the absence of cell death, we have that by definition  $F_1 + F_s + F_2 = 1$ , so only two of the three cell fractions are needed to fully characterise the cell-cycle dynamics.

As shown in Fig. 2, two experimental protocols are tested. At time  $t = 0$ , cells are exposed to either constant radio-biological hypoxia ( $c(t) \equiv c_{RH} \approx 0.1\% O_2$ ,  $t > 0$ ) or periodic cycles of radio-biological hypoxia (2 hr at  $c = c_{RH}$  and 2 hr at  $2\% O_2$ ). Prior to both experiments, the cells are cultured in normoxia ( $21\% O_2$ ) so that the measurements at time  $t = 0$  contain information on this condition. In normoxia, and in absence of competition, the cells are typically in a regime of balanced exponential growth for which the cell fractions  $F_m$  are stationary (i.e., they do not change over time). Hence a single set of measurements is sufficient to fully characterise the cell-cycle dynamics in these environmental conditions. We therefore divide the data into three different sets:  $E_0$  (normoxia),  $E_1$  (constant RH) and  $E_2$  (cyclic RH). The histograms in Fig. 2 summarise the data available from [5], obtained by averaging over between two and

four replicates. As mentioned in §1, our focus is on fluctuating environments (i.e. the scenario  $E_2$ ), however data from  $E_0$  and  $E_1$  are also used to assist with parameter estimation.

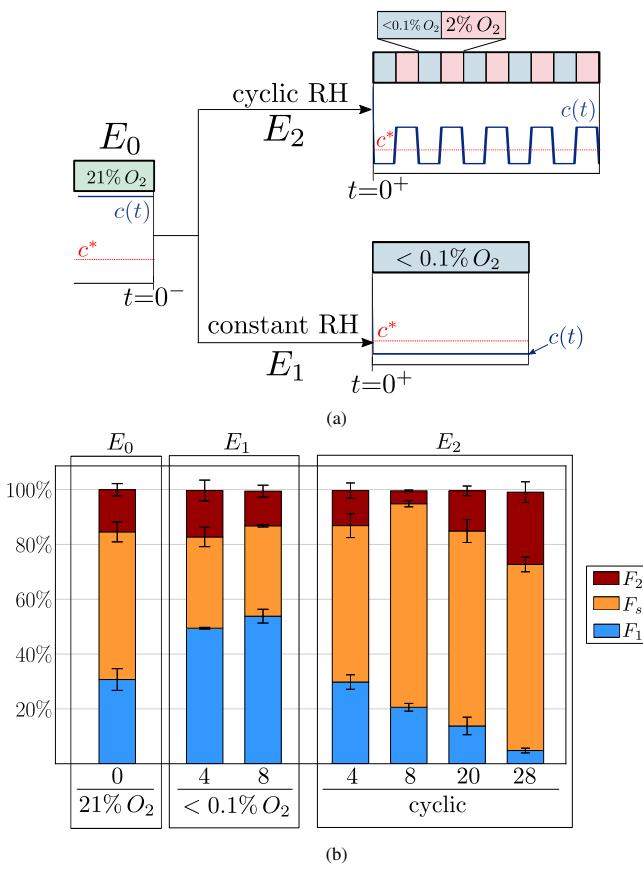


Figure 2: (a) Data are split into three sets corresponding to the three experiments : normoxia ( $E_0$ ), constant RH ( $E_1$ ) and cyclic RH ( $E_2$ ). We plot the evolution of the oxygen levels,  $c(t)$  (see blue curve), and compare it to the threshold  $c^*$  at which DNA synthesis is impaired (see red horizontal line). (b) Histogram summarising the data from [5]. At each time point we report the mean cell fractions estimated from multiple (between two and four) measurements. Error bars indicate one standard deviation in the estimated values. The complete data sets can be found in Appendix B.

The data in Fig. 2 show that both constant ( $E_1$ ) and cyclic RH ( $E_2$ ) redistribute the cells along the cycle. For experiment  $E_1$ , cells tend to accumulate in the G1 phase (as seen by the increase in the fraction  $F_1$ ), whereas in experiment  $E_2$ , we observe a decrease in the fraction  $F_1$ . In this case, cells initially concentrate in the S phase (as seen by the increase in the fraction  $F_s$ ), while they accumulate in the G2/M phase at later times (see time  $t = 28$  hr). For experiment  $E_1$  we only consider the early dynamics since on longer time scales ( $t > 20$  hr) cell-death becomes significant. By contrast, cells can survive for longer periods in cyclic hypoxia, with cell death negligible for up to 28 hours.

As mentioned in the introduction, and based on current experimental evidence [4, 27, 49], three mechanisms may influence the cell-cycle dynamics under cyclic hypoxia: the reduction in the rate of DNA synthesis (Mechanism 1) and variation in the timing of the G1-S transition (Mechanism 2), both due to dNTP shortage and, the arrest of damaged cells in the G2 phase

upon completion of the S phase (Mechanism 3). Here, we assume that there is an oxygen threshold,  $c^*$ , such that dNTP levels drop (due to the impaired activity of RNR enzyme) when oxygen levels fall below  $c^*$ . Based on the experiments in [27], we estimate that  $0.1\% O_2 < c^* < 2\% O_2$ . We therefore fix  $c^*$  at an intermediate value (i.e.,  $c^* \approx 1\% O_2$ ). Our aim is to develop a mathematical model that captures these three mechanisms and that can be used to investigate whether they can explain the experimental data in both constant RH ( $E_1$ ) and cyclic RH ( $E_2$ ) conditions.

### 3. The mathematical model

We propose a 5-compartment partial differential equation (PDE) model to describe cell-cycle dynamics in cyclic RH. For simplicity, we assume that the cells are in a well-mixed, spatially-homogeneous environment where the oxygen concentration  $c = c(t)$  is externally prescribed. We also assume that cell death is negligible since this is supported by the experimental observations.

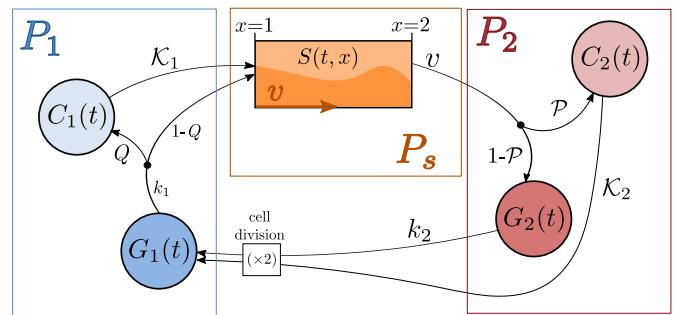


Figure 3: Schematic representation of the 5-compartment model of the cell-cycle given by Equations (1). Here the two independent variables are time  $t$  and the cell DNA content  $x$  (measured in number of copies), while  $P_1$ ,  $P_2$  and  $P_s$  are as defined in Fig. 1. The sub-population  $P_1$  groups cells in the  $G_1$  and  $C_1$  compartments, and similarly  $P_2$  comprises cells in the  $G_2$  and  $C_2$  compartments. Cells in the  $G$  compartments are progressing along the cell-cycle as usual, while the  $C$  compartments are “checkpoint” compartments where cells arrest. The black dots on the arrows correspond to redistribution of cell fluxes according to a given probability (for example in  $P_1$  the influx  $k_1 G_1$  is redistributed with probability  $Q$  into  $C_1$  and  $1 - Q$  into  $S$ ). To account for DNA synthesis during the  $S$ -phase we structure cells in the  $S$ -compartment according to their DNA content  $x$ , which is synthesised at velocity  $v$ .

As illustrated in Fig. 3, we subdivide the population into 5 compartments. Here  $G_1 = G_1(t)$  and  $G_2 = G_2(t)$  denote, respectively, the number of cells at time  $t$  in the G1 and G2/M phases of the cycle that are actively proceeding along their cycle. On the other hand,  $C_1 = C_1(t)$  and  $C_2 = C_2(t)$  are, respectively, the number of cells in the G1 and G2/M phases that are arrested at time  $t$ . We define the latter as checkpoint compartments and assume the rates of entry into and exit from these compartments change over time in response to the current and previous environmental conditions (i.e., oxygen levels). Finally, we structure cells in the S phase according to their DNA content,  $x$ , so that  $S = S(x, t)$  represents the number of cells with DNA content  $x$  at time  $t$ . Here  $x$  is a dimensionless variable corresponding to the relative DNA content of a cell, scaled so that  $1 \leq x \leq 2$ .

Cells start the S phase with  $x = 1$  and exit it upon completion of DNA duplication with  $x = 2$ .

Referring to the schematic in Fig. 3 and applying the principle of mass balance, we obtain the following system of equations for the time evolution of the model variables:

$$\frac{dG_1}{dt} = 2(k_2 G_2 + \mathcal{K}_2(t)C_2) - k_1 G_1, \quad t > 0, \quad (1a)$$

$$\frac{dC_1}{dt} = Q(t)k_1 G_1 - \mathcal{K}_1(t)C_1, \quad t > 0, \quad (1b)$$

$$\frac{\partial S}{\partial t} + \frac{\partial}{\partial x} (v(t)S) = 0, \quad x \in (1, 2], \quad t > 0, \quad (1c)$$

$$\frac{dG_2}{dt} = -k_2 G_2 + (1 - \mathcal{P}(t))v(t)S(t, 2), \quad t > 0, \quad (1d)$$

$$\frac{dC_2}{dt} = \mathcal{P}(t)v(t)S(t, 2) - \mathcal{K}_2(t)C_2, \quad t > 0. \quad (1e)$$

The factor of 2 in Eq. (1a) arises because cell division produces two daughter cells. The positive constants  $k_1, k_2$  [hr<sup>-1</sup>] represent the rates at which cells leave  $G_1, G_2$ , respectively. In Eq. (1c), we account for DNA synthesis by assuming that cells in the S compartment are advected along the  $x$ -axis at velocity  $v(t) > 0$  [hr<sup>-1</sup>] (i.e., they produce DNA at a rate  $v(t)$ ). In Eq. (1b), we assume that cells exiting the  $G_1$  compartment arrest (i.e. enter the  $C_1$  compartment) with probability  $Q \in [0, 1]$ , while they proceed to enter the S phase with probability  $1 - Q$ . We further assume that arrested cells re-enter the cycle at rate  $\mathcal{K}_1$  [hr<sup>-1</sup>]. In a similar manner, in Eq. (1e) we assume that cells exiting the S compartment arrest in  $C_2$  with probability  $\mathcal{P} \in [0, 1]$  and arrested cells may re-enter the cycle at a rate  $\mathcal{K}_2$  [hr<sup>-1</sup>]. We account for the effect that different oxygen levels have on cell behaviour by allowing certain model parameters (specifically  $\mathcal{K}_1, \mathcal{K}_2, v, \mathcal{P}$  and  $Q$ ) to vary over time in response to the oxygen levels  $c$ . The boundary condition for Equation (1c) is derived by applying conservation of cell flux at  $x = 1$ . Under the assumption that  $v(t) > 0$  for all  $t > 0$ , we find:

$$v(t)S(t, 1) = (1 - Q(t))k_1 G_1(t) + \mathcal{K}_1(t)C_1(t). \quad (1f)$$

We close Eqs. (1a)-(1f) by imposing the following initial conditions:

$$\begin{aligned} G_1(0) &= G_{1,0}, & C_1(0) &= C_{1,0}, \\ S(x, 0) &= S_0(x) \quad x \in (1, 2], & & (1g) \\ G_2(0) &= G_{2,0}, & C_2(0) &= C_{2,0}, \end{aligned}$$

where  $G_{i,0}, C_{i,0}$  are constants and  $S_0(x)$  is a prescribed function. Since cell numbers can not be negative, we have that  $G_{i,0} \geq 0$ ,  $C_{i,0} \geq 0$  and  $S_0(x) \geq 0$  for all  $x \in [1, 2]$ .

In order to compare the model output to flow cytometry data, we need to express cell fractions (see Fig. 1) in terms of our model variables. As shown in Fig. 1, cells can be divided in three sub-populations depending on the cell-cycle phase they are in:  $P_1$ ,  $P_s$  and  $P_2$ . Cells in G1 belong to  $P_1$ , cells in G2/M to  $P_2$  and cell in S to  $P_s$ . The total number of cells,  $N = N(t)$ , is then obtained by summing the number of cells in each sub-population:

$$N(t) = P_1(t) + P_s(t) + P_2(t), \quad (2a)$$

where (as illustrated in Fig. 3)

$$\begin{aligned} P_1(t) &= C_1(t) + G_1(t), \\ P_s(t) &= \int_1^2 S(x, t) dx, \\ P_2(t) &= G_2(t) + C_2(t). \end{aligned} \quad (2b)$$

Differentiating Eq. (2a) with respect to time and using (1), we find that  $N(t)$  satisfies:

$$\frac{dN}{dt} = k_2 G_2 + \mathcal{K}_2(t)C_2, \quad (3a)$$

$$\text{with } N(0) = \sum_{i=1}^2 (G_{i,0} + C_{i,0}) + \int_1^2 S_0(x) dx := N_0. \quad (3b)$$

The population proliferation rate,  $\omega$ , is given by:

$$\omega(t) = \frac{1}{N} \frac{dN}{dt} = \frac{k_2 G_2 + \mathcal{K}_2 C_2}{N}, \quad (4)$$

while the cell fractions  $f_m$  are defined by:

$$f_m(t) = \frac{P_m(t)}{N(t)}, \quad m \in \{1, s, 2\}, \quad (5)$$

and correspond to the probability that a cell randomly chosen from the total population belongs to the sub-population  $P_m$ . We note that the fractions  $f_m$  are not independent; Eq. (2) implies that  $\sum_m f_m = 1$ . We further introduce the distribution  $s = s(x, t)$ :

$$s(x, t) = \frac{S(x, t)}{N(t)}, \quad 1 < x < 2, \quad (6)$$

which corresponds to the probability that a cell in the sub-population  $P_s$  has a DNA content  $x \in [1, 2]$  at time  $t$ . A summary of the model variables is given in Table 1.

A key feature of our model is that the velocity  $v$  is assumed to depend on the oxygen levels  $c$  in order to account for impaired DNA synthesis at low oxygen levels ( $c < c^*$ ). In particular, we propose the following piece-wise linear ODE to describe how the advection velocity  $v$  adapts to changes in local oxygen levels  $c(t)$ :

$$\frac{dv}{dt} = \begin{cases} -\mathcal{R}_-(v - v_-), & c(t) < c^*, \\ \mathcal{R}_+(v_+ - v), & c(t) \geq c^*, \end{cases} \quad (7a)$$

$$\text{with } v(0) = v_0. \quad (7b)$$

In Eqs. (7),  $\mathcal{R}_\pm$ ,  $v_0$  and  $v_\pm$  are positive constants, while  $c^*$  is the threshold oxygen level below which RNR activity is impaired (see §2 for details). Here  $v_+$  and  $v_-$  represent, respectively, the equilibrium velocities in well-oxygenated ( $c > c^*$ ) and severely hypoxic ( $c < c^*$ ) environments and  $0 < v_- < v_+$ . It is straightforward to show that if  $v_0 \in [v_-, v_+]$ , then  $v(t) \in [v_-, v_+]$  at all times  $t > 0$ , so that  $v_+$  and  $v_-$  can also be viewed, respectively, as the maximum and minimum rates of DNA synthesis. When oxygen levels drop below the critical value  $c^*$ , the rate of DNA synthesis decreases (i.e.  $dv/dt \leq 0$  for  $c < c^*$ ), capturing the effect of dNTP shortage. As observed in [27], cells

Table 1: Summary of the model variables in Eqs. (1) and (8).

| Description |  |
|-------------|--|
| $t$         | time [hr]  |
| $x$         | copies of DNA in a cell ( $1.0 \leq x \leq 2.0$ )  |
| $c$         | externally controlled oxygen levels [%]  |
| $G_1(t)$    | number of cells in $P_1$ active in the cycle   |
| $C_1(t)$    | number of cells in $P_1$ that are arrested   |
| $S(x, t)$   | number density of cells in $P_s$ with DNA content $x$  |
| $G_2(t)$    | number of cells in $P_2$ active in the cycle   |
| $C_2(t)$    | number of cells in $P_2$ that are arrested   |
| $N(t)$      | total cell number  |
| $f_m(t)$    | predicted fraction of cells ( $f_m \in [0, 1]$ ) in the $P_m$ sub-population $m \in \{1, s, 2\}$               |
| $F_m(t)$    | experimentally measured fraction of cells ( $F_m \in [0, 1]$ ) in the $P_m$ sub-population $m \in \{1, s, 2\}$ |
| $s(x, t)$   | DNA distribution of cells in the $P_s$ sub-population  |
| $\tau_S(t)$ | duration of the S phase for cells that exit the S compartment at time $t$ [hr]                                 |
| $\omega(t)$ | population proliferation rate [ $\text{hr}^{-1}$ ]   |

maintain a minimum level of DNA synthesis even after prolonged exposure to severe hypoxia ( $c < c^*$ ); we therefore assume that  $v_- > 0$ , which guarantees that  $S(t, 1)$ , as defined by Eq. (1f), remains finite. Once  $c > c^*$ , dNTP levels are restored and the rate of DNA synthesis increases (i.e.,  $dv/dt > 0$ ), towards its maximum value  $v_+$ . The rates  $\mathcal{R}_\pm$  are assumed constant for simplicity although, in practice, they may depend on oxygen levels. In the absence of suitable data to identify these dependencies, we proceed with the simplest form that captures the currently available data and postpone investigation of more complex functional forms to future work.

We introduce the checkpoint compartment  $C_1$  to account for transient arrest of cells in the  $P_1$  phase due to dNTP shortage. The rates of transition into ( $Q$ ) and out of ( $\mathcal{K}_1$ ) the compartment depend on current oxygen levels. For simplicity, we adopt the following functional forms:

$$\mathcal{K}_1(t) = K_1 H_\epsilon(c(t) - c^*), \quad (8a)$$

$$Q(t) = q H_\epsilon(c^* - c(t)), \quad (8b)$$

where  $K_1$  and  $q$  are positive constants with  $q \in [0, 1]$  and  $c^*$  is the oxygen threshold introduced above. The function  $H_\epsilon$  is the standard continuous approximation of the Heaviside function:

$$H_\epsilon(\cdot) = \frac{1}{2} \left[ \tanh\left(\frac{\cdot}{\epsilon}\right) + 1 \right]. \quad (8c)$$

where  $0 < \epsilon \ll 1$  is a small parameter, here set to  $\epsilon = 0.01\%$  oxygen. Note that Eqs. (8a)-(8b) ensure that cells arrest when  $c < c^*$  and that they resume cycling by entering the S phase

once  $c > c^*$ . We considered alternative functional forms for  $\mathcal{K}_1$  and  $Q$  (such as assuming them to be proportional to  $v$ ) but found that the resulting model was unable to capture the experimentally observed dynamics (results not shown).

In Eq. (7), prolonged or frequent exposure to RH slows DNA replication and, therefore, increases the time  $\tau_S$  that cells spend in the S phase. More precisely, we denote by  $\tau_S = \tau_S(t)$  the amount of time a cell exiting the S phase at time  $t$  has taken to complete DNA synthesis (we will explain how  $\tau_S(t)$  is computed in §3.1). Consequently, the larger  $\tau_S$ , the more cells have been damaged during the S phase due to re-oxygenation and replication stress (in RH). Since the accumulation of damage regulates the arrest of cells in the G2 phase (here captured by cells transitioning into the  $C_2$  compartment), we assume that the probability,  $\mathcal{P}(t)$ , of arrest in G2 increases with  $\tau_S(t)$ . In particular, we assume that, when cells exit the S phase, they arrest with a probability  $\mathcal{P} \in [0, 1]$  where:

$$\mathcal{P}(t) = p H_{\epsilon_p}(\tau_S(t) - T). \quad (8d)$$

In Eq. (8d), we fix the small parameter  $\epsilon_p$  so that  $\epsilon_p = 0.1$  hr, while  $p \in [0, 1]$  and  $T > 0$  are unknown parameters. Here  $p$  is the maximum probability that a cell enters the  $C_2$  compartment while  $T$  captures the critical duration of the S phase after which cells are likely to arrest in  $C_2$ . Based on Eq. (7), the timescale for completing the S phase,  $\tau_S$ , satisfies  $v_+^{-1} \leq \tau_S \leq v_-^{-1}$ . When oxygen levels are sufficiently high, (i.e.  $c(t) > c^*$  for all  $t$ ), neglecting an initial transient in the case  $v_0 \neq v_+$ , we have that  $\tau_S \equiv v_+^{-1}$ . Since we do not expect cells to arrest in an oxygen-rich environment we require  $\mathcal{P}(t) \approx 0$  when  $\tau_S(t) \equiv v_+^{-1}$ . Consequently we set  $T \gg v_+^{-1}$ . As for the  $C_1$  compartment, we assume that, once oxygen levels rise above the critical value  $c^*$ , cells can repair any damage they have accumulated and re-enter the cell-cycle at a rate  $\mathcal{K}_2$ :

$$\mathcal{K}_2(t) = K_2 H_\epsilon(c(t) - c^*), \quad (8e)$$

where  $K_2$  is a non-negative constant and  $\epsilon = 0.01\%$  oxygen as in Eq. (8a).

