A hierarchical Bayesian framework for understanding the spatiotemporal dynamics of the intestinal epithelium

O.J. Maclaren[1]*, A. Parker[2], C. Pin[2], S.R. Carding[2,3], A.J.M. Watson[2,3], A.G. Fletcher[4], H.M. Byrne[6] and P.K. Maini[6]

1 Department of Engineering Science, University of Auckland, Auckland, New Zealand

2 Gut Health and Food Safety Research Programme, Institute of Food Research, Norwich, United Kingdom

3 Norwich Medical School, University of East Anglia, Norwich, United Kingdom

4 School of Mathematics and Statistics, University of Sheffield, Sheffield, United Kingdom

5 Bateson Centre, University of Sheffield, Sheffield, United Kingdom

6 Wolfson Centre for Mathematical Biology, Mathematical Institute, University of Oxford, Oxford, United Kingdom

***oliver.maclaren@auckland.ac.nz**

Abstract

Our work addresses two key challenges, one biological and one methodological. First, we aim to understand how proliferation and cellular migration rates in the intestinal epithelium are related under healthy, damaged (Ara-C treated) and recovering conditions, and how these relations can be used to identify mechanisms of repair and regeneration. We analyse new data, presented in more detail in a companion paper, in which BrdU/IdU cell-labelling experiments were performed under these respective conditions. Second, in considering how to more rigorously process these data and

interpret them using mathematical models, we develop a probabilistic, hierarchical framework. This framework provides a best-practice approach for systematically modelling and understanding the uncertainties that can otherwise undermine drawing reliable conclusions - uncertainties in experimental measurement and treatment, difficult-to-compare mathematical models of underlying mechanisms, and unknown or unobserved parameters. Both discrete and continuous mechanistic models are considered and related via hierarchical conditional probability assumptions. This allows the incorporation of features of both continuum tissue models and discrete cellular models. We perform model checks on both in-sample and out-of-sample datasets and use these checks to illustrate how to test possible model improvements and assess the robustness of our conclusions. This allows us to consider - and ultimately decide against - the need to retain finite-cell-size effects to explain a small misfit appearing in one set of long-time, out-of-sample predictions. Our approach leads us to conclude, for the present set of experiments, that a primarily proliferation-driven model is adequate for predictions over most time-scales. We describe each stage of our framework in detail, and hope that the present work may also serve as a guide for other applications of the hierarchical approach to problems in computational and systems biology more generally.

Author Summary

The intestinal epithelium serves as an important model system for studying the dynamics and regulation of multicellular populations. It is characterised by rapid rates of self-renewal and repair; failure of the regulation of these processes is thought to explain, in part, why many tumours occur in the intestinal and similar epithelial tissues. These features have led to a large amount of work on estimating rate parameters in the intestine. There still remain, however, large gaps between the raw data collected, the experimental interpretation of these data and speculative mechanistic models for underlying processes. In our view hierarchical statistical modelling provides an ideal - but currently underutilised - method to begin to bridge these gaps. This approach makes essential use of the distinction between

'measurement', 'process' and 'parameter' models, giving an explicit framework for combining experimental data and mechanistic modelling in the presence of multiple sources of uncertainty. As we illustrate, the hierarchical approach also provides a suitable framework for addressing other methodological questions of broader interest in systems biology: how to systematically relate discrete and continuous mechanistic models; how to formally interpret and visualise statistical evidence; and how to represent the notion of causal mechanism as invariance under intervention.

Introduction

Motivation

The intestinal epithelium provides crucial barrier, transport and homeostatic functions. These requirements lead it to undergo constant repair and regeneration, and dysfunctions can result in pathologies such as tumorigenesis [1–7]. While much work has been carried out on estimating the rate parameters in the intestine and other epithelia [1, 8–10], attempts to interpret these experimental data using mechanistic modelling remain inconclusive (see e.g. [11–14]). A key issue in drawing reliable conclusions is the lack of consistent and reproducible frameworks for comparing models representing conjectured biological mechanisms, both to each other and to experimental data.

This challenge goes in both directions: using experimental data (taken to be 'true') to parameterise and test mathematical or computational formalisations of mechanistic theories, and using these models (taken to be 'true') to predict, interpret and question experimental results. Both experimental measurements and mathematical models are subject to uncertainty, and we hence need systematic ways of quantifying these uncertainties and attributing them to the appropriate sources. Furthermore, establishing a common framework for analysing experimental results, formulating mechanistic models and generating new predictions has many potential advantages for enabling interdisciplinary teams to communicate in a common language and efficiently discover and follow promising directions as and when they arise.