### 3.1. Model reduction to a system of Delay Differential Equations (DDEs)

Here we show how Eqs. (1) can be rewritten as a system of ODEs coupled to a (state-dependent) DDE with a non-constant delay  $\tau_S(t)$  which we view as a state variable. Given that the velocity  $v$  is always positive, shocks can not form and a straightforward application of the *method of characteristics* to Eq. (1c) yields

$$S(x, t) = \begin{cases} S_0 \left( x - \int_0^t v(\ell) d\ell \right), & \int_0^t v(\ell) d\ell \leq x - 1, \\ \frac{(1 - Q)k_1 G_1 + \mathcal{K}_1 C_1}{v} \Big|_{t - \tau(t, x)}, & \int_0^t v(\ell) d\ell > x - 1, \end{cases} \quad (9)$$

where the function  $\tau = \tau(t, x)$  is implicitly defined by:

$$x - 1 = \int_{t - \tau(t, x)}^t v(\ell) d\ell, \quad \text{if } \int_0^t v(\ell) d\ell > x - 1, \quad (10)$$

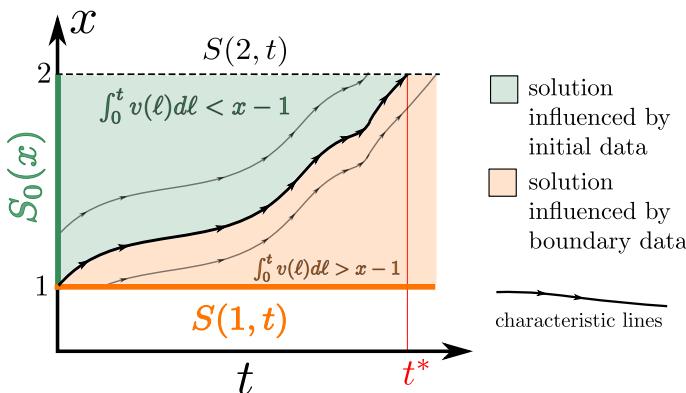


Figure 4: Schematic illustration of the characteristic curves in the  $(t, x)$  plane. The bold curve divides the plane into two distinct regions: in the green region (where  $\int_0^t v(\ell) d\ell < x - 1$ ) the solution  $S(x, t)$  is determined by the initial data; in the orange region (where  $\int_0^t v(\ell) d\ell > x - 1$ )  $S(x, t)$  is instead determined by the boundary data.

and indicates the amount of time a cell with DNA content  $x$  at time  $t$  has spent in the  $S$  compartment.

As illustrated in Fig. 4, the condition  $\int_0^t v(\ell) d\ell = x - 1$  divides the  $(t, x)$ -plane into two regions depending on whether the characteristics propagate from the boundary data curve (orange curve) or the initial data curve (green curve). When  $t > t^*$  (i.e. when  $S(2, t)$  is influenced by the boundary data) the total time spent in the  $S$  compartment is given by  $\tau_S(t) = \tau(t, 2)$ . Using Eq. (10), we find that  $\tau_S$  is implicitly defined by:

$$1 = \int_{t-\tau_S(t)}^t v(\ell) d\ell, \quad \text{if } t > t^*, \quad (11a)$$

where  $t^*$  is defined implicitly by the integral equation

$$\int_0^{t^*} v(\ell) d\ell = 1. \quad (11b)$$

However, more information about the behaviour of cells prior to the beginning of the experiment is needed to estimate  $\tau_S(t)$  for  $0 \leq t < t^*$ . Let us consider a cell that exits the  $S$  phase at time  $t < t^*$ . Then, at time  $t = 0$ , its DNA content was:

$$x_0(t) = 2 - \int_0^t v(\ell) d\ell, \quad (12)$$

as we can write  $\tau_S = t + \tau_0$ , where  $t$  denotes the time spent in the  $S$  compartment since the beginning of the experiment and  $\tau_0$  the time spent in the  $S$  phase prior to the beginning of the experiment. The functional form of  $\tau_0$  will depend on the conditions in which the cells were grown for  $t < 0$ . It is therefore part of the initial data that must be specified in order to fully define the model. In §4, we will show that if cells are cultured in an oxygen-rich environment for  $t < 0$ , then  $\tau_0$  depends only on the DNA content  $x_0(t)$  (i.e.  $\tau_0(t) = \tau_0(x_0(t))$ ). In summary we have:

$$\tau_S(t) = t + \tau_0(t), \quad 0 \leq t < t^*. \quad (13)$$

Differentiating Eq. (11a) with respect to time, and given Eq. (13), we deduce that the variable  $\tau_S(t)$  satisfies:

$$\frac{d\tau_S}{dt} = 1 - \frac{v(t)}{v(t - \tau_S)}, \quad t \geq t^*, \quad (14a)$$

$$\text{with } \tau_S(t) = t + \tau_0(t), \quad 0 \leq t \leq t^*, \quad (14b)$$

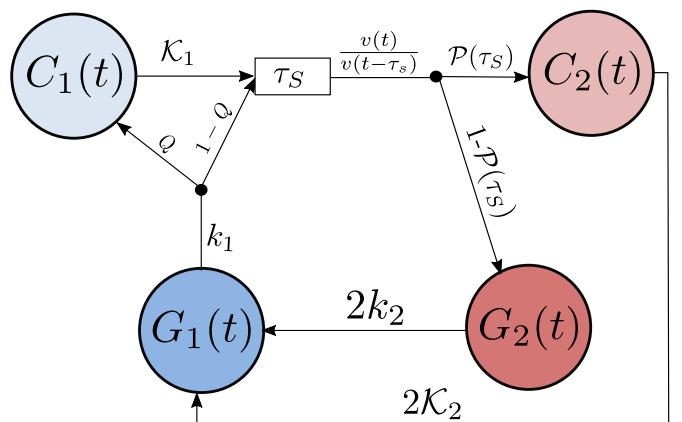


Figure 5: Schematic representation of the 4-compartment model with delay (see Equations (17)) derived by solving Eq. (1) via the method of characteristics. The black nodes correspond to the conservation of the cell fluxes, where the input is redistributed with probabilities  $P(\tau_S)$  ( $Q(c)$ ) and  $1 - P(\tau_S)$  ( $1 - Q(c)$ ) in  $C_2$  ( $C_1$ ) and  $G_2$  ( $G_1$ ), respectively.

### 3.2. Summary of the cell-cycle model with delay

Before presenting the full model, we perform the following re-scaling:

$$\hat{G}_i = \frac{G_i}{N_0}, \quad \hat{C}_i = \frac{C_i}{N_0}, \quad \hat{N} = \frac{N}{N_0}. \quad (15)$$

Since our model is linear, this re-scaling does not affect the form of the governing equations; it only alters the initial conditions. Consequently we can determine the system dynamics without data on the initial cell number. This is helpful since we only have experimental data on the initial cell fractions, while  $N_0$  is unknown. We further assume that cells are initially actively cycling (i.e. none of them is arrested) and DNA synthesis is proceeding at maximum speed so that:

$$v_0 = v_+, \quad C_{i,0} = 0, \quad f_{i,0} = \frac{G_{i,0}}{N_0}, \quad i \in \{1, 2\}, \quad (16)$$

where  $f_{i,0}$ , with  $i \in \{1, 2\}$ , denote the cell fractions at time  $t = 0$ . As discussed in §4, this is a reasonable assumption if cells are cultured in oxygen-rich environments prior to the start of the experiments.

Applying Eq. (15) to Eqs. (1)-(3), using Eqs. (7)-(8), (1c) and (14) and dropping the hat notation, the governing equations become

$$\frac{dG_1}{dt} = 2(k_2 G_2 + \mathcal{K}_2(t) C_2) - k_1 G_1, \quad t > 0, \quad (17a)$$

$$\frac{dC_1}{dt} = Q(t) k_1 G_1 - \mathcal{K}_1(t) C_1, \quad t > 0, \quad (17b)$$

$$\frac{dG_2}{dt} = -k_2 G_2 + (1 - P(t)) I(t - \tau_S(t), t), \quad t > 0, \quad (17c)$$

$$\frac{dC_2}{dt} = P(t) I(t - \tau_S(t), t) - \mathcal{K}_2(t) C_2, \quad t > 0, \quad (17d)$$

$$\frac{dN}{dt} = k_2 G_2 + \mathcal{K}_2(t) C_2, \quad t > 0, \quad (17e)$$

$$\frac{dv}{dt} = \begin{cases} -\mathcal{R}_-(v - v_-), & c(t) < c^*, \\ \mathcal{R}_+(v_+ - v), & c(t) \geq c^*, \end{cases} \quad t > 0, \quad (17f)$$

$$\frac{d\tau_S}{dt} = \frac{v(t - \tau_S) - v(t)}{v(t - \tau_S)}, \quad t \geq t^*, \quad (17g)$$

wherein

$$I(s, t) = \begin{cases} v(t) s_0 \left( 2 - \int_0^t v(\ell) d\ell \right), & t < t^*, \\ v(t) \frac{(1 - Q(s)) k_1 G_1(s) + \mathcal{K}_1(s) C_1(s)}{v(s)}, & t \geq t^*, \end{cases} \quad (17h)$$

$$\mathcal{K}_1(t) = K_1 H_\epsilon(c(t) - c^*), \quad (17i)$$

$$Q(t) = q H_\epsilon(c^* - c(t)), \quad (17j)$$

$$\mathcal{P}(t) = p H_{\epsilon_p}(\tau_S(t) - T), \quad (17k)$$

$$\mathcal{K}_2(t) = K_2 H_\epsilon(c(t) - c^*), \quad (17l)$$

and

$$\begin{aligned} v(0) &= v_+, \quad \tau_S(\xi) = \xi + \tau_0(\xi), \quad 0 \leq \xi \leq t^*, \\ N(0) &= 1, \quad G_i(0) = f_{i,0}, \quad C_i(0) = 0 \quad i \in \{1, 2\}. \end{aligned} \quad (17m)$$

In Eqs. (17h)-(17m) it remains to specify the distribution  $s_0 = S_0/N_0$  and the function  $\tau_0$  (see §3.1). The former is subject to the constraint:

$$\int_1^2 s_0(x) dx = f_{s,0} = 1 - f_{1,0} - f_{2,0}. \quad (18)$$

In Fig. 5, we present a schematic summary of the model (as in Fig. 3) where the  $S$ -compartment is replaced by the time delay  $\tau_S$ .

As listed in Table 2, our model has 12 parameters. As discussed in §2, we fix  $c^* = 1\%$  oxygen in line with experimental evidence. Further, based on the experimental estimates of the rate of DNA synthesis in [49], we expect  $v_-$  to be small. Since the limit  $v_- \rightarrow 0$  is non-singular, it can be shown that the solution is not sensitive to the precise value of  $v_-$  provided that  $v_- \sim O(10^{-3})$  (results not shown). We therefore fix  $v_- = 0.005 \text{ hr}^{-1}$ , which is sufficiently small to exhibit the correct qualitative behaviour. With  $c^* = 1$  and  $v_- = 0.005$ , there are ten unknown parameters, which we split into two classes. While  $\theta = [k_1, v_+, k_2]$  comprises parameters associated with oxygen-independent mechanisms,  $\ell = [K_1, \mathcal{R}_+, \mathcal{R}_-, p, q, T, K_2]$  contains parameters associated with oxygen-dependent mechanisms. Given the large number of unknown parameters and the expected variability for different cell lines, we here focus on the RKO cancer cell line and the data in Fig. 2 to estimate the model parameters. Discussion on how we estimate the model parameters is postponed to §6. First, we present some characteristic predictions of our model obtained by solving Eqs. (17) numerically using the Python *ddeint* package to integrate delay-differential equations. As we will justify in §5.2, we assume the  $C_2$  checkpoint to be irreversible by setting  $K_2 = 0$ . This is likely to be the case for the RKO cancer cell line considered, but may not be true for other cell lines, particularly if they are p53-deficient, given the role of p53 in mediating G2 arrest in several cancer cells [57].

#### 4. Cell-cycle progression in a static normoxic environment

Let us first consider the case in which cells are exposed to a constant, oxygen-rich environment (i.e.,  $c(t) \equiv \bar{c} \gg c^*$ ). Given Eqs. (17f)-(17g) and (17m), we have that  $v \equiv v_+$  and  $\tau_S \equiv v_+^{-1}$ . In this case,  $Q \equiv 0$ ,  $P \equiv 0$  and  $C_1(t) = C_2(t) \equiv 0$  for all  $t > 0$ , so that there are no arrested cells. We conclude that when  $c \equiv \bar{c} > c^*$ , Eqs. (17) reduce to the following system:

$$\frac{dG_1}{dt} = 2k_2 G_2 - k_1 G_1, \quad t > 0, \quad (19a)$$

$$\frac{dG_2}{dt} = -k_2 G_2 + k_1 G_1(t - v_+^{-1}), \quad t > 0, \quad (19b)$$

$$\frac{dN}{dt} = k_2 G_2, \quad t > 0, \quad (19c)$$

subject to

$$G_1(\xi) = \begin{cases} f_{1,0}, & \xi = 0, \\ v_+ k_1^{-1} s_0 (2 - \xi v_+), & v_+^{-1} < \xi < 0, \end{cases} \quad (19d)$$

$$N(0) = 1, \quad G_2(0) = f_{2,0},$$

where  $s_0$  satisfies (18). We note that Eqs. (19) is analogous to models previously proposed in the literature, such as the model by Basse et al. [8, 9] when dispersion is neglected. Since Eqs. (19) are linear with a constant delay, they can be solved exactly via superposition of exponential functions  $e^{\Lambda_i t}$  where  $\Lambda_i$  ( $i = 1, 2, \dots$ ) are the complex roots of the characteristic polynomial (see Eq. (A.3) in Appendix A). In the case of DDEs, the characteristic polynomial is a transcendental equation with an infinite number of roots so that the computation of  $\Lambda_i$  is non-trivial. To investigate the transient dynamics, it is therefore more convenient to solve Eqs. (19) numerically. In Fig. 6 we present numerical solutions for two sets of initial conditions: cells are initially synchronised in either the  $G_1$  (Fig. 6(a)) or  $G_2$  (Fig. 6(b)) compartment. This corresponds to setting  $s_0 \equiv 0$  with  $f_{1,0} = 1$  and  $f_{2,0} = 0$  (for panel (a)), or  $f_{2,0} = 1$  and  $f_{1,0} = 0$  (for panel (b)).

As shown in Fig. 6, the evolution of the cell fractions  $f_m$  in panels (a.1) and (b.1) differs only up to time  $t \approx 20 \text{ hr}$ ; after this first transient the cell fractions evolve to constant values, denoted by  $\bar{f}_m$ , which are independent of the initial conditions. When looking at Fig. 6(a.2)-(b.2), we see that the evolution of  $G_1$ ,  $G_2$  and  $N$ , differs even at long times. From time  $t > 20 \text{ hr}$ , the variables have a similar qualitative behaviour, but their values remain higher for scenario (b) than for scenario (a). While cells in scenario (b) start proliferating at the beginning of the simulations, cells in scenario (a) are delayed since they need to complete the  $S$ -phase before they can replicate. Once cells enter the  $G_2$  compartment, we see an increase in the cell number  $N$ . We also note that from time  $t > 60 \text{ hr}$ , for both scenarios,  $N(t)$ ,  $G_1(t)$  and  $G_2(t)$  increase exponentially at a constant rate. This agrees with the results in Fig. 7, which show that the population proliferation rate  $\omega(t)$  (see Eq. (4)) asymptotes to a constant value  $\lambda$  for both sets of simulations.

Fig. 8 shows the evolution of the distribution  $s(x, t)$ . We note that in Fig. 8(a), cells are initially highly synchronised in the

Table 2: Summary of the parameters that appear in Eqs. (1)-(8), together with their typical values for the RKO cancer cell line.

| Description     | Typical value(s)                   |
|-----------------|------------------------------------|
| $c^*$           | $\approx 1\% O_2$                  |
| $v_-$           | $5 \times 10^{-3} \text{ hr}^{-1}$ |
| $v_+$           | $0.083 \text{ hr}^{-1}$            |
| $k_1$           | $0.195 \text{ hr}^{-1}$            |
| $k_2$           | $0.22 \text{ hr}^{-1}$             |
| $K_1$           | $0.0 - 2.5 \text{ hr}^{-1}$        |
| $K_2$           | $0.0 - 2.0 \text{ hr}^{-1}$        |
| $T$             | $15 - 20 \text{ hr}$               |
| $p$             | $0.0 - 1.0$                        |
| $q$             | $0.0 - 1.0$                        |
| $\mathcal{R}_+$ | $10^{-3} - 2.0 \text{ hr}^{-1}$    |
| $\mathcal{R}_-$ | $10^{-3} - 2.0 \text{ hr}^{-1}$    |

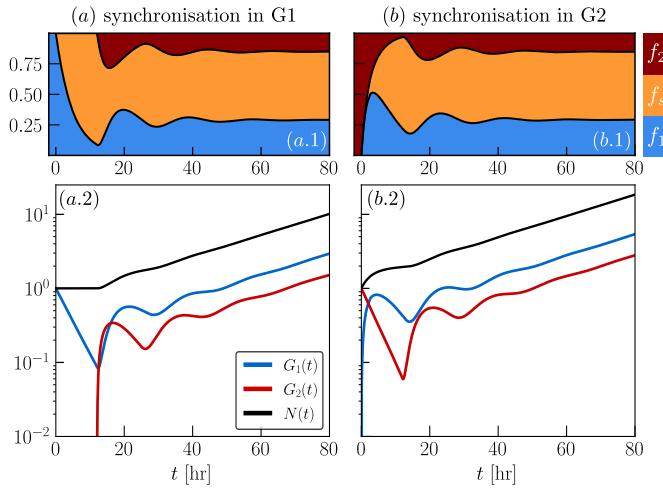


Figure 6: Numerical simulations of Eqs. (19) for two sets of initial conditions: (a)  $s_0 \equiv 0$ ,  $f_{1,0} \equiv 0$  and  $f_{2,0} = 1$ ; (b)  $s_0 \equiv 0$ ,  $f_{1,0} \equiv 1$  and  $f_{2,0} = 0$ . For both simulations, the parameters  $k_1$ ,  $k_2$  and  $v_+$  are set as in Table 4. In panels (a.1) and (b.1), we plot the cell-cycle dynamics with cumulative plots of the cell fractions  $f_m$  with  $m \in \{1, s, 2\}$ ; in panels (a.2) and (b.2) we show instead the predicted time evolution of the model variables  $G_1$ ,  $G_2$  and  $N$  on a semi-logarithmic scale.

S compartment, with the formation of a front that propagates at velocity  $v_{max}$  (note the steep gradient in the profiles at time  $t = 5$  and  $t = 10$ ). This is because of the discontinuity in the initial data for  $G_1$  (when  $f_{1,0} = 1$  and  $S_0 \equiv 0$ ,  $G_1(\xi)$  in Eqs. (19) is discontinuous at  $\xi = 0$ ). The discontinuity propagates along the  $x$  axis but it quickly smooths out, due to the de-synchronisation of cells in the  $G_1$  and  $G_2$  compartments. By contrast, in Fig. 8(b), there is no discontinuity in the initial data for  $G_1$  and therefore the profile of  $s$  is smoother. Despite these large differences in the distributions at early times,  $s$  eventually evolves to the same stationary distribution (in Fig. 8, the curves for  $t = 60$  and  $t = 80$  are almost indistinguishable) and the time scales required to approach the stationary distribution for the two initial conditions are comparable.

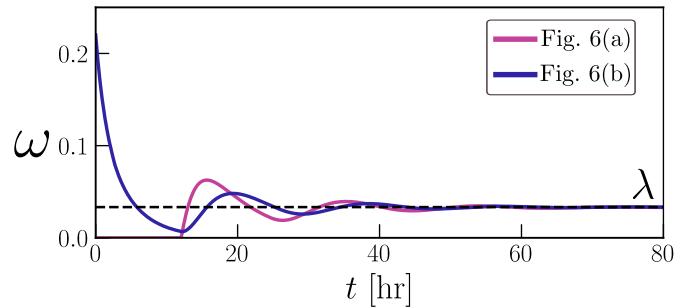


Figure 7: Comparison of the population proliferation rate  $\omega$  (see Eq. (4)) for the two scenarios in Fig. 6. We see that on the long time scale, the proliferation rate settles to a constant value,  $\lambda$ , independently of the initial conditions chosen.

Following the notation introduced in [54], we term the asymptotic solution of the model a *phase stationary solution* (PSS) to indicate that, in this regime, the cell fractions  $f_1$ ,  $f_s$  and  $f_2$ , and the distribution  $s(x, t)$  remain constant in time. This is similar to predictions from other models in the literature [8, 10, 13, 21, 54] in the context of unperturbed growth; this regime is usually referred to as balanced, or asynchronous, exponential growth [9, 14]. As mentioned previously, we can write the solution to Eqs. (19) as a superposition of exponential functions  $e^{\Lambda_i t}$ . At long times, the behaviour is dominated by the exponential whose eigenvalue  $\Lambda_i$  has the largest real part (here denoted by  $\lambda$ ). It is possible to prove that this eigenvalue  $\lambda$  is real and positive (see Appendix A for details). We conclude that for  $t \gg 1$  the system approaches a regime of exponential growth (as observed in the numerical results in Fig. 6) in which the model variables take the form:

$$\begin{aligned} G_{1,2}(t) &\approx \bar{f}_{1,2} \xi_N e^{\lambda t}, \\ N(t) &\approx \xi_N e^{\lambda t}, \quad t \gg 1. \\ S(x, t) &\approx k_1 \bar{f}_1 v_+^{-1} e^{-\lambda v_+^{-1}(x-1)} \xi_N e^{\lambda t}, \end{aligned} \quad (20)$$

where  $\bar{f}_{1,2} \geq 0$  and  $\xi_N > 0$  are constant. Substituting Eqs. (20) into Eqs. (5), we obtain that for  $t \gg 1$ ,  $f_m(t) \rightarrow \bar{f}_m$  for

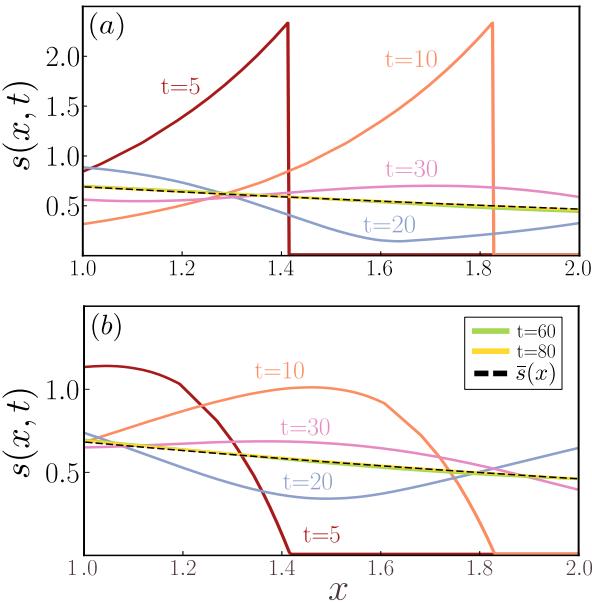


Figure 8: Evolution of the distribution of cells in  $P_s$  with respect to  $x$ . We estimate the distribution  $s(x, t)$  by numerically solving Eq. (9) for the two cases in Fig. 6: (a) initial synchronisation in  $G_2$  (as in Fig. 6a) and (b) initial synchronisation in  $G_1$  (as in Fig. 6b). We compare the long time behaviour with the analytically computed solution from the phase stationary solution (see black dashed line). Here the colour scheme in (a) and (b) is the same.

$m \in \{1, s, 2\}$ . Therefore  $\bar{f}_1$ ,  $\bar{f}_s$  and  $\bar{f}_2$  indicate, respectively, the stationary values of cell fraction in the G1, S and G2/M phase of the cycle. Focusing on the distribution of cells  $s(x, t)$ , this converges to the stationary distribution  $\bar{s}(x) = v_+^{-1} k_1 \bar{f}_1 e^{-\lambda v_+^{-1}(x-1)}$ , which is monotonically decreasing in  $x$  (see black line in Fig. 8). This indicates that, for  $t \gg 1$ , cells in S phase are more likely to be starting DNA synthesis (with  $x \approx 1$ ) rather than close to completing it (with  $x \approx 2$ ). We note that the longer the duration of the S phase, or equivalently the smaller  $v_+$ , the steeper is the  $\bar{s}(x)$  curve and, therefore, the larger is the number of cells in  $P_s$  that are concentrated around  $x \approx 1$ . In the limit where the DNA synthesis velocity is high, i.e.  $\lambda v_+^{-1} \rightarrow 0$ , the distribution  $\bar{s}(x)$  flattens and cells are equally spread along the  $P_s$  phase.

We remark that while such balanced exponential growth is difficult to observe *in vivo*, where non-linear effects due to competition affecting cell proliferation, it can be observed experimentally when cells are grown in a nutrient-rich environment in the absence of contact inhibition [8, 54].

## 5. Cell-cycle progression in a dynamic environment

Having discussed the model predictions for a well-oxygenated environment, we now investigate the behaviour it exhibits under radio-biological hypoxia (RH). We start by comparing how continuous ( $E_1$ ) and cyclic RH ( $E_2$ ) affect RKO cells originally in a regime of exponential growth (i.e., the phase stationary solution computed in §4). To replicate the oxygen dynamics in the experiments from [5], we use the following

functional form for the oxygen levels  $c = c(t)$  at time  $t > 0$ :

$$c(t) = c_+ + (c_+ - c_{RH}) \sum_{i=1} \left[ H_\epsilon(t - t_i^{(R)}) - H_\epsilon(t - t_i^{(H)}) \right], \quad (21)$$

where  $H_\epsilon$  is defined in Eq. (8c),  $t_i^{(R)}$  and  $t_i^{(H)}$  are the times at which oxygen levels decrease and increase across the threshold  $c = c^*$ , respectively. By fixing  $t_i^{(R)} = 4(i-1)+2$  and  $t_i^{(H)} = 4(i-1)$ ,  $c_{RH} \approx 0.1\%$  oxygen and  $c_+ = 2\%$  oxygen, we can reproduce the 2hr+2hr cycle corresponding to experiment  $E_2$  in Fig. 2. Fixing  $t_1^{(H)} = 0$  and  $t_1^{(R)} = \infty$ , we obtain the constant RH ( $E_1$ ) conditions with  $c < c^*$  for  $t > 0$ .

**Initial conditions.** Under standard culture conditions, cells *in vitro* are typically in a regime of balanced growth. We therefore initiate our simulations by assuming that cells are growing according to Eqs. (20). Recalling that we have re-scaled the model so that  $N(0) = 1$ , we have the following initial conditions:

$$\begin{aligned} G_1(0) &= \bar{f}_1, & s_0(x) &= \tau_s k_1 \bar{f}_1 e^{-\lambda \tau_s(x-1)}, \\ G_2(0) &= \bar{f}_2, & C_{1,2}(0) &= 0. \end{aligned} \quad (22a)$$

Finally, to complete Eqs. (14), we must specify the function  $\tau_0$  (see §3.1). Since we assume DNA is synthesised at a constant rate  $v_+$  for  $t < 0$ , it is straightforward to show that cells with DNA content  $x$  at time  $t = 0$  have spent a period  $\tau_0(s) = \frac{s-1}{v_+}$  in the S phase. Using Eq. (12), we have that

$$\tau_0(t) = \frac{1}{v_+} \left( 1 - \int_0^t v(\ell) d\ell \right), \quad t \leq 0. \quad (22b)$$

### 5.1. Numerical results

We start by considering scenario  $E_1$  where cells are exposed to constant RH for about 15 hours (see Fig. 9a and pink curve in Fig. 10). As mentioned in §2, at longer times, cell start dying and our model stops being valid, therefore we run simulations only up to this time. Fig. 10(a.1) shows that  $f_1(t)$  rapidly increases in the first 5 hours, while it appears to settle to a value of  $\approx 50\%$  at longer times. This suggests that cells tend to accumulate in the G1 phase. By contrast, both  $f_s(t)$  and  $f_2(t)$  decrease. While  $f_s(t)$  decreases monotonically over time, the decrease in  $f_2(t)$  is delayed by a couple of hours, during which time its value remains approximately constant. Focussing now on the evolution of the model variables (see Fig. 10(a.2)), we see that the number of cells in the  $C_1$  compartment increases monotonically, but the rate of increase tends to slow after about 10 hours. By contrast,  $G_1(t)$  slightly increases in the first few hours ( $\approx 4$  hours) while it decreases rapidly at later times. Similarly, the  $G_2$  compartment starts to empty only after a couple of hours from the beginning of the simulation. As the velocity  $v$  decreases (see Fig. 10(b)), the flux of cells out of the S phase (see Eqs. (17)) also decreases, contributing to the reduction in  $G_2(t)$ . Even though  $\tau_s$  quickly increases above the threshold  $T \approx 17$  (see Fig. 10(d)), cells do not accumulate in  $C_2$ . They instead remain trapped in the S phase due to the reduction in the rate of DNA synthesis. As expected, given the trend in  $G_2(t)$ , the overall proliferation rate decreases monotonically (see Fig. 10(a)),

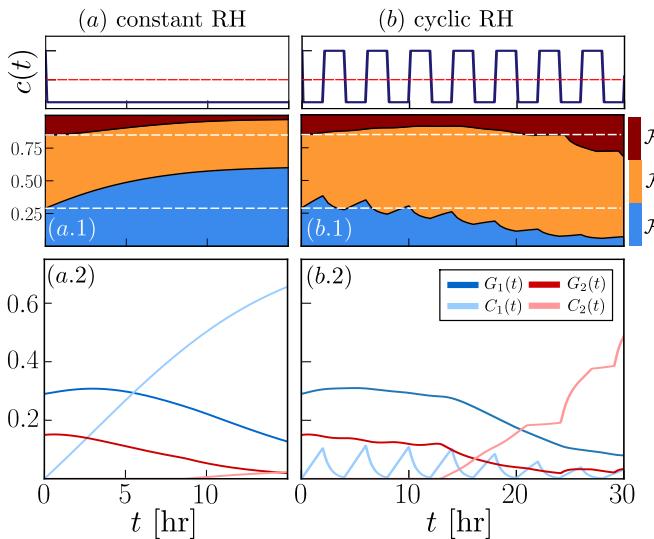


Figure 9: Numerical simulations of cell-cycle evolution as predicted by Eqs. (17) and (21)-(22) in constant RH (a) and cyclic RH (b). The top row illustrates the evolution of the oxygen levels  $c$  in the two simulations. The middle row panels show the evolution of cell fractions over time in the form of a cumulative diagram. The white lines indicate the composition at time  $t = 0$ , i.e., the phase stationary solution  $\bar{f}_1$ ,  $\bar{f}_s$  and  $\bar{f}_2$ . Panels (a.2) and (b.2) present the predicted evolution of cell numbers in the different model compartments:  $G_1(t)$ ,  $C_1(t)$ ,  $G_2(t)$  and  $C_2(t)$ . The parameters  $k_1$ ,  $v_{\pm}$  and  $k_2$  are as in Table 1 while  $K_2 = 0$  and the remaining parameter values are the estimated mean parameter values for model  $\mathcal{M}_1$  reported in Table C.9.

driving the gradual flattening of the population growth curve  $N(t)$  (see Fig. 10(c)).

In the case of cyclic RH ( $E_2$ ), the dynamics are quite different. As shown in Fig. 9(b), while initially  $f_s(t)$  increases and  $f_2(t)$  decreases, the opposite occurs at later times ( $t \approx 25$  hr) when  $f_2(t)$  increases while  $f_1(t)$  decreases. Overall, the fraction  $f_1(t)$  decreases, albeit non-monotonically, so that, at the end of the simulation, its value is almost negligible (i.e.,  $0 < f_1 \ll 1$ ). Looking at the evolution of the number of cells in each compartment illustrated in Fig. 9(b.2), we see that cells transiently accumulate in the  $C_1$  compartment during exposure to RH (light blue curve) and resume cycling during re-oxygenation. The evolution of  $G_2$  and  $G_1$  is qualitatively similar; both remain approximately constant up to  $t \approx 15$  hr, after which time they start to decrease. At the same time, the  $C_2$  checkpoint is activated (Fig. 10(d) shows that  $\tau_S(t) \approx T = 17.03$  hr so that  $P \approx p/2$  for  $t \approx 15$  hr). Since the  $C_2$  checkpoint prevents cells from replicating, its activation results in a rapid decrease in the population proliferation rate  $\omega$  (see the blue curve in Fig. 10(a)-(b)). Under cyclic conditions, the rate of DNA synthesis falls below  $v_+$  but remains well above its minimum value  $v_- \approx 5 \times 10^{-3}$ . Despite the marked fluctuations in the velocity  $v$ ,  $\tau_S$  increases almost steadily (albeit at a lower rate than for  $E_1$ ) until it plateaus at a maximum value of  $\approx 22$  hours.

As shown in Fig. 11, constant and cyclic RH also affect the evolution of the distribution  $s(x, t)$ . Under constant RH ( $E_1$ , see pink curves), cells tend to accumulate near  $x \approx 1$ . Due to the low rate of DNA synthesis, the profile appears approximately stationary. In particular, comparison of the pink curves in pan-

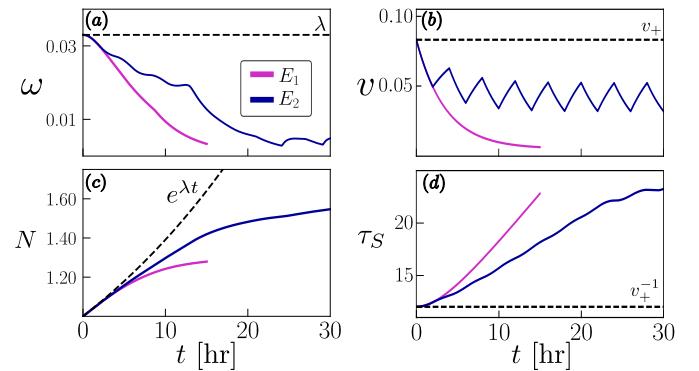


Figure 10: We compare the predicted population dynamics for the two simulations in Fig. 9: (a) population proliferation rate  $\omega$  as defined by Eq. (4), (b) the total cell number  $N$ , (c) the DNA synthesis velocity  $v$  and (d) the duration of the S phase  $\tau_S$ . The parameters are chosen as in Fig. 9.

els (b) and (c) suggests that the discontinuity in the profile has not moved significantly. While for  $E_1$ ,  $s(x, t)$  is characterised by a single slow-moving front, for  $E_2$  (see blue curves) the front is followed by a series of asymmetric spikes which propagate along the  $x$ -axis with a faster velocity  $v$  (see Fig. 9(e)). Each spike corresponds to cells in the  $C_1$  compartment quickly re-entering the cell-cycle at the S phase after re-oxygenation; these cells remain highly synchronised as they proceed through the S phase. Since there is no re-oxygenation in  $E_1$ , spikes are not observed. Focusing on the blue curve in Fig. 11(c), and moving from left ( $x = 1$ ) to right ( $x = 2$ ), the peak value decreases (as the spikes become wider). However, at later times (see Fig. 11(e)), the left-most spikes have lower peaks due to the depletion of cells in the  $P_1$  population (see Fig. 9).

Overall, our results suggest that, while both constant and cyclic RH lead to inhibition of proliferation, the mechanisms driving arrest are distinct. In the first case, cells arrest in the  $C_1$  compartment and DNA synthesis is almost completely inhibited. On the other hand, under cyclic hypoxia ( $E_2$ ), DNA synthesis proceeds, albeit at a lower rate. This leads to an increase in  $\tau_S$  and cell accumulation in the S phase. Despite being able to complete the S phase, cells later arrest in the  $C_2$  checkpoint. This, however, is evident only at long times ( $\approx 24$  hr), when we see a large accumulation of diploid cells. Our findings are in line with the experimental data in Fig. 2 (which are taken from [5]), indicating that our model can capture the experimentally observed cell-cycle dynamics.

Next, we use our model to investigate modes of cyclic RH not yet tested in the laboratory. For example, in Fig. 12, we fix the length of the RH phase to 2 hours and compare the growth curves for different periods of reoxygenation. We find that when the reoxygenation periods are significantly longer than the time cells spend in RH, cells have time to recover and continue proliferating, albeit at a lower rate. As shown in Fig. 13, even when cyclic RH does not result in inhibition of proliferation (i.e., cycles 2+6 and 2+8 in Fig. 12), it can still affect the cell-cycle distribution when compared to the predictions for the phase stationary solution (represented by the dashed line). For the example in Fig 13(a), cyclic RH eventually leads to  $f_s(t)$  being above its initial value  $\bar{f}_s$ . This is relevant when thinking

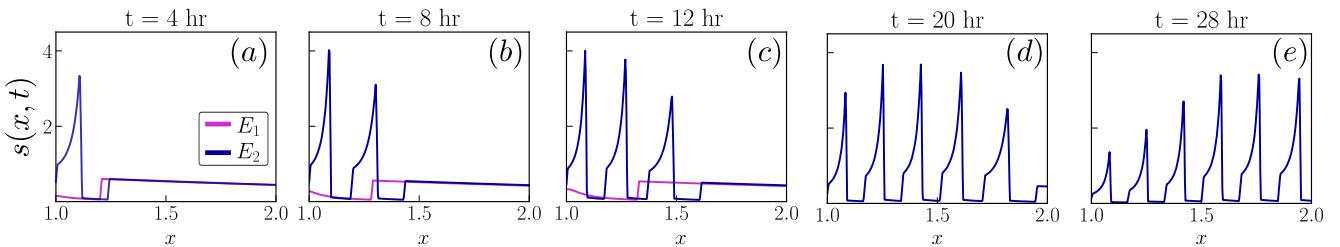


Figure 11: Evolution of the DNA distribution  $s(x, t)$  computed by numerically solving Eq. (9) for the two cases in Fig. 9: acute RH (pink curve- $E_1$ ) and cyclic RH (blue curve- $E_2$ ). The parameters are chosen as in Fig. 9.

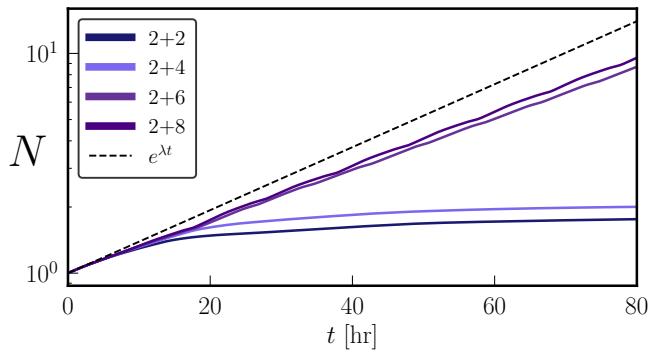


Figure 12: Evolution of the total cell number  $N(t)$  for different oxygen cycles as predicted by Eqs. (17). While the time spent in RH is constant for all experiments (i.e., 2 hr), we change the length of the reoxygenation phase (from 2 hr as in Fig. 9(b) to 8 hr). We also include, as a control case, the PSS (black dotted line). Parameters are chosen as in Fig. 9.

about cell-cycle specific treatment. Since cyclic RH changes the distribution of cells along the cycle, it can impair or favour treatment efficacy. For example, we know that cells in different phases of the cell-cycle have different responses to radiotherapy [48]. In this case, cells in the G2/M phase have been shown to be the most sensitive to radiotherapy [48]. Referring to Fig. 13(a), we note that persistent exposure to cyclic hypoxia biases the cell-cycle distribution to the S phase. As such, it could decrease the overall sensitivity of cells to RT, even during reoxygenation, when oxygen levels do not directly increase cell radio-resistance. When considering longer re-oxygenation periods, such as in Fig. 13(b), we observe larger fluctuations in the cell fractions. Towards the end of the re-oxygenation phase,  $f_2$  is just above the value  $\bar{f}_2$ . This suggests that applying treatment at this time could improve treatment efficacy. However, if treatment is not timed accurately and it is applied when  $f_2$  is at its minimum or when cells are in RH, then cyclic hypoxia could instead favour radio-resistance. This highlights the possible use of mathematical models to predict cell-cycle dynamics and how this can affect treatment outcomes when testing protocols *in vitro* accounting also for other mechanisms (such as oxygen) that can affect treatment outcomes. In order to achieve this, robust calibration of the model to experimental data is needed.

### 5.2. A class of models: comparison of different modelling assumptions

The model defined by Eqs. (17) is complex in its response to variable oxygen levels. Based on experimental evidence from

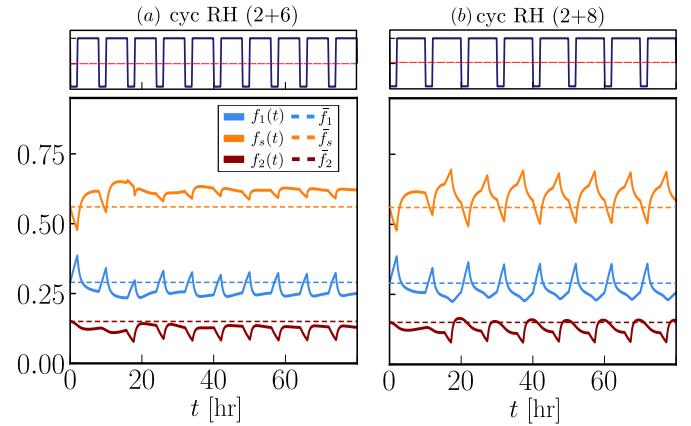


Figure 13: Evolution of the phase fraction  $f_m$  (see solid lines) with  $m \in \{1, s, 2\}$  as predicted by the model for two different cyclic protocols in Fig. 12: (a) 2 hr (< 0.2%) + 6 hr (2%) or (b) 2 hr (< 0.2%) + 8 hr (2%). We compare the dynamics in cyclic RH with the PSS ( $\bar{f}_m$ ) indicated by the dotted line. The parameters are chosen as in Fig. 9.