Approach

We address the above issues by developing a best-practice hierarchical Bayesian framework for combining measurements, models and inference procedures, and applying it to a tractable set of experiments targeting mechanisms of repair and regeneration in the intestinal epithelium. These experiments were performed ourselves and are presented in more detail in [15]. The aim of these experiments was to identify how proliferation rates, tissue growth and cellular migration rates are related under healthy, damaged (Ara-C treated) and recovering conditions, and how these relations can be used to identify mechanisms of repair and regeneration.

A notable feature of the Bayesian approach to probabilistic modelling is that all sources of uncertainty are represented via probability distributions, regardless of the source of uncertainty (e.g. physical or epistemic) [16–18]. We will adopt this perspective here, and thus we consider both observations and parameters to be random variables. Within a modelling or theoretical context, uncertainty may be associated with (at least): parameters within a mechanistic model of a biological or physical process, the mechanistic model of the process itself and the measurements of the underlying process. This leads to the existence (at least in principle) of a full joint probability distribution for observable, unobservable/unobserved variables, parameters and data.

Another key feature of the Bayesian perspective, of particular interest here, is that it provides a natural way of decomposing such full joint models in a *hierarchical* manner, e.g. by separating out processes occuring on different scales and at different analysis stages. A given set of hierarchical assumptions corresponds to assuming a factorisation of the full joint distribution mentioned above, and gives a more interpretable and tractable starting point.

Our overall factorisation follows that described in [18–21]. This comprises: a 'measurement model', which defines the observable (sample) features to be considered reproducible and to what precision they are reproducible (the measurement scale); an underlying 'process' model, which captures the key mechanistic hypotheses of spatiotemporal evolution, and a prior parameter model which defines a broad class of

a priori acceptable possible parameter values.

This hierarchical approach is being increasingly adopted - especially in areas such as environmental and geophysical science [22, 23], ecological modelling [24, 25], as well as in Bayesian statistical modelling and inverse problems more generally [17–21, 26]. In our view, however, many of the advantages of hierarchical Bayesian modelling remain under-appreciated and it offers many opportunities for formulating more unified frameworks for model-data and model-model comparison. Furthermore, we note that a similar hierarchical approach has recently received significant development in the context of the non-Bayesian 'extended-likelihood' statistical modelling framework [27–29]. Thus, in our view, many of the benefits of the present approach can be attributed to its hierarchical aspect in particular ([20] also emphasise this point).

As illustration of some of the modelling benefits of the hierarchical approach, we show how both discrete and continuous process models can be derived and related using the hierarchical perspective. This thus provides a clear, unified approach for understanding different - yet related - models and derivations, and for embedding models within a data-driven parameter estimation framework. The hierarchical perspective also provides a natural guide to exploring robustness to model closure assumptions, investigating model misfit and checking for invariance of proposed mechanisms under in/out-of-sample and treated/untreated conditions. These considerations are necessary since, despite the utility of statistical methods for obtaining parameter estimates and model predictions based on an *assumed-to-be-true* model and *assumed-to-be-true* datasets, it is also well-known that 'all models are wrong' [30] and that purely statistical modelling is generally inadequate for drawing causal conclusions without additional assumptions [31, 32]. We adopt a straightforward conditional/hierarchical interpretation of causal modelling (see [33, 34] for clear overviews) and emphasise the distinct roles of (Bayesian) predictive distributions vs. parameter distributions for model checking and the assessment of evidence, respectively (see [17, 30, 35–37] for clear discussion of these distinctions).

Conclusions

For predictions over moderate time-scales a primarily proliferation-driven model appears adequate as a sound starting point. Our framework offers a systematic, hierarchical modelling perspective for exploring and testing possible improvements, and using Bayesian inference for estimating parameters and generating model predictions. Our interdisciplinary approach, mixing new experimental work with mathematical and computational modelling, enables us to more reliably consider and test various hypotheses than would likely be obtained using only one or the other approach.

Structure

Our article is structured as follows. We first outline the experimental procedures (which are also described in more detail in [15]). We then show in detail how we build up our mathematical model, beginning from the data available, the resolution of the measurement process and a hypothesised underlying 'true' or ideal population from which samples are assumed to be drawn. We then consider both discrete and continuous representations of the spatiotemporal evolution of the underlying population, their relation to each other and some of the strengths and weaknesses of different representations. We then present the methods and results for our baseline and Ara-C-treated experiments, using our modelling and exploratory data analysis to interpret them. Each stage is described in quite some detail in order to demonstrate how the present framework may be adapted and used for other problems of computational biology more generally.

Materials and methods I: Experimental treatments

Homeostasis mouse model

To obtain estimates of intestinal epithelial proliferation and migration rates under normal, homeostatic conditions in healthy mice, we used standard methods of

proliferative cell labelling and tracing [1, 8–10] (see also [15] for full details). Actively proliferating cells in the intestinal crypts were labelled by single injection of the thymine analogue 5-bromo-2-deoxyuridine (BrdU) and labelled cells detected by immunostaining of intestinal sections collected from different individuals over time. Migration of labelled cells traced from the base of crypts to villus tips was monitored over the course of 96 hours (5760 min).

Blocked-proliferation mouse model

To assess the effects of proliferation inhibition on crypt/villus migration, migrating and proliferating epithelial cells were monitored by double labelling with two thymine analogues (BrdU and IdU), administered sequentially a number of hours apart and subsequently distinguished by specific immunostaining in longitudinal sections of small intestine. Following initial IdU labelling of proliferating cells at t= -17h (-1020 min, relative to Ara-C treatment), mice were then treated with cytosine arabinoside (Ara-C) at 250 mg/kg body weight, a dose reported to inhibit proliferation without causing major crypt cell atrophy (see [15] and references therein for full details). Tissues were collected over 24 hours, with BrdU administered one hour prior to collection to check for residual proliferation. Successful inhibition of proliferation following treatment with Ara-C was confirmed by an absence of BrdU (S-phase) and phospho-Histone H3 (pH3) staining (M-phase) in longitudinal sections of small intestine (again, see [15] for full details).

Recovering-proliferation mouse model

The above Ara-C treatment effect was observed to last for at least 10h (600 min). Cell proliferation resumed to near normal levels in samples obtained 18h (1080 min) post-Ara-C treatment. We hence considered samples collected at least 10h post-Ara-C treatment as corresponding to 'recovering-proliferation' conditions.

Materials and methods II: Mathematical framework

Overview

Here we build up a hierarchical probability model on the basis of conditional probability assumptions which are used to factor out a measurement model, a mechanistic model and a parameter model. We will interpret and visualise this hierarchical model in an operational manner i.e. as defining various ensembles (e.g. prior and/or posterior parameter and/or predictive ensembles) consisting of sequences of draws from the component distributions of the full product (see e.g. [38–40] for discussion of simulation-based interpretations of Bayesian models). We first discuss the model components, after which we summarise the overall model structure and computational implementation of the model in the 'Computational methods' section that follows.

Measurement model

Observable features and underlying population

Our aim is to connect experimental measurements to mechanistic models. Rather than begin from mechanistic hypotheses before considering the connection to available data, here we explicitly begin by thinking about the available data, how it is obtained and what the key features of interest are. These key features will then be considered to define an ideal 'underlying population' from which samples are considered to be drawn or to which they are to be compared. This means that we begin the construction of our hierarchical model in a 'top-down' (data-to-parameter) manner.

With reference to Fig 1, we consider a collection of one-dimensional 'strips' of cells. These strips run from the base of the crypt to the tip of the villus, i.e. along the so-called 'crypt-villus' axis, and correspond to how strips are collected experimentally. Each of these strips can be considered as arising from a 'measurement' from the (hypothesised) 'true underlying population' of such strips. Equivalently, the underlying population can be considered to define the 'reference frame' with respect to which the (finite) samples are analysed (and without which samples have little meaning).

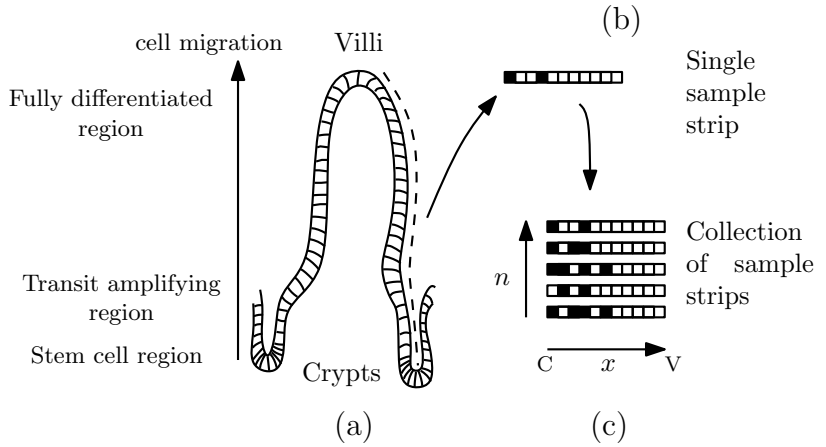

Figure 1. The (a) intestinal epithelium, (b) individual measurements as strips of cells and (c) collection of strips, where 'C' and 'V' indicated 'crypt' and 'villus' respectively.

Measurements can be given a spatial location index i and a time label t; we will represent these by a single index parameter $s := (i, t)$, when notationally convenient.