[5, 27, 49], we argued in §2 that three distinct mechanisms play a key role in determining cell-cycle evolution in cycling RH. While we have shown that the model agrees qualitatively with the experimental data, it is unclear whether all three mechanisms are necessary to recapitulate the data. Further, in §5, we fixed  $K_2 = 0$  based on experimental observations; now we want to test whether this assumption is justified, based on the data available. To answer these questions, we construct a class of models,  $\mathbb{M}$ , obtained by systematically reducing the complexity of the full model. A list of the models considered is presented in Table 3. While all models reduce to Eqs. (19) under normoxia, they differ under RH. The alternative models are derived from the full model, named  $\mathcal{M}_0$ , by setting  $K_2 = 0$  ( $\mathcal{M}_1$ ) and fixing either  $C_1 \equiv 0$  ( $\mathcal{M}_2$ ) or  $C_2 \equiv 0$  ( $\mathcal{M}_3$ ). Here models  $\mathcal{M}_2$  and  $\mathcal{M}_3$  test whether the checkpoint compartments are needed to describe the data. The last two rows of Table 3 list the unknown model parameters associated with each model; while all models share the parameters  $\theta$ , they vary in the number of parameters  $\ell^{(k)}$  associated with RH. In particular, the number of unknown parameters associated with model  $\mathcal{M}_k$  decreases as  $k$  increases (i.e., as the model complexity reduces). As we will detail in §6, we calibrate our models to the data in Fig. 2, and, in Fig. 14 and Fig. 15, we compare the resulting fits.

In Fig. 14, we compare predictions when cells are exposed to constant RH ( $E_1$ ). In this case all models, except  $\mathcal{M}_2$ , pre-

Table 3: Schematics showing the biological mechanisms (see §3 for the explanation of the mechanisms (1)-(3)) included in each model  $\mathcal{M}$  belonging to the class of models  $\mathbb{M}$ . We also indicate the list of unknown parameter sets  $\theta$  and  $\Theta$  associated to each model  $\mathcal{M}_k$  where the first set determines the phase stationary solution, while the second set plays a role in the oxygen dependent mechanisms (Mechanism 1)-(Mechanism 3). The full model,  $\mathcal{M}_0$ , has the largest number of parameters (i.e. higher complexity) but it accounts for all the mechanisms we expect to play a role in cell-cycle dynamics under cyclic hypoxia.

|             | $\mathcal{M}_0$                                     | $\mathcal{M}_1$                                | $\mathcal{M}_2$                        | $\mathcal{M}_3$                          |
|-------------|---|--|--|--|
| Mechanism 1 | True  | True   | True                                   | True                                     |
| Mechanism 2 | True  | True   | False ( $C_1 \equiv 0$ )               | True                                     |
| Mechanism 3 | True  | True ( $K_2 = 0$ )                             | True ( $K_2 = 0$ )                     | False ( $C_2 \equiv 0$ )                 |
| $\theta$    | $[k_1, v_+, k_2]$                                   | $[k_1, v_+, k_2]$                              | $[k_1, v_+, k_2]$                      | $[k_1, v_+, k_2]$                        |
| $\Theta$    | $[K_1, \mathcal{R}_+, \mathcal{R}_-, p, q, T, K_2]$ | $[K_1, \mathcal{R}_+, \mathcal{R}_-, p, q, T]$ | $[\mathcal{R}_+, \mathcal{R}_-, p, T]$ | $[K_1, \mathcal{R}_+, \mathcal{R}_-, q]$ |

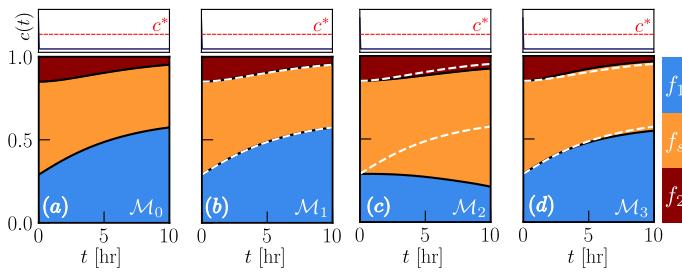


Figure 14: Comparison of the cell-cycle dynamics predicted by solving numerically the models listed in Table 3 for constant RH (scenario  $E_1$ ). We here use the best fit obtained for each model. The white dotted lines in the panels indicate the evolution predicted by model  $\mathcal{M}_0$  (see first panel) here used as a reference. The parameters  $k_1, v_+$  and  $k_2$  are as in Table 1 while the oxygen-dependent parameters  $\ell$  are taken to be the estimated mean values reported in Table C.9.

dict that  $f_1$  increases in line with the experimental observations from [5]. In contrast, model  $\mathcal{M}_2$  (Fig. 14(c)) is incompatible with the experimental data, predicting a monotonic decrease in  $f_1(t)$  and cell accumulation in the S phase (observe the large increase in  $f_s(t)$  over time). This suggests that in constant RH, the  $C_1$  checkpoint plays a key role while the  $C_2$  checkpoint can be neglected given that model  $\mathcal{M}_3$  qualitatively reflects the same trends observed experimentally. However, neglecting the  $C_1$  checkpoint would not significantly impact predictions under cyclic RH. Indeed, looking at Fig. 15(c), we note that model  $\mathcal{M}_2$  is in good agreement with model  $\mathcal{M}_0$  and  $\mathcal{M}_1$  (indicated by the white dotted line). While  $\mathcal{M}_2$  does not capture the rapid fluctuations in  $f_1(t)$  predicted by  $\mathcal{M}_0$ , the overall trend is the same, with  $f_1(t)$  decreasing after each cycle. Focusing on Fig. 15(d), we see good agreement between  $\mathcal{M}_3$  and  $\mathcal{M}_0$  only for the first 20 hours. At later times,  $\mathcal{M}_3$  predicts values for  $f_2$  significantly lower than those predicted by  $\mathcal{M}_0$  and observed experimentally (see Fig. 2). This suggests that the  $C_2$  checkpoint is needed for the model to capture the delayed accumulation of cells in the G2/M phase observed experimentally.

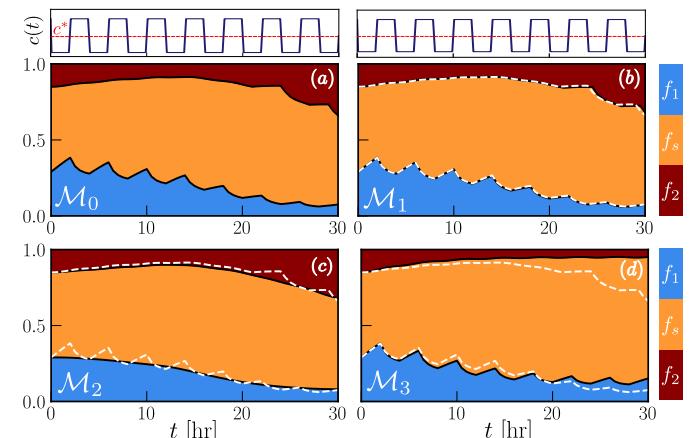


Figure 15: Comparison of the cell-cycle dynamics predicted by solving numerically the models listed in Table 3 for cyclic RH (scenario  $E_2$ ). We here use the characteristic fits obtained for each model (see Table in Appendix C). The white dotted lines in the panels indicate the evolution predicted by model  $\mathcal{M}_0$  (see first panel) here used as a reference. Parameters are chosen as in Fig. 14.

The results from Figs. 14 and 15 highlight the inability of models  $\mathcal{M}_2$  and  $\mathcal{M}_3$  to recapitulate the experimental data. Further, the predictions from  $\mathcal{M}_0$  and  $\mathcal{M}_1$  are indistinguishable, which suggests that the value of the parameter  $K_2$  does not significantly influence the dynamics. Discriminating between models  $\mathcal{M}_0$  and  $\mathcal{M}_1$  is, therefore, not straightforward and requires consideration of other metrics, in addition to how well the model fits the experimental data. These questions will be addressed in the next section where we implement *Bayesian model selection*.

## 6. Parameter fitting and model selection

So far we have presented model predictions for point estimates of the model parameter values. In this section, we explain how such estimates were obtained and investigate how the re-

sults in §5 change when we account for uncertainty in the estimates of “oxygen-dependent” parameters. We start by using the data from the balanced exponential growth experiments ( $E_0$ ) to determine the “oxygen-independent” parameters  $\theta$  using the results from §4. As mentioned in §5.2, all models  $\mathcal{M} \in \mathbb{M}$  reduce to the same set of equations (Eqs. (19)) in normoxia. As such, they share the same value  $\theta$ . We then focus on estimating the remaining parameters  $\Theta$ , which determine the system response to dynamic oxygen conditions. Here we will compare different modelling assumptions by applying model selection methods. We start by fitting each model  $\mathcal{M}_k$  in Table 3 to experimental data from both  $E_1$  and  $E_2$  simultaneously, using Bayesian inference and Monte Carlo methods to estimate the posterior distribution,  $\pi_{ps}$ , for the parameters  $\Theta$  (more details follow). We then select the “best” candidate model by using the deviance information criterion (DIC) as an estimate of model performance and briefly discuss parameter identifiability based on posterior profiles.

### 6.1. Estimation of oxygen independent parameters

We recall from §4, that the regime of unperturbed exponential growth is characterised by the value of the constants ( $\bar{f}_1, \bar{f}_2, \lambda$ ). As explained further in Appendix A, these constants uniquely define the values of the parameters ( $k_1, k_2, v_+$ ) given by Eqs. (A.7). In practice, we can estimate the stationary values  $\bar{f}_1$  and  $\bar{f}_2$  using flow cytometry data. However, additional data are needed to determine the proliferation rate  $\lambda$ . This parameter can be related to the population doubling time,  $T_{doub}$  [8], which is known for most cell lines when cultured in standard media and in the absence of competition (i.e., low confluence). Indeed, it is straightforward to show that  $T_{doub} = \lambda^{-1} \ln 2$ .

Given that, prior to any of the experiments in [5], the cells were cultured at 21%  $O_2$ , with re-plating in order to minimise any effects due to contact inhibition, we assume that initially the cells are undergoing exponential growth. We can, therefore, use the cell fractions data reported at time  $t = 0$  in Fig. 2 to estimate the values of  $\bar{f}_1$  and  $\bar{f}_2$ . For simplicity, we suppose that median values provide a good approximation to the ‘true’ cell fractions so that  $\bar{f}_1 \sim 0.29$  and  $\bar{f}_2 \sim 0.15$ . In this way, we can obtain point estimates for  $\theta$  which facilitates the identification of the remaining model parameters  $\Theta$ . In previous experiments, the doubling time of RKO cells has been estimated to be about 21 hours [65]. The corresponding parameter estimates (obtained

Table 4: Estimates of the cell-cycle parameters for the RKO cell line obtained using the phase stationary solution (PSS) and the experimental data from [5].

| transition rates [1/hr] | average time in the compartment [hr]   | cell fractions (from [5]) |
|-------------------------|--|---------------------------|
| G1 $k_1 = 0.195$        | $\tau_1 \approx k_1^{-1} \approx 5.14$ | $\bar{f}_1 = 0.29$        |
| S $v_+ = 0.083$         | $\tau_S = v_+^{-1} \approx 12$         | $\bar{f}_s = 0.56$        |
| G2/M $k_2 = 0.22$       | $\tau_2 \approx k_2^{-1} \approx 4.54$ | $\bar{f}_2 = 0.15$        |

using Eq. (A.7)) are listed in Table 4 together with estimates of the duration of each cell cycle phase (given by the inverse of the rates of  $k_1, k_2$  and  $v_+$  [8]). We note that the RKO cell line

is characterised by a particularly long S phase. By contrast, the durations of the G1 and G2/M phases are almost half the duration of the S phase, with G2/M being the shortest.

### 6.2. Calibration of the candidate models to time-series flow-cytometry data

The second step concerns estimation of the parameters  $\Theta$  which are associated with the oxygen-dependent mechanisms. Here we use Bayesian inference techniques [41], which allow us to account for measurement errors and to assess uncertainty in the parameter estimates.

Given the small amount of data available, we calibrate the model by pooling the data from all of the experiments in constant RH ( $E_1$ ) and cyclic RH ( $E_2$ ). We therefore postpone model validation until more data will be available. In Fig. 2, for each time point, we reported the mean and standard deviation for multiple repeats of the experiment; however, for the estimation of  $\Theta$  we consider individual experimental measurements, instead of summary statistics. The complete data set can be found in Appendix B. We denote by  $\mathbf{F}^{(i,j)}$  the  $i$ -th measurement performed at time  $t^{(i,j)}$  during experiment  $E_j$  with  $j \in \{1, 2\}$ . Given that  $F_1, F_s$  and  $F_2$  are not independent (recall  $F_s = 1 - F_1 - F_2$ ), we define the values of  $\mathbf{F} = [F_1, F_2] \in \mathbb{R}^2$ . We collect the data in the set  $\mathcal{E} = \{(t^{(i,1)}, \mathbf{F}^{(i,1)})\}_{i=1}^{L_1} \cup \{(t^{(i,2)}, \mathbf{F}^{(i,2)})\}_{i=1}^{L_2}$ , where  $L_1 = 4$  and  $L_2 = 14$  are the number of measurements in sets  $E_1$  and  $E_2$ , respectively.

*The error model.* Details of the error between the experimental measurements and the predictions of the mathematical model  $\mathcal{M}$ , with model parameters  $\Theta$ , are captured in the *likelihood* function  $\mathcal{L}$ . We assume the measurement errors are additive and independent of the oxygen protocol  $c = c(t)$  used in the *in vitro* experiments. We further suppose that the difference between each measurement and model prediction ( $\mathbf{e}^{(i,j)} = \mathbf{F}^{(i,j)} - \mathcal{M}(t^{(i,j)}, \Theta, c_j)$ ) to be independent and normally distributed, with covariance matrix  $\Sigma$ , i.e.,  $\mathbf{e}(t) \sim \mathcal{N}(0, \Sigma)$ ; therefore  $\mathcal{L}$  has the form:

$$\mathcal{L}(\mathcal{E}|\Theta, \Sigma, \mathcal{M}) = \left( \frac{1}{2\pi\sigma_1\sigma_2} \right)^{L_1+L_2} \exp \left[ -\frac{\chi^2}{2} \right], \quad (23a)$$

$$\chi^2 = \sum_{j=1}^2 \sum_{i=1}^{L_j} \mathbf{e}^{(i,j)} \cdot (\Sigma^{-1} \mathbf{e}^{(i,j)}). \quad (23b)$$

Here the covariance matrix  $\Sigma$  is considered to be constant in time and diagonal (i.e.,  $\Sigma = \text{diag}(\sigma_1^2, \sigma_2^2)$ ), as we assume independence between errors in the fractions  $F_1$  and  $F_2$ . Instead of specifying the values of  $\sigma_1$  and  $\sigma_2$ , we treat them as unknown parameters that are learnt from the data. Since the cell fractions are normalised, we assume that  $\sigma_1$  and  $\sigma_2$  are uniformly distributed in  $[0, 1]$  ( $\pi_{pr}(\Sigma) \sim \mathcal{U}(0, 1) \times \mathcal{U}(0, 1)$ ).

*Calibration and model selection.* In brief, given a mathematical model  $\mathcal{M} \in \mathbb{M}$ , the main goal of Bayesian inference is to sample, via Monte Carlo methods, from the posterior probability distribution for the model ( $\Theta$ ) and error parameters ( $\Sigma$ ) conditioned to the data ( $\mathcal{E}$ ) (i.e.,  $\pi_{ps}(\Theta, \Sigma|\mathcal{E}, \mathcal{M})$ ). Using Bayes’

Theorem, the posterior distribution can be expressed in terms of the *likelihood* function ( $\mathcal{L}(\mathcal{E}|\Theta, \Sigma, \mathcal{M})$ ) and the prior distributions ( $\pi_{pr}(\Theta)$ ,  $\pi_{pr}(\Sigma)$ ):

$$\begin{aligned}\pi_{ps} &= \frac{\pi_{pr}(\Theta)\pi_{pr}(\Sigma)\mathcal{L}(\mathcal{E}|\Theta, \Sigma, \mathcal{M})}{\pi(\mathcal{E}|\mathcal{M})} \\ &\sim \pi_{pr}(\Theta)\pi_{pr}(\Sigma)\mathcal{L}(\mathcal{E}|\Theta, \Sigma, \mathcal{M}),\end{aligned}\quad (24a)$$

where  $\pi(\mathcal{E})$  is a normalising factor,  $\pi_{pr}$  is the prior knowledge on parameter values, and  $\mathcal{L}$  is as defined in (23b).

In the absence of prior information about the values of model parameters  $\Theta$ , we assume that each component of the vector  $\Theta$  (i.e.,  $\Theta_i$  for  $i = 1, \dots, \kappa$ ), is uniformly distributed on the intervals  $[\Theta_i^{min}, \Theta_i^{max}]$ , where the extremes of the intervals are taken from Table 2; hence  $\pi_{pr}(\Theta)$  is given by:

$$\pi_{pr}(\Theta) = \prod_{i=1}^{\kappa} (\Theta_i^{max} - \Theta_i^{min})^{-1}. \quad (24b)$$

Due to the dimensions of our parameter space, it is not possible to compute Eqs. (24a) analytically. Instead we rely on Markov Chain Monte Carlo (MCMC) methods to estimate  $\pi_{ps}$  using the python package PINTS (Probabilistic Inference on Noisy Time-Series) for Bayesian inference [19]. More details on the numerical technique are included in Appendix C.

For model selection among the class of models  $\mathbb{M}$ , we compute the deviance information criterion (DIC) [28, 41]. The DIC represents a trade-off between model complexity (as measured by the over-fitting bias  $k_{DIC}$ ) and model accuracy (as measured by the likelihood). Given a model  $\mathcal{M}$ , the DIC is defined as follows [41]:

$$DIC(\mathcal{M}) = 2k_{DIC} - 2\ln(\hat{\mathcal{L}}(\mathcal{M})), \quad (25a)$$

$$\text{with } k_{DIC} = 2\text{Var}[\ln \mathcal{L}|\mathcal{E}, \mathcal{M}], \quad (25b)$$

where  $\hat{\mathcal{L}}(\mathcal{M}) = \mathcal{L}(\mathcal{E}|\hat{\Theta}, \hat{\Sigma}, \mathcal{M})$  is the likelihood value estimated at the expected value of the parameters, (i.e.,  $\hat{\Theta} = \mathbb{E}[\Theta|\mathcal{E}, \mathcal{M}]$  and  $\hat{\Sigma} = \mathbb{E}[\Sigma|\mathcal{E}, \mathcal{M}]$ ) and  $\text{Var}$  is the variance of the log-likelihood,  $\ln \mathcal{L}$ , here approximated via sampling from the estimated posterior. We note that the DIC penalises models with large values of  $k_{DIC}$ , to account for the fact that complex models are more likely to fit data well. At the same time, more complex models tend to have more parameters which can lead to higher posterior uncertainty if the model is too complex for the data (i.e., it is over-fitted) [41]. This results in a larger variability in the expected log-likelihood (i.e., larger  $k_{DIC}$ ). In other words, the term  $k_{DIC}$  in Eq. (25a) corrects for over-fitting. When comparing models, we are interested in the relative value of the DIC, and favour the model with the smaller DIC.

For a model  $\mathcal{M}$ , predictive posterior estimates are obtained by sampling 800 parameter sets,  $(\Theta^{(i)}, \Sigma^{(i)})$ , from the estimated posterior  $\pi_{ps}$ . For each set, we run the model  $\mathcal{M}_k$  forwards to generate 800 predictive curves for each phase fraction  $f_m^{(i)}(t)$  with  $m \in \{1, 2\}$  and  $i = 1, \dots, 800$ . This gives posterior distributions for the “true” cell fraction. To obtain posterior distributions for the *measured* cell fractions  $F_{1,2}^{(i)}(t)$ , we add to the simulated cell fractions  $f_1^{(i)}(t)$  and  $f_2^{(i)}(t)$  the corresponding

measurement errors  $\mathbf{e}(t) \sim \mathcal{N}(0, \Sigma^{(i)})$ . We then estimate  $F_s^{(i)}$  as  $F_s^{(i)} = 1 - F_1^{(i)} - F_2^{(i)}$  and  $f_s^{(i)}$  as  $f_s^{(i)} = 1 - f_1^{(i)} - f_2^{(i)}$ . At each time point, we compute the mean of the 800 predictive curves (for either the “true” or measured fractions) and the corresponding 68%- and 95%-confidence intervals. For the plots in §5 and §5.2, we used the expected values  $\hat{\Theta} = \mathbb{E}[\Theta|\mathcal{E}, \mathcal{M}]$  as representative of characteristic model fits. Additional results on the estimated posterior distributions can be found in Appendix C.