In our case we assume each strip is independent of the others as, in general, strips are taken from different crypt-villus units and/or animals after 'identical preparation'. Thus we do not ever directly possess, for example, measurements of a particular crypt with dimensions given in terms of a certain number of strips, though we will distinguish between the underlying population of strips (full ensemble) and a particular sample of strips at hand. Characterising a 'typical' reference crypt-villus unit by the two vectors $(\mathbf{L}, \mathbf{n})$, where $\mathbf{L}$ is the vector of labelled fractions at each grid point and $\mathbf{n}$ is the vector of number of samples at each grid point, gives a useful reduction of the (spatially) two-dimensional dynamics to a (spatially) one-dimensional problem. In general the dynamics of strips in a given crypt may be affected by those in the same crypt but we will not consider this additional complexity at present.

Given our choice of key observable as $\mathbf{y}$, the vector of counts of labelled cells at each grid point, and key 'ideal characteristics' of comparison $(\mathbf{L}, \mathbf{n})$, we will assume that all observations at a given grid point s are exchangeable (see [16, 17] for a formal definition and further discussion) conditional on $(\mathbf{L}, \mathbf{n})$. Satisfying this exchangebility condition implies the existence of a Bayesian probability model and is, in essence, a statistical reduction/symmetry assumption [16, 17]. Thus we will assume the existence of some $p(\mathbf{y}|\mathbf{L}, \mathbf{n})$, which links observables to an underlying population. Note that we will consider different experimental treatment conditions E, but the measurement

component will be taken to be independent of E - i.e. treatment will be taken to affect the underlying *process parameters* only (this is discussed in more detail in 'Summary of hierarchical structure').

A slight strenghtening [16, 17] of the general exchangeability assumption (which leads to a pure existence theorem) to the independence assumption discussed above leads to a particular model for the number of labelled cells y_s drawn in a collection of n_s cells at grid location s. This corresponds to a binomial distribution $\mathcal{B}$ with probability L_s at each grid location, giving the factorisation

$$p(\mathbf{y}|\mathbf{L},\mathbf{n}) = \prod_{s=0}^{S-1} \mathcal{B}(n_s, L_s). \tag{1}$$

Note that this certainly does not imply that the underlying parameters $L(i,t)$ are independent, rather that *if* the true parameters are known at each location then observations can be made independently at those locations (as discussed in the above).

Likelihood and normal approximation

The above justifies a measurement component $p(\mathbf{y}|\mathbf{L},\mathbf{n})$ of the full sampling model for the probability of a set of observed labelled cells $\mathbf{y}$ in samples of sizes $\mathbf{n}$ given the vector of modelled underlying labelled fractions $\mathbf{L}$. This also gives a likelihood function $\mathcal{L}$ for this measurement model, which is proportional to the probability given by the sampling model evaluated for the observed data and considered as a function of $\mathbf{L}$, i.e.

$$\mathcal{L}(\mathbf{L};\mathbf{y},\mathbf{n}) = \prod_{s=0}^{S-1} L_s^{y_s}(1-L_s)^{n_s-y_s} \propto p(\mathbf{y}|\mathbf{L},\mathbf{n}) = \prod_{s=0}^{S-1} \mathcal{B}(n_s, L_s). \tag{2}$$

It is also helpful to note that, for each s, if n_s is sufficiently large and L_s is not too close to 0 or 1 (e.g. $n_s L_s$ and $n_s(1-L_s) > 5$ is typical) then we may replace the binomial distributions $\mathcal{B}(n_s, L_s)$ by their normal approximations $\mathcal{N}(n_s L_s, n_s L_s(1-L_s))$. In this case we can use, denoting the set of all measured labelled fractions through the (useful, but slightly non-standard) notation $\mathbf{y}/\mathbf{n} := (y_1/n_1, ..., y_S/n_S)$,

$$\mathcal{L}(\mathbf{L};\mathbf{n},\frac{\mathbf{y}}{\mathbf{n}}) = p(\frac{\mathbf{y}}{\mathbf{n}}|\mathbf{L},\mathbf{n}) = \prod_{s=0}^{S-1} \frac{1}{\sigma_s\sqrt{2\pi}} \exp\left(-\frac{(\frac{y_s}{n_s} - L_s)^2}{2\sigma_s^2}\right) \quad (3)$$

where the standard deviations are given by $\sigma_s = \sqrt{\frac{L_s(1-L_s)}{n_s}}$. Since these standard deviations depend on the L_s in addition to the n_s we cannot drop them from the likelihood function. This normal approximation formulation is useful for the construction of approximate uncertainty intervals and, as we illustrate later, checking model misfit.