### 6.2.1. Numerical results

The estimated DICs for the models  $\mathcal{M} \in \mathbb{M}$  are reported in Table 5. Based on these estimates, model  $\mathcal{M}_1$  has the lowest DIC and is, therefore, the “best” model in the class  $\mathbb{M}$ . As expected from the results presented in §5.2 (see Fig. 14 and 15),  $\mathcal{M}_2$  and  $\mathcal{M}_3$  have the worst performance. In this case, the difference in the estimated DICs is rooted in the value of  $\hat{\mathcal{L}}(\mathcal{M})$ , which is markedly reduced for  $\mathcal{M}_2$  and  $\mathcal{M}_3$ , suggesting that the two models fail to fit the experimental data. We conclude that including the  $C_1$  and  $C_2$  compartments (i.e. cell-cycle arrest in the G1 and G2 phases) is necessary for our model to reproduce the experimental cell-cycle dynamics. However, we can not exclude the possibility that other mechanisms (i.e., modelling assumptions) not considered here, might also explain the data.

Table 5: Comparison based on the deviance information criterion (DIC) of the models in class  $\mathbb{M}$  that were introduced in §5.2. The last column indicates the relative *DIC* score with respect to model  $\mathcal{M}_0$ , i.e.,  $\Delta\text{DIC} = \text{DIC}(\mathcal{M}) - \text{DIC}(\mathcal{M}_0)$ .

|                 | $k_{DIC}$ | $\ln(\hat{\mathcal{L}}(\mathcal{M}))$ | DIC     | $\Delta\text{DIC}$ |
|-----------------|-----------|---------------------------------------|---------|--------------------|
| $\mathcal{M}_0$ | 6.10      | 65.73                                 | -119.27 | 0                  |
| $\mathcal{M}_1$ | 5.13      | 65.33                                 | -120.41 | -1.73              |
| $\mathcal{M}_2$ | 6.04      | 43.83                                 | -75.58  | 43.62              |
| $\mathcal{M}_3$ | 4.25      | 45.84                                 | -83.18  | 37.35              |

Focusing now on  $\mathcal{M}_0$ , we see that its performance is similar to that of  $\mathcal{M}_1$ , whose DIC is only slightly smaller. While the two models yield similar values of  $\hat{\mathcal{L}}(\mathcal{M})$ , the larger estimated value of  $k_{DIC}$  for  $\mathcal{M}_0$  is the main source of the discrepancy. This suggests that the additional complexity due to the introduction of the parameter  $K_2$  is not balanced by a sufficient improvement in the description of the data. This interpretation is confirmed by the profile of the marginal posterior for  $K_2$  (see Fig. 16(a)). The latter is approximately uniform, suggesting that its value can not be identified, given the available data. As shown in Fig. 16(b), the posterior predictions for  $f_2$  have larger confidence intervals at longer times. This suggests that experiments run over longer time periods could improve the estimates of parameter  $K_2$ . In the absence of such data, we view model  $\mathcal{M}_1$  as the best candidate.

*Candidate model  $\mathcal{M}_1$  is in good agreement with the experimental data.* Our analysis shows that model  $\mathcal{M}_1$  gives the best fit to the experimental data. Referring to the posterior predictions in Fig. 17, we note that  $\mathcal{M}_1$  captures the experimental data from both experiments  $E_1$  and  $E_2$ , with all experimental data points falling within the 95%-confidence interval of the posterior predictions for the “measured” fractions,  $F_1$ ,  $F_s$  and  $F_2$ .

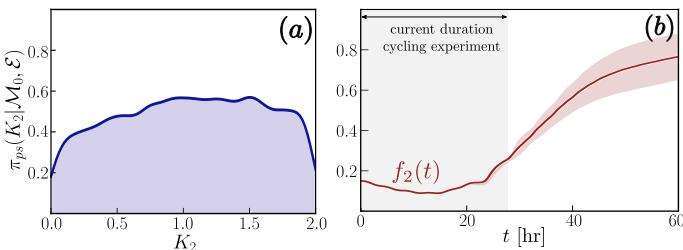


Figure 16: Analysis of the fitting of model  $\mathcal{M}_0$ : (a) marginal posterior distribution for parameter  $K_2$ ; (b) posterior prediction of the time evolution of  $f_2$ , i.e., the fraction of cells in the  $P_2$  phase. The mean (red curve) and 68%-confidence intervals (red shaded area) are computed as outlined in the section on “Calibration and model selection”.

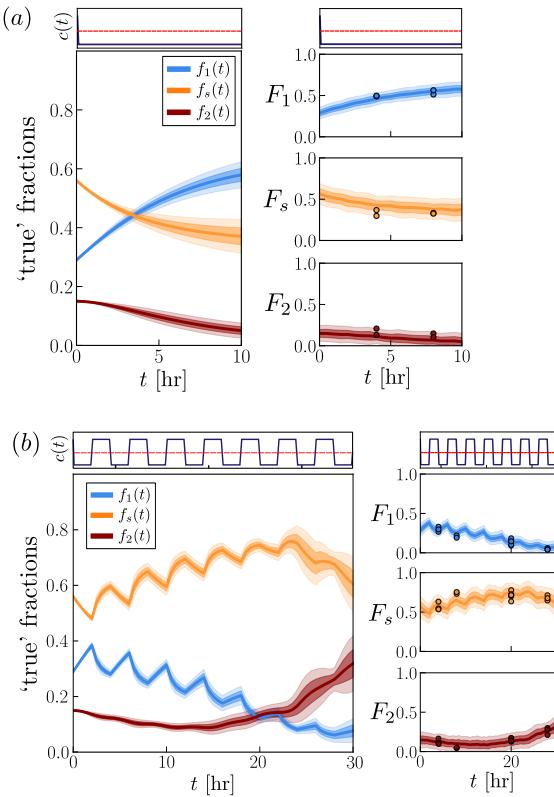


Figure 17: Posterior prediction distribution of the selected model  $\mathcal{M}_1$  for (a) chronic RH and (b) cyclic RH. For both scenarios we plot the predicted ‘true’ cell fractions  $f_m(t)$  and the predicted measured fractions  $F_m(t)$  with  $m \in \{1, s, 2\}$ ; the predicted evolution of  $F_m$  is compared with the experimental data (dots). For each model output, we plot the expected values, together with the 68%- and 95%- confidence intervals, as indicated by the dark and light shaded areas, respectively. The top panels indicate the prescribed oxygen tension  $c(t)$  associated with the two experiments.

Fig. 18 shows that  $\mathcal{M}_1$  can also capture the qualitative shape of the flow cytometric distribution ( $D_f$ ) reported in [5], which indicates cell number over PI-fluorescence intensity (see Fig. 1). Here we estimate  $D_f$  using output from the numerical solution of Eqs. (17) as follows: based on [10], we assume that a population of cells with the same DNA content  $x$  gives rise to a Gaussian-like flow cytometric output (see Fig 1); conse-

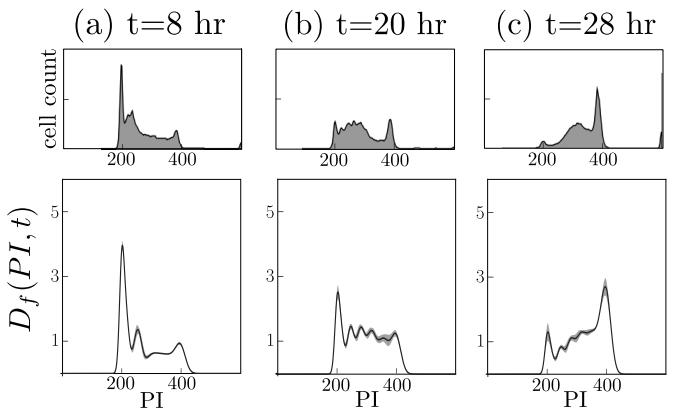


Figure 18: Posterior prediction for the distribution  $D_f$  as defined by Eq. (26a) for cyclic RH ( $E_2$ ). The dark line indicates the average over 800 parameter sets sampled from the estimated posterior, while the shaded grey area indicates the 68% confidence interval. At each time point, we compare the theoretical prediction (bottom panels) with the experimentally measured distribution from Fig. 3 in Bader et al. [5] (top panels). Fluorescence readings of  $PI = 200$  and  $PI = 400$  correspond, respectively, to  $x \approx 1$  and  $x \approx 2$ . This comparison is qualitative, as the vertical axes for the theoretical (bottom panels) and experimental (top panels) distributions are different since  $D_f$  is re-scaled by the initial number of cells  $N_0$ , while the experimental output is in terms of absolute cell counts.

quently, we model the flow cytometric output  $D_f$  as

$$D_f(PI, t) = \int_1^2 f(PI|x) \mathcal{D}(x, t) dx, \quad (26a)$$

where

$$\mathcal{D}(x, t) = P_1(t) \delta(x - 1) + S(x, t) + P_2(t) \delta(x - 2), \quad (26b)$$

where  $\delta(x - y)$  is the delta function, (i.e.,  $\delta = 1$  if  $x = y$  and  $\delta = 0$  otherwise). The term  $f(PI|x)$  in Eq. (26a) is the probability of recording a fluorescence intensity  $PI$  for a cell with DNA content  $x$  and has the form

$$f(PI|x) = \frac{1}{\sqrt{2\pi\alpha\gamma(x)}} \exp\left[-\frac{(PI - \alpha x)^2}{2\alpha^2\gamma(x)^2}\right]. \quad (26c)$$

In Eq. (26c), we assume that the mean PI fluorescence is proportional to the DNA content  $x$ , with constant of proportionality  $\alpha = 200$  PI, so that  $x = 1$  corresponds to a fluorescence intensity  $PI = 200$ . Following [10] again, we suppose that the variance  $\gamma$  depends linearly on  $x$  (i.e.,  $\gamma = (\gamma_2 - \gamma_1)(x - 1) + \gamma_1$ , with  $\gamma_{1,2}$  being the variances associated with DNA contents of  $x = 1$  and  $x = 2$ , respectively). In general,  $\gamma_1$  and  $\gamma_2$  will depend on the cell line of interest. Since here we are interested in qualitative comparisons, we set  $\gamma_1 = 0.04$  and  $\gamma_2 = 0.08$ , for consistency with the estimates from [10].

While the calibration only uses information about cell fractions, we see that our theoretical estimates for  $D_f$  are in good agreement with the experimental observations (see Fig. 18). However, there are some discrepancies, particularly at time  $t = 20$  hr. The model predicts a higher percentage of cells in the late S phase, so that the peak corresponding to  $PI = 400$  is not isolated; in contrast the experimental profile tends to flatten in the vicinity of  $PI \approx 350$ , so that the peak at  $PI \approx 400$  is isolated.

However, we note that the experimental observations relate to only one realisation, whereas we illustrate the expected profile over several model realisations.

*Model Identifiability.* We end our discussion on model calibration by briefly considering practical identifiability of the unknown parameters. Following [23, 35], we define a parameter as practically identifiable if we can constrain its value to a reasonably small region of parameter space, *i.e.*, the posterior distribution has compact support.

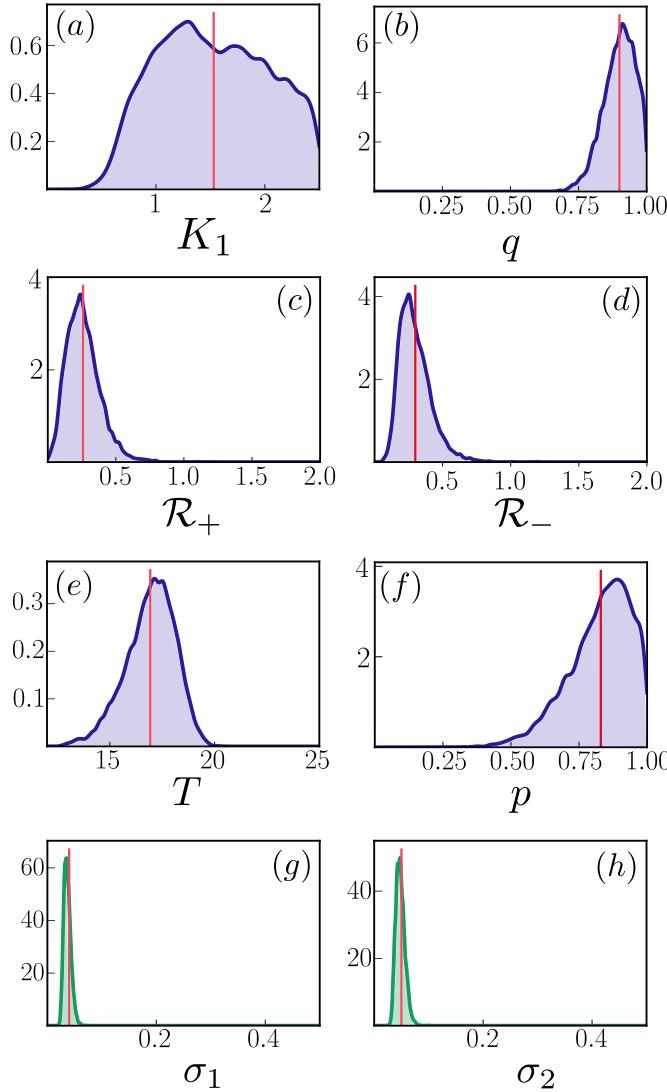


Figure 19: Marginal posterior distributions  $\pi_{ps}(\Theta|\mathcal{M}_1, \mathcal{E})$  for the model parameters  $\Theta \in \{K_1, q, \mathcal{R}_+, \mathcal{R}_-, T, p\}$  and the error parameters  $\pi_{ps}(\sigma_i|\mathcal{M}_1, \mathcal{E})$  for  $i = 1, 2$ . The red vertical line indicates the mean value as reported in Table C.9. We see that for most parameters the posterior distribution has a compact support. This is not the case for  $K_1$  (see (a)) where the large support of the posterior is a sign of practical non-identifiability.

Looking at the profile of the marginal posterior distributions for the different parameters (see Fig. 19), we see that the marginal posteriors have a bell-like shape, with a unique, well-defined, maximum for most parameters. However, the posterior distribution for parameter  $K_1$  tends to flatten at large values,

where  $K_1 > 1$  (see Fig. 19(a)). This indicates greater uncertainty in the estimation of  $K_1$  for which we can identify only a lower, but not an upper, bound. From Eq. (8a), we note that  $K_1$  is only relevant in data from experiment  $E_2$  (*i.e.*, when oxygen levels  $c$  are above the threshold  $c^*$ ). If we consider, for example, the first time at which this happens (*i.e.*, time  $t_1^{(R)} = 2\text{hr}$ ), then, over the period  $t_1^{(R)} = 2\text{hr} < t < t_2^{(H)} = 4\text{hr}$ ,  $Q(t) \approx 0$  and the evolution of  $C_1(t)$  can be determined explicitly by solving Eq. (17b) to obtain:

$$C_1(t) = C_1(t_1^{(R)}) e^{-K_1(t-t_1^{(R)})}, \quad t \in (t_1^{(R)}, t_2^{(H)}). \quad (27)$$

Thus, when the first measurement is taken,  $C_1(t_2^{(H)}) = C_1(t_1^{(R)}) e^{-2K_1}$ . If  $K_1 > 1$ , then  $e^{-2K_1} \ll 1$  and compartment  $C_1$  rapidly empties after re-oxygenation. Therefore, unless  $C_1$  is very large,  $C_1(t_2^{(H)}) \approx 0$ , independently of the value of  $K_1 > 1$ . In order to resolve this fast time-scale we would need to collect experimental data at an earlier time point, say  $t^*$ , for which  $t^* - t_1^{(R)} < K_1^{-1}$ ; alternatively, we could choose an oxygen cycle for which a larger number of cells accumulate in the  $C_1$  compartment. This could be achieved by prolonging the period for which the cells are exposed to severe hypoxia (*i.e.*,  $c < c^*$ ).

## 7. Model predictions in the presence of uncertainty

To conclude, we use the calibrated model from §6 to make predictions on cell-cycle dynamics in different environmental conditions. We start by considering the oxygen cycles in Fig. 13 but now account for uncertainty in our parameter estimates.

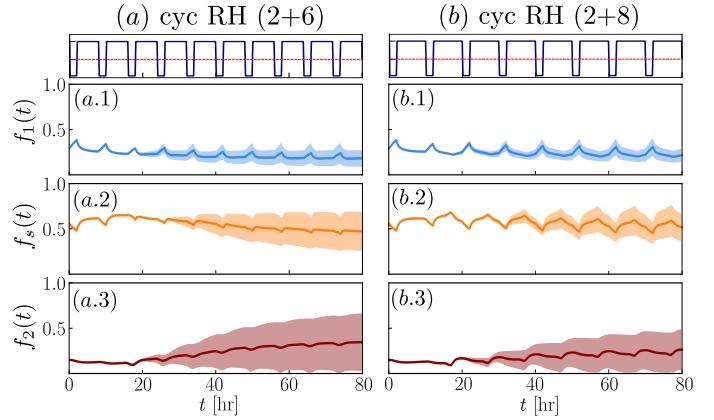


Figure 20: Posterior estimates for the cell fractions  $f_m$  for the two cyclic protocols considered in Fig. 13: (a) cyclic RH with 2 hr in RH + 6 hr in an oxygenated environment; (b) cyclic RH with 2 hr in RH + 8 hr in an oxygenated environment. We plot mean estimates and indicate the 68%-confidence intervals by the shaded regions.

For the 2hr+6hr cycle in Fig. 20(a), in the absence of uncertainty, we predicted a systematic increase in the fraction of cells in the S phase, with no activation of the  $C_2$  compartment (see Fig. 13(a)). When uncertainty is taken into account, we find large variability in model predictions at longer times ( $t > 25$  hr). In particular, the 68%-confidence interval encompasses the possibility of  $f_s(t)$  both increasing or decreasing compared to its initial value when  $t \gg 1$ . Further, activation of  $C_2$  cannot

be ruled out, as indicated by the value of  $f_2(80)$ , which ranges between  $0 \lesssim f_2(80) \lesssim 0.6$  (see Fig. 20(a.3)). Despite the large uncertainty in the value of  $f_2(t)$ , the confidence interval on  $f_1(t)$  remains reasonably small. We observe similar behaviour for cycles with a longer re-oxygenation phase (see 2hr+8hr cycle in Fig. 20(b)). Again, uncertainty in  $f_s$  and  $f_2$  increases over time, even though it is less pronounced than in Fig. 20(a). Again, the 68%-confidence interval allows for the possibility of activation of the  $C_2$  checkpoint (see increase in  $f_2$  as  $t \gg 1$ ) even though this is less probable than in the scenario depicted by Fig. 20(a). For both oxygen cycles considered, the uncertainty in the cell-cycle distribution is reflected in the predictions for the number of cells  $N(t)$  (see Fig. 21). This is particularly evident in Fig. 21(a) where  $N(80)$ , the number of cells at the final time,  $t = 80$  hr, ranges between  $3.5 \lesssim N(80) \lesssim 10$ .

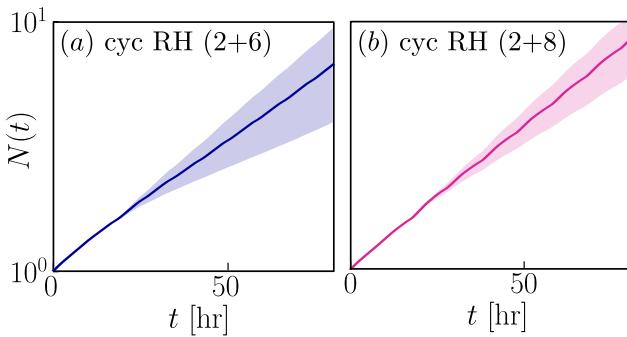


Figure 21: Posterior estimates for the total cell number  $N(t)$  for the two cyclic protocols in Fig. 20: (a) cyclic RH with 2 hr in RH + 6 hr in an oxygenated environment; (b) cyclic RH with 2 hr in RH + 8 hr in an oxygenated environment. We plot mean estimates and indicate the 68%-confidence intervals by the shaded regions.

Based on the results in Fig. 20, we conclude that we can use our calibrated model to predict cell-cycle dynamics on short time scales ( $0 < t < 25$  hr); thereafter the increased uncertainty prevents us from making reliable predictions for the cell-cycle distribution at longer times. However, the predicted uncertainty is still informative when considering experimental design. In this case, the objective is twofold. On the one hand, we want to identify experiments that could facilitate model validation. From this point of view, we focus on times at which the uncertainty in our predictions is small (e.g., on the short time scale for experiments in Fig. 20). On the other hand, we aim also to propose experiments that could improve the accuracy of model parameter estimates. In this case, our attention focuses on scenarios where uncertainty in the model predictions is large and, therefore, new measurements can refine parameter estimates (such as for the long time dynamics in Fig. 20). From this point of view, a scenario like Fig. 20(a) is adequate since it can account for both the validation and refinement steps. Our model also suggests that, if such experiments were performed, information about the total number of cells  $N(t)$  may improve model calibration, given the large variation predicted in its value at the end of treatment.