Process models

Discrete, measurement-grid-level process model

Here we first consider a formulation of our governing 'process' equations as a discrete probabilistic model. This captures the evolution of the (population) 'measurement' probability (labelled fraction) at the scale of the measurement grid. We will later consider an underlying continuous approximation to this model which discards some arbitrary details of the discrete model and so, though an approximation, it may in another sense be more generally applicable. It introduces a finer, continuous, spatial scale than the discrete measurement grid. The relationships between discrete and continuous representations motivates various possible modifications to, and interpretations of, our process equations, which we discuss later.

Again with reference to Fig 1, we consider a collection of one-dimensional 'strips' of cells. For our probabilistic discrete model, we use $l_i \in \{0, 1\}$ as an indicator variable denoting the occupancy status of site i of a given strip. The full state of this strip is then given by the vector $\mathbf{l} = (l_0, l_1, ...l_{S-1})$. We first seek a description of the probabilistic dynamics in terms of a discrete-time Markov chain for the probability distribution of the full state $p(\mathbf{l}, t)$ [39, 41]. Though we will begin from an explicit joint distribution for the full state we will show this reduces here to a description in terms of the set of 'single-site' probability distributions $p(l_i, t)$ for each site i.

For the following derivations it is helpful to adopt an explicit notation. Thus we will denote the probabilities of occupancy and vacancy at site i at time t by $p(l_i(t) = 1)$

and $p(l_i(t) = 0)$ respectively, and note that since $p(l_i(t) = 1) + p(l_i(t) = 0) = 1$ we only need to consider the probability of occupancy to fully characterise the distribution $p(l_i(t))$ (and, of course, similar properties hold for conditional distributions - see below).

The equation of evolution for this probability can be conveniently derived by considering conservation of probability in terms of probability fluxes in and out, giving, to first order in Δt

$$p(l_i(t + \Delta t) = 1) - p(l_i(t) = 1) = \Delta t \sum_{j=0}^{i-1} k_j \left[p(l_{i-1}(t) = 1, l_i(t) = 0) - p(l_{i-1}(t) = 0, l_i(t) = 1) \right]. \quad (4)$$

The first term on the right represents a net 'influx of occupancy probability' due to a single division event somewhere at site $j < i$, each division event having a probability given by $k_j \Delta t$. This flux leads to the value of the state variable $l_i(t) = 0$ being replaced, at the next time step, by the value of $l_{i-1}(t) = 1$. The second term similarly represents a net 'outflux of occupancy probability' due to a division event somewhere at site $j < i$.

We can further simplify this due to the binary nature of the occupancy state. To do this, note that since $l_i(t) = 0$ and $l_i(t) = 1$ partition the event space of $l_i(t)$, we have

$$p(l_{i-1}(t) = 1, l_i(t) = 0) = p(l_{i-1}(t) = 1) - p(l_{i-1}(t) = 1, l_i(t) = 1) \quad (5)$$

and similarly, since $l_{i-1}(t) = 0$ and $l_{i-1}(t) = 1$ partition the event space of $l_{i-1}(t)$, we have

$$p(l_{i-1}(t) = 0, l_i(t) = 1) = p(l_i(t) = 1) - p(l_{i-1}(t) = 1, l_i(t) = 1). \quad (6)$$

This leads to the above-mentioned simplification in terms of only single-site probability distributions

$$p(l_i(t+\Delta t)=1)-p(l_i(t)=1)=\Delta t\sum_{j=0}^{i-1}k_j\left[p(l_{i-1}(t)=1)-p(l_i(t)=1)\right]. \quad (7)$$

Underlying continuous model

Here we introduce a smooth parameter field of 'true' labelled fractions $L(x,t)$, now defined over a continuous space-time domain, and which corresponds to a further idealisation of the 'underlying population' from which we envisage the strips are sampled. The smoothness assumption, while not strictly necessary, means we will be able to give an interpretation of some model properties in terms of local derivatives; it also makes future comparisons with off-lattice and/or continuum models (see [42] for a review of various model types) more directly possible, and reduces arbitrary dependence on discrete grid features.

The position x is taken to be a continuous coordinate passing through the discrete cell indices, e.g. $x=0$ denotes the coordinate of the cell labelled '0' (base of the crypt), while $x=0.5$ is the location halfway between the cell labelled '0' and that labelled 1'. We will indicate sample locations consisting of space-time pairs by $s=(x_s,t_s)$. Then, for sample locations (i,t) corresponding to cell indices and arbitrary times, we match the discrete model and continuous model using

$$p(l_i(t)=1|L(i,t))=L(i,t) \quad (8)$$

i.e. $L(i,t)$ serves as the parameter for a single measurement modelled as a Bernoulli trial at that sample location (as in the above measurement model section).