Experimentally, it might be difficult to study scenarios for which oxygen levels change on timescales faster than two hours due to the time required for oxygen levels to equilibrate *in vitro*

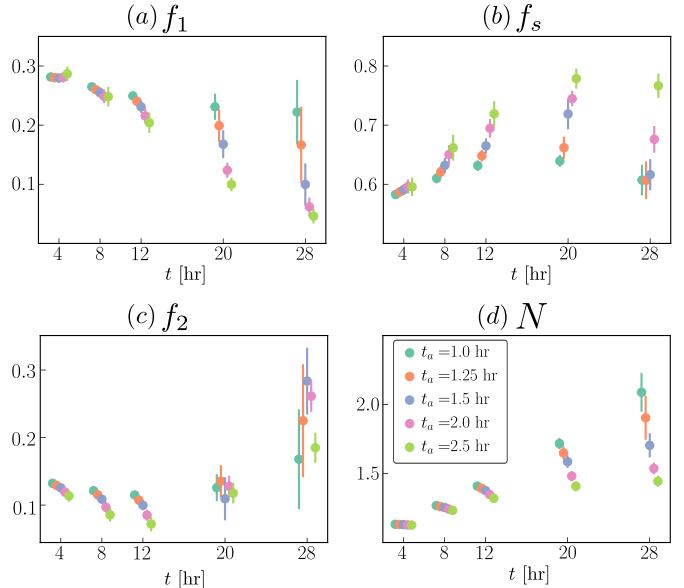


Figure 22: Numerically-estimated values of the cell fractions  $f_m$  and cell number  $N$  for different cyclic protocols. We report values of the variables at 5 time points; for each variable we report the expected value based on the estimated posterior (coloured dots) and the corresponding 68%-confidence interval indicated by the vertical lines. Different colours correspond to different oxygen dynamics, *i.e.*, a different choice of  $t_a$ .

[50]. This, however, is not a limitation of our model. Indeed, we can consider what happens when a 4 hour cycle (as in  $E_2$ ) is split into asymmetric periods of RH and re-oxygenation by setting  $t_i^{(R)} = t_i^{(H)} + t_a$  and  $t_i^{(H)} = 4(i-1)$  in Eq. (21), where  $t_a \in (0, 4)$ . If  $t_a > 2$ , then the cells spend more time exposed to low oxygen levels ( $c < c^*$ ) than to normal values ( $c > c^*$ ); the opposite holds when  $0 < t_a < 2$ . We run numerical simulations up to  $t = 28$  hr and report predictions of the true cell fractions,  $f_m$  with  $m \in \{1, s, 2\}$  and total cell number,  $N$ , at 5 time points.

The results from these numerical simulations are summarised in Fig. 22. We observe that, at all time points considered, the larger the value of  $t_a$  the lower is the total number of cells,  $N$ , (see Fig. 22(d)). When  $t_a = 2.5\text{hr}$  (see light green dots), DNA synthesis is so slow that at the final time ( $t = 28$  hr) only a small fraction of cells has completed duplication and entered the  $G2/M$  phase. As a result, the fraction  $f_s$  is larger than for the smaller values of  $t_a$  while the fraction  $f_2$  is smaller. In this case, the variability in the model predictions remains small (even smaller than for the case  $t_a = 2$  hr which was used for the *in vitro* experiments). Therefore, our model predicts that, on short time scales ( $\approx 30$  hr), cycles with longer periods of RH than re-oxygenation favour the accumulation of cells in the S phase. In particular, exposing cells to cycles with  $t_a = 2.5$  for 20 hours may be sufficient to synchronise them in the S phase of the cell-cycle. By contrast, when we decrease the value of  $t_a$ , our model predicts only a 5% increase in the fraction of cells in the S phase. In line with our intuition, as  $t_a$  decreases, fewer cells are impacted by cyclic levels of oxygen and, therefore, their distribution is more concentrated at specific phases of the cell-cycle. We note, however, that for smaller values of  $t_a$  (e.g.,  $t_a = 1$  or  $t_a = 1.25$ ), there is greater uncertainty in the predic-

tions of  $f_1$  and  $f_2$  at  $t = 28$  hr, due to the uncertainty in whether the  $C_2$  checkpoint will be activated or not (results not shown). This, again, hints at the need to refine our parameter estimates by performing new experiments (such as the cycle experiment in Fig. 20) to obtain more accurate predictions for environmental conditions that deviate from those used here to calibrate our model.

## 8. Discussion and Future Work

In this paper, we have presented a five-compartment model for the cell-cycle which accounts for cell response to variable oxygen levels. We have focussed on the impact of dNTP shortages (in conditions of radio-biological hypoxia) on rates of DNA synthesis. This was achieved by introducing a variable rate,  $v(t)$ , of DNA synthesis and allowing transient cell-cycle arrest in the late G1 phase by transition to a checkpoint compartment  $C_1$ . A second checkpoint compartment,  $C_2$ , accounts for cell-cycle arrest in the G2 phase due to accumulation of replication stress and damage. Under constant oxygen-rich conditions the model reduces to a linear, three-compartment model and analytic expressions for the long time dynamics can be derived (see §4). This analysis predicts that, in the absence of competition, cells evolve to a regime of balanced exponential growth, a result which is consistent with other cell-cycle models [8, 10, 13, 21, 54]. The main novelty of the work is the investigation of the cell-cycle dynamics in cyclic hypoxia (see §5). We show first that the model can recapitulate the experimental data from [5]. We then explore different oxygen dynamics and, in so doing, show different ways in which cyclic hypoxia can dis-regulate cell-cycle dynamics, and lead to a redistribution of cells across the phases of the cell cycle. Further, we identify scenarios in which cyclic hypoxia leads to complete inhibition of proliferation and scenarios in which proliferation is only slowed down. This is of relevance when thinking about cell-cycle specific treatment, for which changes in cell-cycle distribution (even if they are transient), can have a large impact on treatment efficacy. However, in order to use our model as a predictive tool, accurate and robust predictions are needed. In the remainder of the paper, we therefore showed how our modelling framework can be used to predict cell-cycle dynamics and inform the design of *in vitro* experiments (see §3). We started by deriving a class of candidate models  $\mathcal{M}$  based on Eqs. (17) by systematically decreasing model complexity (i.e., the number of unknown parameters). Here our aim was to test different assumptions on the mechanisms driving cell response to cyclic hypoxia. In §6, we used Bayesian model selection to identify the best candidate model from  $\mathcal{M}$  and showed that this model can indeed recapitulate the dynamic data from [5]. Furthermore, by constructing our class of models and applying Bayesian modelling selection we were able to systematically show that the inclusion of the  $C_1$  and  $C_2$  compartments in our model is necessary for it to capture the cell-cycle dynamics observed experimentally. In §7, we used our calibrated model to revisit the results from §5 where we account for uncertainty in parameter estimates. While the model makes precise predictions on short time scales ( $t \approx 30$  hr), we observed a large uncer-

tainty in the cell-cycle dynamics at longer times. We therefore discussed how our model could be used to inform the efficient and effective design of future experiments to refine our parameter estimates, as well as to validate predictions of the calibrated model.

In this work we showed that our model can recapitulate the response to cyclic hypoxia of a specific cancer cell line (i.e., RKO cancer cell line). For this cell line, our model predicts that both constant and cyclic radio-biological hypoxia (RH) perturb the cell-cycle dynamics, but in different ways. In constant RH, cells tend to accumulate in the late G1 phase and proliferation is rapidly altered. During cyclic RH, we predict instead a more diverse range of responses depending on the oxygen dynamics. We can identify regimes where proliferation is inhibited due to accumulation of cells in the G2/M phase. By contrast, when the duration of the re-oxygenation phase is increased, population growth is only mildly slowed down and the fractions of cells in S phase is up-regulated. It would be interesting to test whether our findings can be generalised to other cell lines and how much variability there is in their response to similar cyclic protocols. Based on these observations, an interesting future research direction emerges, namely, investigating the role that cyclic RH can have on the response to cell-cycle dependent treatment, such as radiotherapy. Given that cyclic RH can change the distribution of cells around the phases of the cell-cycle, we expect a differential response to radiotherapy. This could be investigated by extending our model to include radiotherapy and to account for changes in radio-sensitivity in different phases of the cell-cycle. From this point of view, a natural question is whether cell-cycle redistribution is sufficient to explain the increase in radio-resistance due to cyclic hypoxia as reported in the literature [35, 36, 37].

In this paper, our focus was on constructing a minimal model to describe the influence of cyclic hypoxia on the cell-cycle which could be validated against existing experimental data. This guided our assumption that the rates  $\mathcal{R}_\pm$  at which cells adjust their rates of DNA synthesis,  $v$ , to be constant. In practice, these rates may also depend on the level of damage and replication stress accumulated by the cell during cyclic hypoxia. Our model could easily be extended to account for these effects but at the cost of increasing the number of unknown model parameters. *In silico* hypothesis testing (using a Bayesian framework, as in §6), could be used to compare different modelling assumptions (i.e., constant vs variable rates), and to investigate the design of future experiments that can distinguish between the alternative mathematical models.

In several instances we have mentioned that DNA damage plays a key role in mediating cell-cycle progression and cell-cycle arrest in cyclic hypoxia. However, for simplicity, we have accounted for it only implicitly. In principle, our model could be extended by introducing DNA damage as an additional structural variable to describe the cell state. While being more realistic (and of interest from a mathematical perspective), the increased model complexity would make it difficult to fit the model to experimental data. Moreover, introducing DNA damage into the model would enable us to account for radiotherapy in a more realistic manner. Analogously to re-oxygenation, ra-

diotherapy also causes DNA damage. Such a model extension could be used to investigate whether cyclic hypoxia selects for radio-resistant clones, which are less sensitive to damage accumulation. Further model extensions could account for spatial heterogeneity, and bring the model closer to *in vivo* conditions. In this light, we view our work as a first step towards developing a theoretical framework for investigating cyclic hypoxia and its effect on cell-cycle progression, particularly in the context of solid tumours.

## Appendix A. Exponential steady state uniqueness

Starting from Eqs. (19), we note that the evolution of the number of cells,  $N$ , decouples from the equations for  $G_1$  and  $G_2$  so that we can consider the reduced model:

$$\frac{dG_1}{dt} = 2k_2G_2 - k_1G_1, \quad (\text{A.1a})$$

$$\frac{dG_2}{dt} = -k_2G_2 + k_1G_1(t - \tau_s). \quad (\text{A.1b})$$

Since Eqs. (A.1) form a system of linear delay differential equations, their solutions can be written as a superposition of exponential functions  $e^{\Lambda t}$  with corresponding eigenvalue  $\Lambda \in \mathbb{C}$ . We therefore know that the long time behaviour of the system will be dominated by the eigenvalue  $\Lambda$  with largest real part, here denoted by  $\lambda$ . Unlike for ordinary differential equations, the number of eigenvalues  $\Lambda$  for a delay differential equation is infinite and they are defined by the characteristic equation:

$$\det(A(\Lambda) - \Lambda \mathbb{I}) = 0, \quad A(\Lambda) = \begin{bmatrix} -k_1 & 2k_2 \\ k_1 e^{-\Lambda \tau_s} & -k_2 \end{bmatrix} \quad (\text{A.2})$$

where  $\mathbb{I}$  is the identity matrix in  $\mathbb{R}^{2 \times 2}$ . Evaluating the determinant explicitly we obtain the following transcendental equation, whose roots correspond to the eigenvalues  $\Lambda$  of Eqs. (A.1):

$$P(\Lambda) \equiv -(\Lambda + k_1)(\Lambda + k_2) + 2k_1k_2e^{-\Lambda \tau_s} = 0. \quad (\text{A.3})$$

It can be shown that this system has at least one root with  $\text{Re}(\Lambda) > 0$  so that the solution will blow up in time, i.e.  $\lim_{t \rightarrow \infty} G_1, G_2 = \infty$ . We also have that the spectrum of eigenvalues is bounded above, in the sense that there exists an upper limit to the values of  $\text{Re}(\Lambda)$ . This is analogous to the system investigated by Crivelli et al. in [21]. Lemma 1 summarises some of their results as adapted to our model.

**Lemma 1.** *For any value of  $k_1 > 0$ ,  $k_2 > 0$  and  $\tau_s > 0$ , the right-most root of  $P(\Lambda)$  as defined by Eq. (A.3) is real and positive.*

*Proof.* Let us first consider the existence of a real and positive root  $\Lambda = \Lambda_0$ . This is straightforward to prove since  $P(\Lambda)$  is continuous and  $P(0) > 0$  while  $\lim_{\Lambda \rightarrow \infty} P(\Lambda) < 0$ . Given that for  $\Lambda > 0$ ,  $dP/d\Lambda < 0$ , i.e.  $P$  is strictly monotonically decreasing, we have that the zero  $\Lambda = \Lambda_0$  is unique. This implies that for any choice of parameters, the trivial steady state  $(0, 0)$  is unstable.

Let us now consider the complex solution  $\Lambda = \Lambda_R + i\Lambda_I$  for the function  $P$  where  $\Lambda_I \in \mathbb{R} \setminus \{0\}$ :

$$(\Lambda_R + i\Lambda_I + k_1)(\Lambda_R + i\Lambda_I + k_2) = 2k_1k_2e^{-(\Lambda_R + i\Lambda_I)\tau_s}. \quad (\text{A.4})$$

Taking absolute values of the above we obtain that:

$$\begin{aligned} 2k_1k_2e^{-\Lambda_R\tau_s} &= \sqrt{((\Lambda_R + k_1)^2 + \Lambda_I^2)((\Lambda_R + k_2)^2 + \Lambda_I^2)} \\ &\Rightarrow 2k_1k_2e^{-\Lambda_R\tau_s} > (\Lambda_R + k_1)(\Lambda_R + k_2), \end{aligned} \quad (\text{A.5})$$

which implies that  $P(\Lambda_R) > 0 = P(\Lambda_0)$ , where  $\Lambda_0$  is the unique real root of  $P(\Lambda)$ . Since  $P$  is strictly monotonically decreasing we therefore have that  $\Lambda_R > \Lambda_0$ . We therefore have that the rightmost eigenvalue is real and it is positive.  $\square$

Based on Lemma 1, we know that for any choice of model parameters, under unperturbed growth, cells will eventually reach a regime in which they grow exponentially. This is a common result of many cell-cycle models for *in vitro* systems and it is usually referred to as balanced or asynchronous exponential growth:

$$G_{1,2}(t) \approx \xi_{1,2}e^{\bar{\lambda}t}, \quad t \gg 1, \quad (\text{A.6a})$$

where  $\xi_{1,2}$  are positive constants and the character  $\bar{\lambda}$  is used to indicate the asymptotic solution. Having characterised the long time behaviour of  $G_1(t)$  and  $G_2(t)$ , let us discuss what this implies for the other model variables, i.e. the distribution  $S(x, t)$  and the total number of cells,  $N$ . Using Eq. (9) we find that, in the case of unperturbed growth and assuming  $t > \tau_s$ , the long time distribution  $S(x, t)$  takes the form:

$$S(x, t) \approx \tau_s k_1 \frac{\xi_1}{\xi_n} e^{-\bar{\lambda} \tau_s (x-1)} \xi_n e^{\bar{\lambda} t}, \quad t \gg 1, \quad (\text{A.6b})$$

which can be written by separating the DNA and time components, as  $S(x, t) \approx \bar{s}(x) \xi_n e^{\bar{\lambda} t}$ . This implies that the population  $P_s$  also grows exponentially,  $P_s(t) = \xi_s e^{\bar{\lambda} t}$  where  $\xi_s = \int_1^2 \bar{s}(x) dx$ . Combining this with Eq. (A.6a), we can compute the total number of cells and the cell fractions:

$$N(t) \approx (\xi_1 + \xi_2 + \xi_s) e^{\bar{\lambda} t} = \xi_N e^{\bar{\lambda} t}, \quad t \gg 1, \quad (\text{A.6c})$$

$$f_j(t) \approx \bar{f}_j = \frac{\xi_j}{\xi_N}, \quad j \in \{1, s, 2\}, t \gg 1. \quad (\text{A.6d})$$

We therefore find that the long time behaviour is characterised by a stationary DNA-distribution,  $\bar{s}(x)$ , of cells in the S phase, and constant cell fractions,  $\bar{f}_j$ . Following [54] we will denote this specific regime as the *steady phase solution* to highlight the fact that the fraction of cells in each phase of the cycle remains constant.

We now discuss how this *steady phase solution* of the model can be used to estimate the model parameters. Assume that the cells are left growing in an unperturbed environment for sufficiently long time so as to reach the regime of balanced exponential growth. Provided that we know the fraction of cells,  $\bar{f}_j$ , and proliferation rate,  $\lambda$ , of the population of cells (which can be approximated using the doubling time as given in the main

text,  $T_{doub} = \lambda^{-1} \ln 2$ ), we can uniquely identify the parameters  $k_1, k_2, \tau_s$  in the model. Let us substitute the solution (A.6) into (19) and, upon re-writing  $\bar{g}_i = \bar{f}_i \xi_N e^{\lambda t}$ , we obtain the following algebraic system:

$$\lambda \bar{f}_2 = -k_2 \bar{f}_2 + \bar{f}_1 k_1 e^{-\tau_s \lambda}, \quad (\text{A.7a})$$

$$\lambda \bar{f}_1 = 2k_2 \bar{f}_2 - k_1 \bar{f}_1, \quad (\text{A.7b})$$

$$\lambda = k_2 \bar{f}_2. \quad (\text{A.7c})$$

Looking at Eq. (A.7c) it is apparent that  $k_2$  is uniquely identified and it is positive. Substituting  $k_2$  into Eq. (A.7b), we obtain an equation for  $k_1$ :

$$k_1 = \frac{2 - \bar{f}_1}{\bar{f}_2} \lambda, \quad (\text{A.8})$$

where we note that  $k_1 > 0$  since, by definition,  $\bar{f}_1 \leq 1$ . Substituting now the forms of  $k_1$  and  $k_2$  into (A.7a) we obtain an expression for  $\tau_s$ :

$$\tau_s = \frac{1}{\lambda} \ln \left( \frac{2 - \bar{f}_1}{\bar{f}_2 + 1} \right) \quad (\text{A.9})$$

where the physical constraint  $\bar{f}_1 + \bar{f}_2 \leq 1$  guarantees that the argument of the logarithm is always positive. Hence, given the measurements of  $\bar{f}_1, \bar{f}_2, \tau_s$  we can uniquely identify the constant parameters appearing in the model (19) for unperturbed growth.

## Appendix B. Experimental data

Here we present the raw-data for the cell-cycle dynamics in RKO cancer cells as measured in the experiments from [5] discussed in §2 of the main text. Table B.6 summarises the value of the cell fractions in normoxia ( $E_0$ ), while Table B.7 and Table B.8 represent the time series data obtained when cells are exposed to constant ( $E_1$ ) and cyclic ( $E_2$ ) RH, respectively.