Now, we consider how the discrete dynamics of $p(l_i(t)=1)$ can be transferred to the continuous $L(x,t)$ dynamics. In particular, since $L(x,t)$ is a smooth function, we can make the correspondence

$$p(l_{i-1}(t)=1|L(i-1,t))=L(i-1,t)$$

$$\equiv$$

$$p(l_{i-1}(t)=1|L(i,t),L_x(i,t),...,\Delta x)=L(i,t)-\Delta x\frac{\partial L(i,t)}{\partial x}+\frac{\Delta x^2}{2}\frac{\partial^2 L(i,t)}{\partial x^2}-... \quad (9)$$

where $\Delta x = i-(i-1)=1$ is the normalised cell length and we have also conditioned on knowledge of the spatial derivatives at i, $L_x(i,t)=\frac{\partial L(i,t)}{\partial x}$ etc. Here the continuous spatial field can be considered to be interpolating between - i.e. *internal* to - points of the discrete grid, making use of local derivative information. Substituting the above Taylor series, and similar expressions, into the discrete Markov equation 7 leads to

$$\frac{\partial L(i,t)}{\partial t}+v(i)\frac{\partial L(i,t)}{\partial x}=\frac{1}{2}\left(\Delta x v(i)\frac{\partial^2 L(i,t)}{\partial x^2}-\Delta t\frac{\partial^2 L(i,t)}{\partial t^2}\right)+... \quad (10)$$

where we have, for completeness, also retained higher order terms in Δt for the continuous model. We have also similarly assumed the existence of smooth functions $k(x,t)$ and $v(x,t)$ that satisfy the discrete relations

$$v(i,t)=\sum_{j=0}^{i-1}k_j\Delta x=\int_0^i k(x,t)dx+v(0). \quad (11)$$

We will assume $k(x,t)=k(x)$, $v(x,t)=v(x)$ and $v(0)=0$ in what follows.

Now we obtain closure for the continuous model by keeping only the lowest order terms in both time and space, and further asserting that the equation structure obtained should *hold for all (continuous)* x and not just (discrete) i (this can also be motivated by an assumption of grid translation invariance). This gives the advection equation

$$\frac{\partial L(x,t)}{\partial t}+v(x)\frac{\partial L(x,t)}{\partial x}=0 \quad (12)$$

with

$$v(x) = \int_0^x k(x')dx'. \tag{13}$$

The above partial differential equation (PDE) has a straightforward continuum interpretation as the advection of a tracer in an incompressible fluid field with a source, and is sometimes referred to in this context as the 'colour equation' [43]. We will take advantage of this in analysing model properties. Furthermore, it is straightforward to show that incorporating cell death, with discrete rates d_i, leads to the same equations with k replaced by $k - d$, where $d(x,t)$ is defined similarly to $k(x,t)$, and hence we will interpret k in the above as the net cell production rate.

Higher-order spatial effects Our smooth interpolation of the discrete set of equations into a continuous equation is a process of 'continualisation' - the reverse process of discretising a continuous equation to obtain a numerical scheme (see e.g. [44] and [43] Section 8.6 for similar ideas). Thus we might expect that a better continuum approximation could be obtained by retaining higher-order spatial derivatives and hence finite-cell-size effects.

As one possible model improvement and/or perturbation, here we reconsider the continuous approximation of our discrete Markov equation. In particular we consider retaining the higher-order spatial derivative. This naturally gives rise to a Fokker-Planck equation containing a diffusion term [41]. Equations of this (and similar) form have been derived before, also based on continuous approximations to discrete master equations (e.g. [45–48] also contain similar ideas; [42] give additional references).

Reconsidering the expansion in 10, we will again neglect all terms of order Δt; however we now retain the next order spatial derivative leading to

$$\frac{\partial L(x,t)}{\partial t} + v(x)\frac{\partial L(x,t)}{\partial x} = D(x)\frac{\partial^2 L(x,t)}{\partial x^2} \tag{14}$$

where $D(x) = (1/2)\Delta x v(x)$.

Here retaining the second spatial derivative amounts to accounting for spatial effects

due to finite cell sizes. Though we will first evaluate our original 'zeroth-order' (advection) model against our data, we will later examine the extent to which higher-order spatial terms such as those considered here may account for any misfits.

Parameter model

Within a Bayesian approach the choice of prior distribution is as important a part of model construction as the choice of sampling model (measurement and process components) and can be used to include relevant scientific information and constrain or 'regularise' the physical and mathematical characteristics of solutions ([17] provides an applied account of this approach while [37] presents a more philosophical perspective). In particular we can constrain the variability of the (here) spatially varying parameter field *a priori* to help avoid unphysical solutions, as well as vary these prior assumptions to explore the solution dependence on parameter variability.

We represent possible proliferation profiles, varying with cell locations, as realisations from a prior given in terms of a discretised random field (a random vector) $\mathbf{k}$ of length m, modelled as a multivariate Gaussian $\mathcal{N}(\boldsymbol{\mu}, \mathbf{C})$ with joint distribution

$$p(\mathbf{k}) = \frac{1}{(2\pi)^m \sqrt{\det(\mathbf{C})}} \exp(-(\mathbf{k}-\boldsymbol{\mu})^T \mathbf{C}^{-1} (\mathbf{k}-\boldsymbol{\mu})/2). \tag{15}$$

The properties of this distribution are determined by its mean vector $\boldsymbol{\mu}$ and covariance matrix $\mathbf{C}$. Furthermore, we decompose the latter into a standard deviation matrix given by the outer (tensor) product of the standard deviation vector for each variable, $\mathbf{S} = \boldsymbol{\sigma}\boldsymbol{\sigma}^T$, and correlation matrix $\mathbf{R}$. These multiply element-wise to give $C_{ij} = S_{ij} R_{ij}$ (no summation).

Another common, equivalent, representation is $\mathbf{C} = \mathbf{DRD}$ where $\mathbf{D}$ is a diagonal matrix with diagonal entries $D_{ii} = \sigma_i$. Decomposing the covariance into these separate parts makes clearer how we can control the smoothness and magnitude of variations via off-diagonal and diagonal terms, respectively, in addition to the mean response.

We will assume that the correlation matrix $\mathbf{R}$ is obtained from the squared-exponential (Gaussian) correlation function $k(x, x') = \exp(\frac{(x-x')^2}{2l_c^2})$ where l_c is

a parameter controlling the characteristic length-scale of the correlations - e.g. the scale over which the correlation function decays to $1/e$ - and x and x' are two spatial locations. This allows us to control the 'smoothness' of the realisations from the prior, in the sense that as l is increased the values at x and x' tend to be more similar. We generate the matrix $\mathbf{R}$ by evaluating this correlation function at the locations of a discretised grid of proliferation activity (here chosen to be coarser than the cell index grid), and l_c can then also be interpreted as a 'parameter correlation length', a measure number of the parameters over which the correlations decay. We will typically consider correlation lengths of 1-2 parameters.

In line with our approach of providing a 'simulation' or 'sampling-based' interpretation of a given model structure, it is typically most informative to visualise realisations of the whole function from the resulting prior rather than simply give the individual parameters/matrices separately ([18] discusses this in more detail). These will be discussed and displayed in more detail in the Results section below.

Summary of hierarchical structure

Here we show how to piece together the above model components.

We begin from a full joint distribution for a given experimental treatment E and known sample size vector $\mathbf{n}$, decomposing it according to

$$p(\mathbf{y}, \mathbf{L}, \mathbf{k}|\mathbf{n}, E) = p(\mathbf{y}|\mathbf{L}, \mathbf{n})p(\mathbf{L}|\mathbf{k})p(\mathbf{k}|E). \tag{16}$$

This hierarchical factorisation makes clear the conditional independence closure assumptions which separate the dependencies of the various levels, i.e. $p(\mathbf{y}|\mathbf{L}, \mathbf{k}, \mathbf{n}, E) = p(\mathbf{y}|\mathbf{L}, \mathbf{n})$, $p(\mathbf{L}|\mathbf{k}, \mathbf{n}, E) = p(\mathbf{L}|\mathbf{k})$ and $p(\mathbf{k}|\mathbf{n}, E) = p(\mathbf{k}|E)$. Furthermore, a 'causal' - structural invariance - assumption is made in assuming that the experimental treatment condition only affects the process parameter $\mathbf{k}$ but not the structure of the measurement or process models (see e.g [31–34, 49–52] for discussions of causal modelling and structural invariance).

The above factorisation can, and will, be checked to some extent in what follows. In

particular, we distinguish working 'within' the model - e.g. parameter estimation assuming the model and factorisation is valid - and working 'outside' the model - e.g. checking the validity of the model structural assumptions themselves (see [17, 36, 37] for relevant discussion of the distinction). From now on we will suppress the sample size $\mathbf{n}$ as it will be taken to be fixed and known, and will also in general suppress the conditioning on E (keeping in mind that it only affects $\mathbf{k}$).

As implicit in the model derivations above, we use a *deterministic expression of conservation of probability* (as is typical for Markov/master/Fokker-Planck equations; see [41]) for our process model and so the relationship between any set of process parameters $\mathbf{k}$ and the output process variable $\mathbf{L}$, $p(\mathbf{L}|\mathbf{k})$, can be (formally) taken to be a Dirac delta function centred on the relationship described by the 'true' model (see e.g.[18]). It suffices for our purposes to simply replace all functional dependencies on the process variable above with a dependence on the process parameters.