Table B.6: Raw experimental data for the stationary cell-cycle dynamics of RKO cancer cells when exposed to normoxia ( $E_0$ ) for sufficiently long time.

| $F_1$       | $F_s$       | $F_2$       |
|-------------|-------------|-------------|
| 0.291222313 | 0.557013946 | 0.151763741 |
| 0.297764120 | 0.533750132 | 0.168485748 |
| 0.276288660 | 0.602061856 | 0.121649485 |
| 0.391933816 | 0.475698035 | 0.132368149 |
| 0.270000000 | 0.540000000 | 0.190000000 |
| 0.353889645 | 0.498481834 | 0.147628521 |
| 0.296387429 | 0.515714126 | 0.187898445 |
| 0.289529012 | 0.571414043 | 0.139056945 |
| 0.291841004 | 0.560669456 | 0.147489540 |

Table B.7: Raw experimental data for the cell-cycle dynamics of RKO cancer cells when exposed to constant RH ( $E_1$ ).

| Time [hr] | $F_1$       | $F_s$       | $F_2$       |
|-----------|-------------|-------------|-------------|
| 4         | 0.491356731 | 0.297537978 | 0.207438449 |
| 4         | 0.497400877 | 0.36815819  | 0.131485068 |
| 8         | 0.513138837 | 0.334160977 | 0.147941237 |
| 8         | 0.563423645 | 0.325328407 | 0.104679803 |

Table B.8: Raw experimental data for the cell-cycle dynamics of RKO cancer cells when exposed to cyclic RH ( $E_2$ ).

| Time [hr] | $F_1$       | $F_s$       | $F_2$       |
|-----------|-------------|-------------|-------------|
| 4         | 0.330630725 | 0.546067179 | 0.120037907 |
| 4         | 0.266513378 | 0.632211538 | 0.098035117 |
| 4         | 0.297175141 | 0.534011299 | 0.164971751 |
| 8         | 0.220709571 | 0.731332508 | 0.044657591 |
| 8         | 0.191676374 | 0.754315366 | 0.050937202 |
| 20        | 0.159383033 | 0.702313625 | 0.133676093 |
| 20        | 0.093743547 | 0.777617179 | 0.125954987 |
| 20        | 0.144138373 | 0.711189408 | 0.140935298 |
| 20        | 0.100658314 | 0.726587386 | 0.168825653 |
| 20        | 0.184611856 | 0.635248251 | 0.171998624 |
| 20        | 0.145497026 | 0.713785047 | 0.13700085  |
| 28        | 0.060079664 | 0.712547024 | 0.21132994  |
| 28        | 0.044509303 | 0.647087438 | 0.29916089  |
| 28        | 0.03960499  | 0.676611227 | 0.280665281 |

## Appendix C. MCMC sampling algorithm and results

In order to estimate the posterior distribution (see Eq. (24a)), we use Markov Chain Monte Carlo (MCMC) methods, employing the freely available implementation in the python package PINTS [19]. As suggested by Johnstone et al. [38], prior to starting our MCMC routine, we compute a good initial guess by maximizing the likelihood function  $\mathcal{L}$  (see Eq. (23b)). Based on Eq. (24a) and the choice of a uniform prior, maximising the posterior is equivalent to maximising the likelihood function. Sampling random initial guesses from  $\pi_{pr}$ , we solve the optimisation problem for the log-likelihood using the CMA-ES algorithm [19, 34]. We then use the output of the optimization routine to initiate the MCMC simulation (we compute, in total, three chains). We sample from the posterior distribution by using HaarioBardenet MCMC, which is a Metropolis-Hastings algorithm with adaptive covariance, where the first 8000 iterations are performed without adaptation (as suggested by [38]). We compute up to 30000 iterations for each chain and discard the first 10000 as “warm-up”. As in [20], we assess the convergence of the MCMC chains by estimating  $\hat{R}$  [41, chapter 13], where we accept the sampled posterior if  $\hat{R} < 1.05$  (note that  $\hat{R} \rightarrow 1$  as the algorithm converges).

In Figs. C.23-C.26, we show the estimated posterior distributions for our class of models  $M$  (see Table 3). Summary statistics of the marginal posterior distributions are listed in Table C.9. For the point estimates of parameter values in §5, we use the mean of the marginal posterior distributions (see Table C.9).

Table C.9: Summary statistics for the marginal posterior distributions for the family of models  $M$  listed in Table 3 in §5.2. We here report the mean and standard deviation together with the quantiles ( $Q_i$ ) of the distribution. The last column shows the value of  $\hat{R}$  used to estimate the convergence of the MCMC algorithm, where convergence corresponds to  $\hat{R} < 1.05$ .

| Model |                 | mean  | std  | $Q_{25\%}$ | $Q_{50\%}$ | $Q_{75\%}$ | $\hat{R}$ |
|-------|-----------------|-------|------|------------|------------|------------|-----------|
| $M_0$ | $\alpha_1$      | 1.57  | 0.50 | 1.17       | 1.55       | 1.96       | 1.01      |
|       | $\mathcal{R}_+$ | 0.25  | 0.11 | 0.17       | 0.24       | 0.32       | 1.01      |
|       | $\mathcal{R}_-$ | 0.28  | 0.10 | 0.21       | 0.27       | 0.34       | 1.01      |
|       | $p$             | 0.83  | 0.11 | 0.77       | 0.85       | 0.91       | 1.02      |
|       | $q$             | 0.89  | 0.06 | 0.86       | 0.90       | 0.94       | 1.01      |
|       | $T$             | 16.89 | 1.16 | 16.21      | 17.02      | 17.74      | 1.01      |
|       | $K_2$           | 1.04  | 0.55 | 0.59       | 1.06       | 1.51       | 1.00      |
|       | $\sigma_1$      | 0.04  | 0.01 | 0.03       | 0.04       | 0.04       | 1.01      |
|       | $\sigma_2$      | 0.05  | 0.01 | 0.03       | 0.05       | 0.05       | 1.01      |
|       |                 |       |      |            |            |            |           |
| $M_1$ | $\alpha_1$      | 1.53  | 0.50 | 1.13       | 1.50       | 1.94       | 1.00      |
|       | $\mathcal{R}_+$ | 0.26  | 0.12 | 0.17       | 0.25       | 0.33       | 1.00      |
|       | $\mathcal{R}_-$ | 0.30  | 0.13 | 0.22       | 0.28       | 0.58       | 1.00      |
|       | $p$             | 0.83  | 0.11 | 0.76       | 0.85       | 0.91       | 1.00      |
|       | $q$             | 0.90  | 0.06 | 0.86       | 0.91       | 0.95       | 1.00      |
|       | $T$             | 16.93 | 1.18 | 16.21      | 17.05      | 17.78      | 1.00      |
|       | $\sigma_1$      | 0.04  | 0.01 | 0.03       | 0.04       | 0.04       | 1.00      |
|       | $\sigma_2$      | 0.05  | 0.01 | 0.04       | 0.05       | 0.05       | 1.00      |
|       |                 |       |      |            |            |            |           |
|       |                 |       |      |            |            |            |           |
| $M_2$ | $\mathcal{R}_+$ | 0.11  | 0.08 | 0.05       | 0.10       | 0.15       | 1.01      |
|       | $\mathcal{R}_-$ | 0.14  | 0.06 | 0.10       | 0.13       | 0.17       | 1.00      |
|       | $p$             | 0.71  | 0.16 | 0.59       | 0.71       | 0.83       | 1.01      |
|       | $T$             | 15.29 | 0.87 | 14.75      | 15.27      | 15.86      | 1.01      |
|       | $\sigma_1$      | 0.14  | 0.02 | 0.14       | 0.15       | 0.19       | 1.00      |
|       | $\sigma_2$      | 0.04  | 0.01 | 0.04       | 0.04       | 0.05       | 1.00      |
|       |                 |       |      |            |            |            |           |
|       |                 |       |      |            |            |            |           |
|       |                 |       |      |            |            |            |           |
|       |                 |       |      |            |            |            |           |
| $M_3$ | $\alpha_1$      | 1.63  | 0.53 | 1.22       | 1.66       | 2.08       | 1.01      |
|       | $\mathcal{R}_+$ | 0.21  | 0.19 | 0.08       | 0.16       | 0.27       | 1.04      |
|       | $\mathcal{R}_-$ | 0.53  | 0.31 | 0.32       | 0.45       | 0.66       | 1.03      |
|       | $q$             | 0.93  | 0.05 | 0.89       | 0.93       | 0.97       | 1.01      |
|       | $\sigma_1$      | 0.04  | 0.01 | 0.04       | 0.04       | 0.05       | 1.01      |
|       | $\sigma_2$      | 0.12  | 0.02 | 0.11       | 0.12       | 0.14       | 1.01      |
|       |                 |       |      |            |            |            |           |
|       |                 |       |      |            |            |            |           |
|       |                 |       |      |            |            |            |           |
|       |                 |       |      |            |            |            |           |

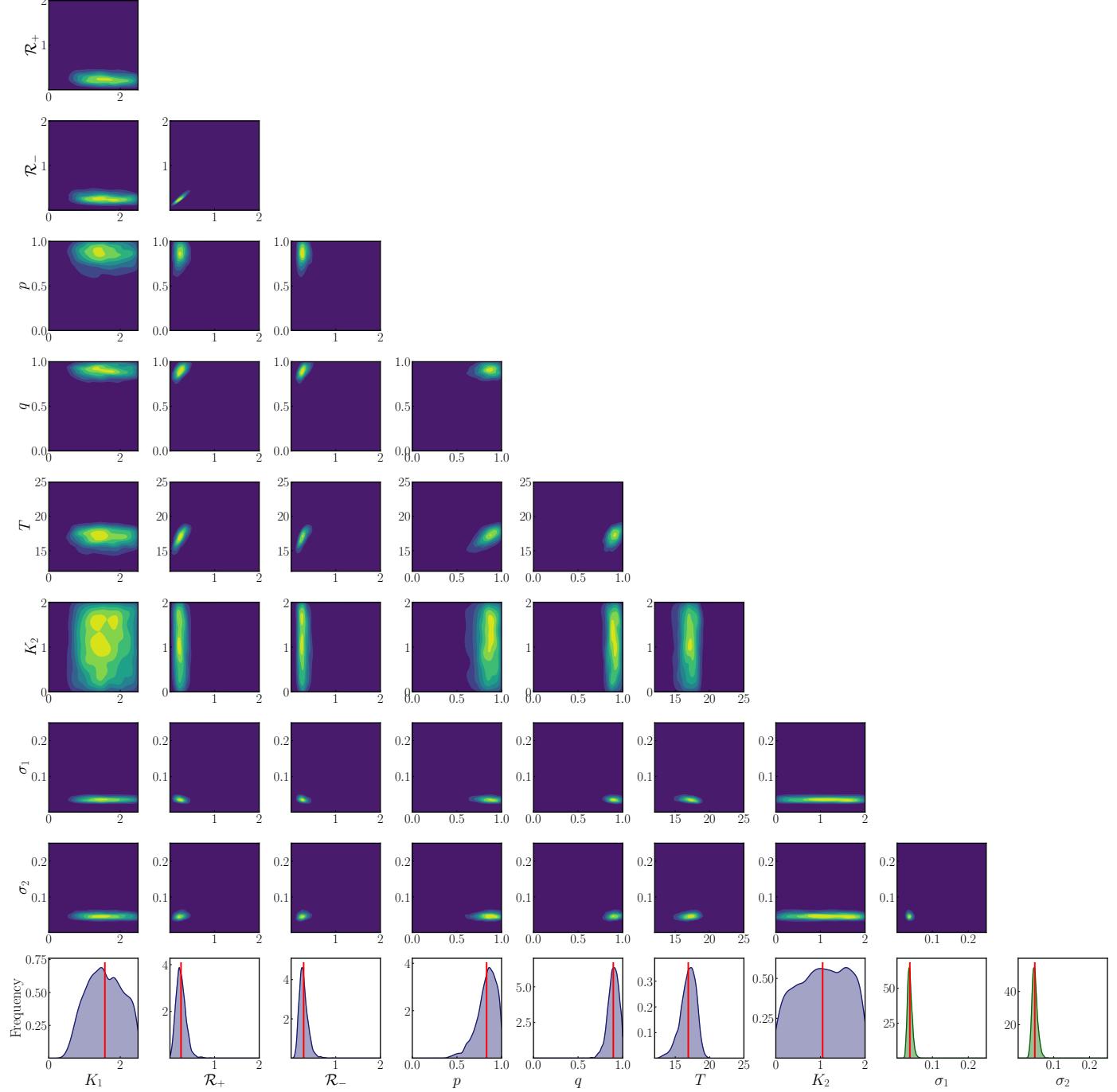


Figure C.23: Approximation of the joint (surface plot) and marginal (last row) posterior distributions for model  $\mathcal{M}_0$  from Table 3. The distributions were obtained by using the MCMC samples generated as discussed in the text. In the surface plot for joint distributions, yellow areas correspond to higher posterior probability in contrast to the blue areas which correspond to low probability. In the last row, the red vertical line indicates the mean of the marginal posterior distribution as reported in Table C.9, where additional summary statistics extrapolated from the marginal distribution are also given.

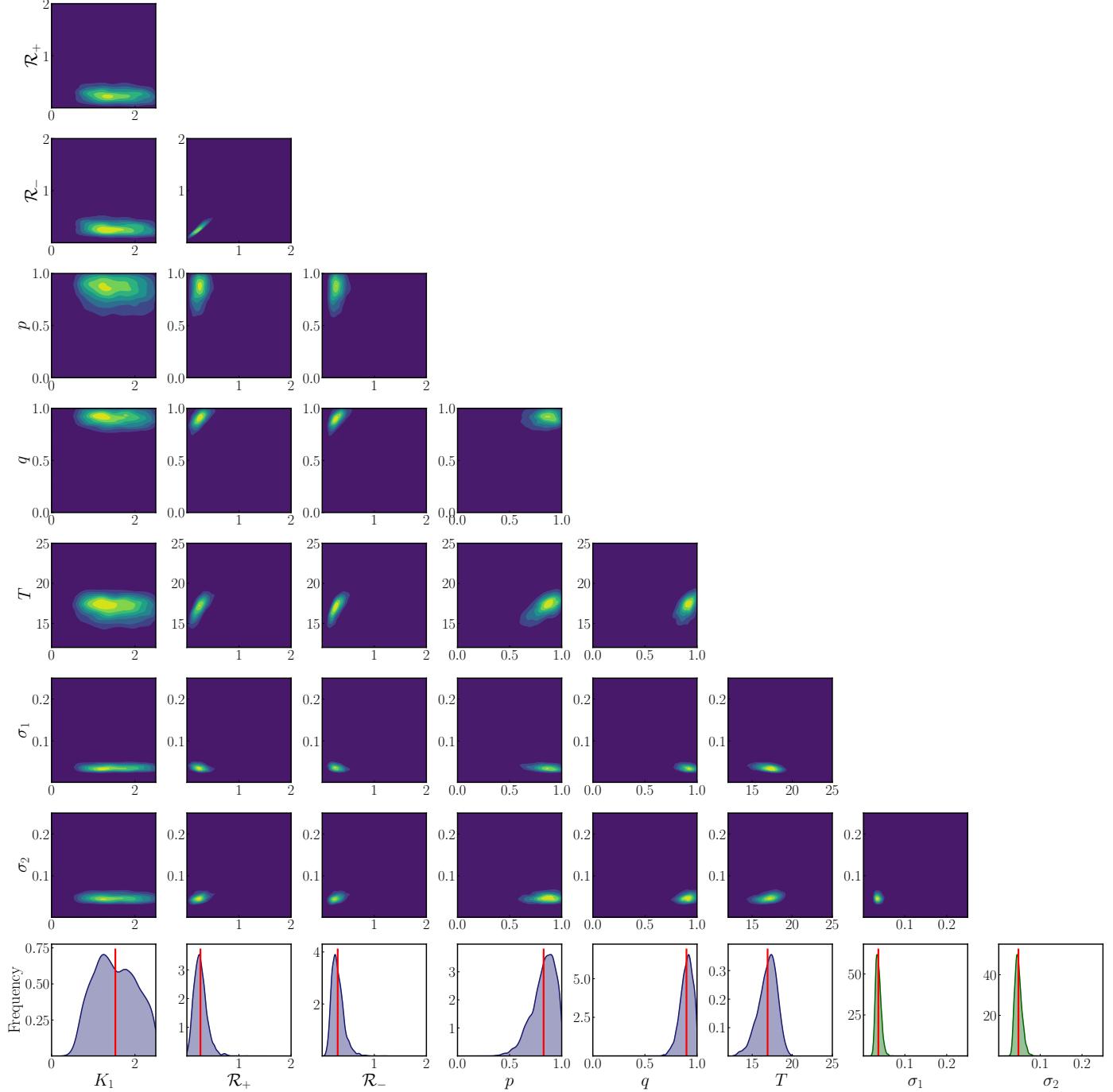


Figure C.24: Approximation of the joint (surface plot) and marginal (last row) posterior distributions for model  $\mathcal{M}_1$  from Table 3. The distributions were obtained by using the MCMC samples generated as discussed in the text. In the surface plot for joint distributions, yellow areas correspond to higher posterior probability in contrast to the blue areas which correspond to low probability. In the last row, the red vertical line indicates the mean of the marginal posterior distribution as reported in Table C.9, where additional summary statistics extrapolated from the marginal distribution are also given.

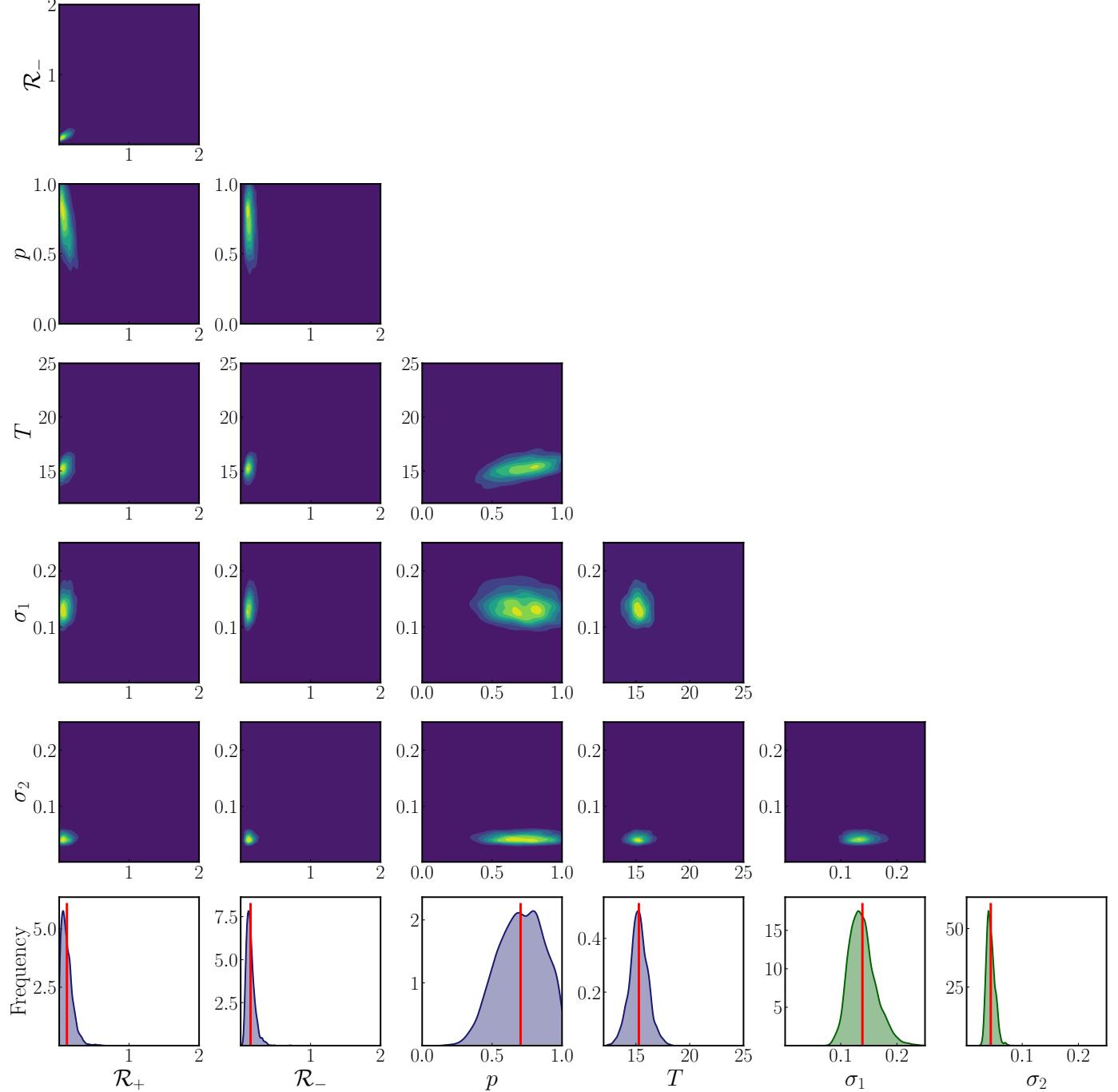


Figure C.25: Approximation of the joint (surface plot) and marginal (last row) posterior distributions for model  $\mathcal{M}_2$  from Table 3. The distributions were obtained by using the MCMC samples generated as discussed in the text. In the surface plot for joint distributions, yellow areas correspond to higher posterior probability in contrast to blue areas which correspond to low probability. In the last row, the red vertical line indicates the mean of the marginal posterior distribution as reported in Table C.9, where additional summary statistics extrapolated from the marginal distribution are also given.

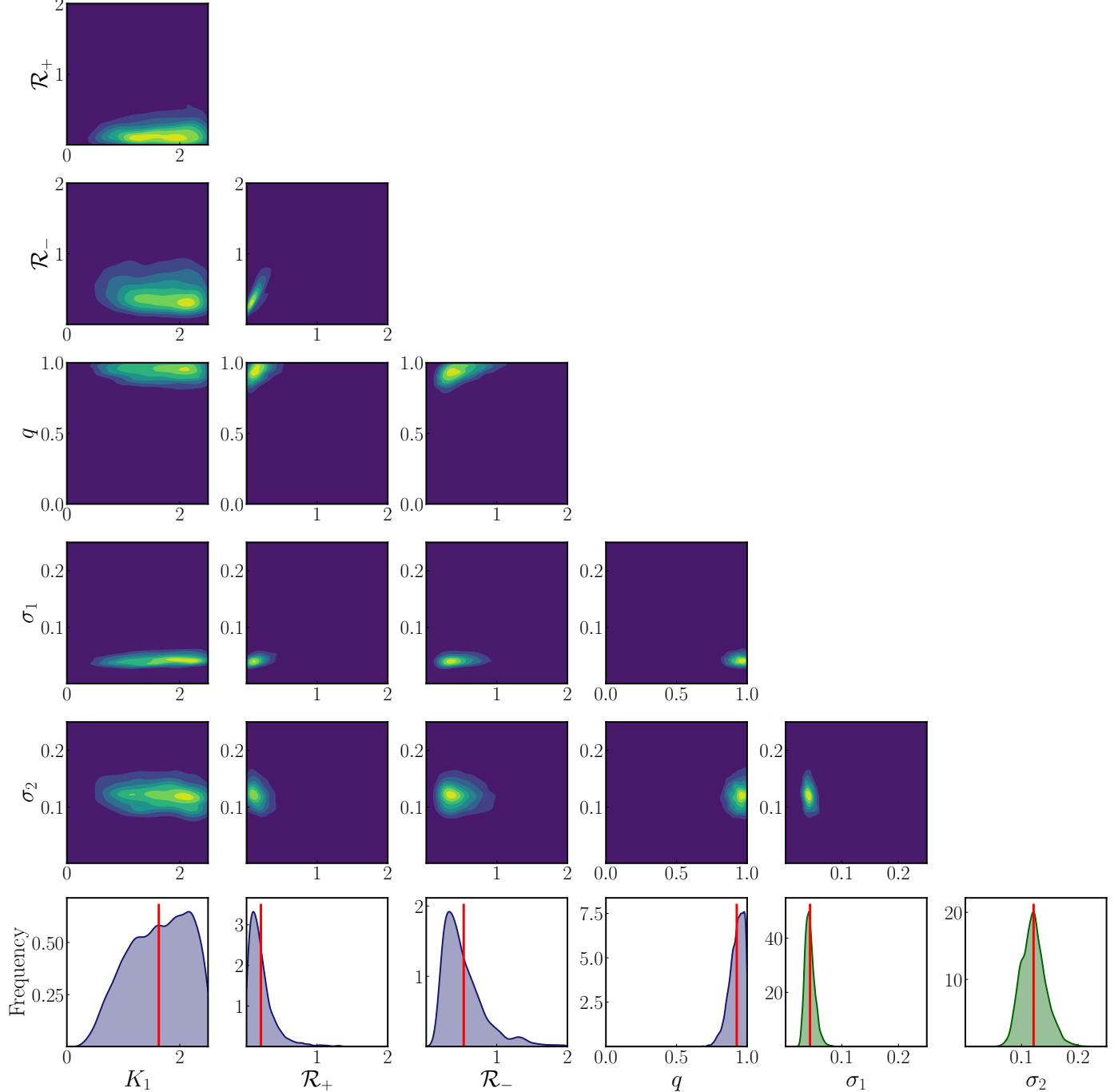


Figure C.26: Approximation of the joint (surface plot) and marginal (last row) posterior distributions for model  $\mathcal{M}_3$  from Table 3. The distributions were obtained by using the MCMC samples generated as discussed in the text. In the surface plot for joint distributions, yellow areas correspond to higher posterior probability in contrast to the blue areas which correspond to low probability. In the last row, the red vertical line indicates the mean of the marginal posterior distribution as reported in Table C.9, where additional summary statistics extrapolated from the marginal distribution are also given.

## References

[1] Alarcón, T., Byrne, H.M., Maini, P.K., 2004. A mathematical model of the effects of hypoxia on the cell-cycle of normal and cancer cells. *Journal of Theoretical Biology* 229, 395–411. doi:10.1016/j.jtbi.2004.04.016.

[2] Alfieri, R., Bartocci, E., Merelli, E., Milanesi, L., 2011. Modeling the cell cycle: From deterministic models to hybrid systems. *BioSystems* 105, 34–40. doi:10.1016/j.biosystems.2011.03.002.

[3] Altinok, A., Gonze, D., Lévi, F., Goldbeter, A., 2011. An automaton model for the cell cycle. *Interface Focus* 1, 36–47. doi:10.1098/rsfs.2010.0009.

[4] Bader, S.B., Dewhirst, M.W., Hammond, E.M., 2021a. Review cyclic hypoxia: An update on its characteristics, methods to measure it and biological implications in cancer. *Cancers* 13, 1–20. doi:10.3390/cancers13010023.

[5] Bader, S.B., Ma, T.S., Simpson, C.J., Liang, J., Maezono, S.E.B., Olcina, M.M., Buffa, F.M., Hammond, E.M., 2021b. Replication catastrophe induced by cyclic hypoxia leads to increased APOBEC3B activity. *Nucleic Acids Research* 49, 7492–7506. doi:10.1093/nar/gkab551.

[6] Barker, H., Paget, J., Khan, A., Harrington, K., 2015. The tumour microenvironment after radiotherapy: Mechanisms of resistance and recurrence. *Nature reviews. Cancer* 15, 409–25. doi:10.1038/nrc3958.

[7] Bartkova, J., Hořejší, Z., Koed, K., Krämer, A., Tort, F., Zleger, K., Guldberg, P., Sehested, M., Nesland, J.M., Lukas, C., Orntoft, T., Lukas, J., Bartek, J., 2005. DNA damage response as a candidate anti-cancer barrier in early human tumorigenesis. *Nature* 434, 864–870. doi:10.1038/nature03482.

[8] Basse, B., Baguley, B.C., Marshall, E.S., Joseph, W.R., Van Brunt, B., Wake, G., Wall, D.J., 2003. A mathematical model for analysis of the cell cycle in cell lines derived from human tumors. *Journal of Mathematical Biology* 47, 295–312. doi:10.1007/s00285-003-0203-0.

[9] Basse, B., Baguley, B.C., Marshall, E.S., Joseph, W.R., Van Brunt, B., Wake, G., Wall, D.J., 2004. Modelling cell death in human tumour cell lines exposed to the anticancer drug paclitaxel. *Journal of Mathematical Biology* 49, 329–357. doi:10.1007/s00285-003-0254-2.

[10] Basse, B., Baguley, B.C., Marshall, E.S., Wake, G.C., Wall, D.J., 2005. Modelling the flow of cytometric data obtained from unperturbed human tumour cell lines: Parameter fitting and comparison. *Bulletin of Mathematical Biology* 67, 815–830. doi:10.1016/j.bulm.2004.10.003.

[11] Basse, B., Ubezio, P., 2007. A generalised age- and phase-structured model of human tumour cell populations both unperturbed and exposed to a range of cancer therapies. *Bulletin of Mathematical Biology* 69, 1673–1690. doi:10.1007/s11538-006-9185-6.

[12] Begg, K., Tavassoli, M., 2020. Inside the hypoxic tumour: reprogramming of the DDR and radioresistance. *Cell Death Discovery* 6, 1234567890. doi:10.1038/s41420-020-00311-0.

[13] Begg, R.E., Wall, D.J., Wake, G.C., 2008. The steady-states of a multi-compartment, age-size distribution model of cell-growth. *European Journal of Applied Mathematics* 19, 435–458. doi:10.1017/S0956792508007535.

[14] Bell, G.I., 1968. Cell growth and division: III. conditions for balanced exponential growth in a mathematical model. *Biophysical Journal* 8, 431–444. doi:10.1016/S0006-3495(68)86498-7.

[15] Bristow, R.G., Hill, R.P., 2008. Hypoxia and metabolism: Hypoxia, DNA repair and genetic instability. *Nature Reviews Cancer* 8, 180–192. doi:10.1038/nrc2344.

[16] Chao, H.X., Poovey, C.E., Privette, A.A., Grant, G.D., Chao, H.Y., Cook, J.G., Purvis, J.E., 2017. Orchestration of DNA Damage Checkpoint Dynamics across the Human Cell Cycle. *Cell Systems* 5, 445–459.e5. doi:10.1016/j.cels.2017.09.015.

[17] Chapman, S.J., Plank, M.J., James, A., Basse, B., 2008. A nonlinear model of age and size-structured populations with applications to cell cycles. *ANZIAM Journal* 49, 151–169. doi:10.1017/S144618110001275X.

[18] Chiorino, G., Lupi, M., 2002. Variability in the timing of G1/S transition. *Mathematical Biosciences* 177–178, 85–101. doi:10.1016/S0025-5564(02)00085-8.

[19] Clerx, M., Robinson, M., Lambert, B., Lei, C.L., Ghosh, S., Mirams, G.R., Gavaghan, D.J., 2019. Probabilistic inference on noisy time series (PINTS). *Journal of Open Research Software* 7, 23. doi:10.5334/jors.252.

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

[21] Crivelli, J.J., Földes, J., Kim, P.S., Wares, J.R., 2012. A mathematical model for cell cycle-specific cancer virotherapy. *Journal of Biological Dynamics* 6, 104–120. doi:10.1080/17513758.2011.613486.

[22] Csikász-Nagy, A., 2009. Computational systems biology of the cell cycle. *Briefings in Bioinformatics* 10, 424–434. doi:10.1093/bib/bbp005.

[23] Daly, A.C., Gavaghan, D., Cooper, J., Tavener, S., 2018. Inference-based assessment of parameter identifiability in nonlinear biological models. *Journal of The Royal Society Interface* 15, 20180318. doi:10.1098/rsif.2018.0318.

[24] Ducrot, A., Le Foll, F., Magal, P., Murakawa, H., Pasquier, J., Webb, G.F., 2011. An in vitro cell population dynamics model incorporating cell size, quiescence, and contact inhibition. *Mathematical Models and Methods in Applied Sciences* 21, 871–892. doi:10.1142/S0218202511005404.

[25] Fadda, S., Cincotti, A., Cao, G., 2012. A novel population balance model to investigate the kinetics of in vitro cell proliferation: Part I. model development. *Biotechnology and Bioengineering* 109, 772–781. doi:10.1002/bit.24351.

[26] Fauré, A., Naldi, A., Chaouiya, C., Thieffry, A., 2006. Dynamical analysis of a generic Boolean model for the control of the mammalian cell cycle. *Bioinformatics* 22, 124–131. doi:10.1093/bioinformatics/btl210.

[27] Foskolou, I.P., Jorgensen, C., Leszczynska, K.B., Olcina, M.M., Tarhonskaya, H., Haisma, B., D'Angiolella, V., Myers, W.K., Domene, C., Flashman, E., Hammond, E.M., 2017. Ribonucleotide Reductase Requires Subunit Switching in Hypoxia to Maintain DNA Replication. *Molecular Cell* 66, 206–220. doi:10.1016/j.molcel.2017.03.005.

[28] Gelman, A., Carlin, J., Stern, H., Dunson, D., Vehtari, A., Rubin, D., 2013. *Bayesian Data Analysis*. Chapman & Hall/CRC Texts in Statistical Science, CRC Press.

[29] Good, J., Harrington, K., 2013. The hallmarks of cancer and the radiation oncologist: Updating the 5rs of radiobiology. *Clinical Oncology* 25, 569 – 577. doi:10.1016/j.clon.2013.06.009.

[30] Gorgoulis, V.G., Vassiliou, L.V.F., Karakaidos, P., Zacharatos, P., Kotsinas, A., Liloglou, T., Venere, M., DiTullio, R.A., Kastrinakis, N.G., Levy, B., Kletsas, D., Yoneta, A., Herlyn, M., Kittas, C., Halazonetis, T.D., 2005. Activation of the DNA damage checkpoint and genomic instability in human precancerous lesions. *Nature* 434, 907–913. doi:10.1038/nature03485.

[31] Goto, T., Kaida, A., Miura, M., 2015. Visualizing cell-cycle kinetics after hypoxia/reoxygenation in HeLa cells expressing fluorescent ubiquitination-based cell cycle indicator (Fucci). *Experimental Cell Research* 339, 389–396. doi:10.1016/j.yexcr.2015.10.019.

[32] Gyllenberg, M., Webb, G.F., 1990. A nonlinear structured population model of tumor growth with quiescence. *Journal of Mathematical Biology* 28, 671–694. doi:10.1007/BF00160231.

[33] Hanahan, D., Weinberg, R.A., 2000. The hallmarks of cancer. *Cell* 100, 57 – 70. doi:10.1016/S0092-8674(00)81683-9.

[34] Hansen, N., Müller, S.D., Koumoutsakos, P., 2003. Reducing the time complexity of the derandomized evolution strategy with covariance matrix adaptation (cma-es). *Evolutionary Computation* 11, 1–18. doi:10.1162/106365603321828970.

[35] Hines, K.E., Middendorf, T.R., Aldrich, R.W., 2014. Determination of parameter identifiability in nonlinear biophysical models: A bayesian approach. *Journal of General Physiology* 143, 401–416. doi:10.1085/jgp.20131116.

[36] Hsieh, C.H., Lee, C.H., Liang, J.A., Yu, C.Y., Shyu, W.C., 2010. Cycling hypoxia increases u87 glioma cell radioresistance via ros induced higher and long-term hif-1 signal transduction activity. *Oncology Reports* 24, 1629–1636. doi:10.3892/or.00001027.

[37] Hsieh, C.H., Wu, C.P., Lee, H.T., Liang, J.A., Yu, C.Y., Lin, Y.J., 2012. NADPH oxidase subunit 4 mediates cycling hypoxia-promoted radiation resistance in glioblastoma multiforme. *Free Radical Biology and Medicine* 53, 649–658. doi:10.1016/j.freeradbiomed.2012.06.009.

[38] Johnstone, R.H., Chang, E.T., Bardenet, R., de Boer, T.P., Gavaghan, D.J., Pathmanathan, P., Clayton, R.H., Mirams, G.R., 2016. Uncertainty and variability in models of the cardiac action potential: Can we build trustworthy models? *Journal of Molecular and Cellular Cardiology* 96, 49–62. doi:10.1016/j.jmcc.2015.11.018.

[39] Kirkpatrick, J., Cárdenas-Navia, L., Dewhirst, M., 2004. Predicting the effect of temporal variations in po2 on tumor radiosensitivity. *International Journal of Radiation Oncology, Biology, and Physics* 61, 133–140. doi:10.1016/j.ijrobp.2003.11.030.

tional Journal of Radiation Oncology-Biology-Physics 59, 822 – 833. doi:10.1016/j.ijrobp.2004.02.015.

[40] Koumenis, C. and Wouters, B.G., 2006. “Translating” tumor hypoxia: Unfolded protein response (UPR)-dependent and UPR-independent pathways. doi:10.1158/1541-7786.MCR-06-0150.

[41] Lambert, B., 2018. A Student’s Guide to Bayesian Statistics. SAGE Publications.

[42] Leszczynska, K.B., Dobrynin, G., Leslie, R.E., Ient, J., Boumelha, A.J., Senra, J.M., Hawkins, M.A., Maughan, T., Mukherjee, S., Hammond, E.M., 2016. Preclinical testing of an Atr inhibitor demonstrates improved response to standard therapies for esophageal cancer. Radiotherapy and Oncology 121, 232–238. doi:10.1016/j.radonc.2016.10.023.

[43] Maad Sasane, S., 2016. An age structured cell cycle model with crowding. Journal of Mathematical Analysis and Applications 444, 768–803. doi:10.1016/j.jmaa.2016.06.065.

[44] Michiels, C., Tellier, C., Feron, O., 2016. Cycling hypoxia: A key feature of the tumor microenvironment. Biochimica et Biophysica Acta (BBA) - Reviews on Cancer 1866, 76 – 86. doi:10.1016/j.bbcan.2016.06.004.

[45] Negrini, S., Gorgoulis, V.G., Halazonetis, T.D., 2010. Genomic instability an evolving hallmark of cancer. Nature Reviews Molecular Cell Biology 11, 220–228. doi:10.1038/nrm2858.

[46] Olcina, M., Lecane, P.S., Hammond, E.M., 2010. Targeting hypoxic cells through the DNA damage response. Clinical Cancer Research 16, 5624–5629. doi:10.1158/1078-0432.CCR-10-0286.

[47] Olcina, M.M., Foskolou, I.P., Anbalagan, S., Senra, J.M., Pires, I.M., Jiang, Y., Ryan, A.J., Hammond, E.M., 2013. Replication stress and chromatin context link ATM activation to a role in DNA replication. Molecular Cell 52, 758–766. doi:10.1016/j.molcel.2013.10.019.

[48] Pawlik, T.M., Keyomarsi, K., 2004. Role of cell cycle in mediating sensitivity to radiotherapy. International Journal of Radiation Oncology\*Biology\*Physics 59, 928–942. doi:<https://doi.org/10.1016/j.ijrobp.2004.03.005>.

[49] Pires, I.M., Bencokova, Z., Milani, M., Folkes, L.K., Li, J.A., Stratford, M.R., Harris, A.L., Hammond, E.M., 2010. Effects of acute versus chronic hypoxia on DNA damage responses and genomic instability. Cancer Research 70, 925–935. doi:10.1158/0008-5472.CAN-09-2715.

[50] Place, T.L., Domann, F.E., Case, A.J., 2017. Limitations of oxygen delivery to cells in culture: An underappreciated problem in basic and translational research. Free Radical Biology and Medicine 113, 311–322. doi:10.1016/J.FREERADBIOMED.2017.10.003.

[51] Ron, A., Deán-Ben, X.L., Gottschalk, S., Razansky, D., 2019. Volumetric optoacoustic imaging unveils high-resolution patterns of acute and cyclic hypoxia in a murine model of breast cancer. Cancer Research 79, 4767–4775. doi:10.1158/0008-5472.CAN-18-3769.

[52] Saxena, K., Jolly, M.K., 2019. Acute vs. Chronic vs. cyclic hypoxia: Their differential dynamics, molecular mechanisms, and effects on tumor progression. Biomolecules 9. doi:10.3390/biom9080339.

[53] Sherer, E., Tocce, E., Hannemann, R., Rundell, A., Ramkrishna, D., 2008. Identification of age-structured models: Cell cycle phase transitions. Biotechnology and Bioengineering 99, 960–974. doi:10.1002/bit.21633.

[54] Simms, K., Bean, N., Koerber, A., 2012. A Mathematical Model of Cell Cycle Progression Applied to the MCF-7 Breast Cancer Cell Line. Bulletin of Mathematical Biology 74, 736–767. doi:10.1007/s11538-011-9700-2.

[55] Singhania, R., Sramkoski, R.M., Jacobberger, J.W., Tyson, J.J., 2011. A Hybrid Model of Mammalian Cell Cycle Regulation. PLoS Computational Biology 7. doi:10.1371/journal.pcbi.1001077.

[56] Spinelli, L., Torricelli, A., Ubezio, P., Basse, B., 2006. Modelling the balance between quiescence and cell death in normal and tumour cell populations. Mathematical Biosciences 202, 349–370. doi:10.1016/j.mbs.2006.03.016.

[57] Taylor, W.R., Stark, G.R., 2001. Regulation of the G2/M transition by p53. Oncogene 2001 20:15 20, 1803–1815. doi:10.1038/sj.onc.1204252.

[58] Ubezio, P., 2004. Unraveling the complexity of cell cycle effects of anticancer drugs in cell populations. Discrete and Continuous Dynamical Systems - Series B 4, 323–335. doi:10.3934/dcdsb.2004.4.323.

[59] Ubezio, P., Lupi, M., Branduardi, D., Cappella, P., Cavallini, E., Colombo, V., Matera, G., Natoli, C., Tomasoni, D., D’Incalci, M., 2009. Quantitative assessment of the complex dynamics of G1, S, and G2-M checkpoint activities. Cancer Research 69, 5234–5240. doi:10.1158/0008-5472.CAN-08-3911.

[60] Vittadello, S.T., McCue, S.W., Gunasingh, G., Haass, N.K., Simpson, M.J., 2018. Mathematical Models for Cell Migration with Real-Time Cell Cycle Dynamics. Biophysical Journal 114, 1241–1253. doi:10.1016/j.bpj.2017.12.041.

[61] Vittadello, S.T., McCue, S.W., Gunasingh, G., Haass, N.K., Simpson, M.J., 2019. Mathematical models incorporating a multi-stage cell cycle replicate normally-hidden inherent synchronization in cell proliferation. Journal of The Royal Society Interface 16, 20190382. doi:10.1098/rsif.2019.0382.

[62] Vittadello, S.T., McCue, S.W., Gunasingh, G., Haass, N.K., Simpson, M.J., 2021. A novel mathematical model of heterogeneous cell proliferation. Journal of Mathematical Biology 82, 34. doi:10.1007/s00285-021-01580-8, [arXiv:2003.03024](https://arxiv.org/abs/2003.03024).

[63] Weber, T.S., Jaehnert, I., Schichor, C., Or-Guil, M., Carneiro, J., 2014. Quantifying the Length and Variance of the Eukaryotic Cell Cycle Phases by a Stochastic Model and Dual Nucleoside Pulse Labelling. PLoS Computational Biology 10. doi:10.1371/journal.pcbi.1003616.

[64] West, C.M., Slevin, F., 2019. Tumour Hypoxia. Clinical Oncology 31, 595–599. URL: <https://doi.org/10.1016/j.clon.2019.06.008>, doi:10.1016/j.clon.2019.06.008.

[65] Witzel, F., Fritzsche-Guenther, R., Lehmann, N., Sieber, A., Blüthgen, N., 2015. Analysis of impedance-based cellular growth assays. Bioinformatics 31, 2705–2712. doi:10.1093/bioinformatics/btv216.