Computational methods

Interpretation and implementation of model updating and inference

The above provides a 'predictive' or 'generative' probabilistic model. Thus, given an initial specification ('prior parameter model' and 'forward/process model') of each of these parts we can construct 'prior predictions' according to (see e.g. [17, 42])

$$\mathbf{y} \sim p(\mathbf{y}) = \int p(\mathbf{y}|f(\mathbf{k}))p(\mathbf{k})d\mathbf{k} \tag{17}$$

where we have made use of the aforementioned deterministic link between a given sample of process parameters and the output process variable, $\mathbf{L} = f(\mathbf{k})$. Here we follow statistical standard notation where $\sim$ denotes 'distributed as', or more relevantly, 'samples drawn according to'.

If, or when, new data $\mathbf{y_0}$ become available one can update the parameters of the model to improve the predictive model and hence pass to a 'posterior predictive' model (again, see e.g. [17])

$$\mathbf{y}|\mathbf{y_0} \sim p(\mathbf{y}|\mathbf{y_0}) = \int p(\mathbf{y}|f(\mathbf{k}))p(\mathbf{k}|\mathbf{y_0})d\mathbf{k} \tag{18}$$

where we have introduced the conditional probability closure assumption $p(\mathbf{y}|f(\mathbf{k}), \mathbf{y_0}) = p(\mathbf{y}|f(\mathbf{k}))$. This closure assumption may be interpreted as maintaining our same mechanistic model despite new observations, and thus again connects well with current theories of causality as based on ideas of structural invariance [31–34, 49–52].

The logical flow is depicted in Fig 2. If we consider proceeding from the 'lowest' level to the 'highest' level, then each lower level draw gives a parameter (or set of parameters) to be plugged into the higher level distributions and generate further draws, ultimately producing predictions to be compared to actual data. Distributions are updated in the 'reverse' manner by conditioning at the highest level and propagating the implications of this back down the hierarchy. The conditional independence probability statements separating the levels may be considered as testable modelling 'closure' assumptions.

To implement the updating from prior to posterior parameter distributions, given measurements, we used Monte Carlo Markov Chain (MCMC) sampling (see [54] for a comprehensive reference). In particular, we used the (open source) Python package *emcee* (http://dan.iel.fm/emcee/) which implements an 'affine-invariant ensemble sampler' MCMC algorithm and has been applied in particular to astrophysics problems (see [55] for details). Given samples from the resulting prior and posterior parameter distributions, respectively, prior and posterior predictive distributions were obtained by forward simulation as described above. Note that each candidate proliferation rate vector $\mathbf{k}$ is connected to the measurements $\mathbf{y}$ via the latent vector $\mathbf{L}$; since this step is deterministic, however, no additional sampling steps are required for the process model component.

Interpretation of statistical evidence

As discussed above, our interpretation of mechanism/causality is as structural invariance under different scenarios. Even given this, we still need to provide an

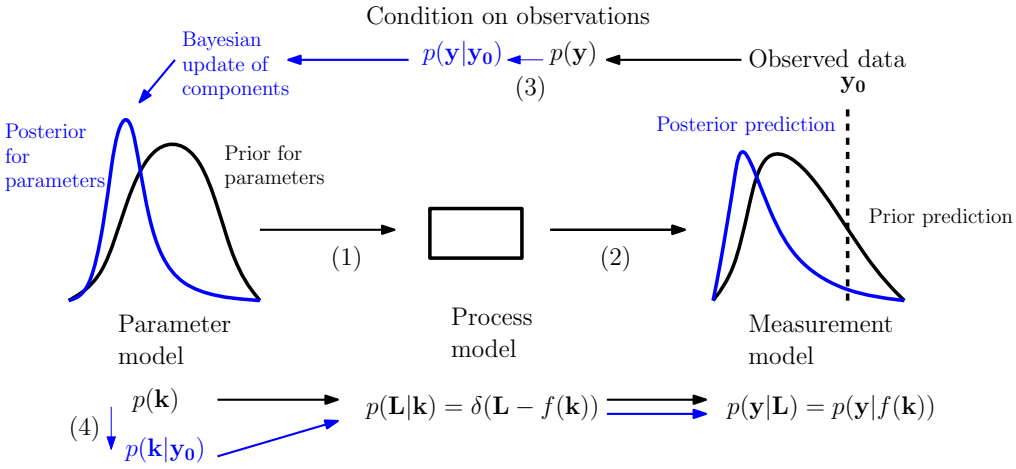

Figure 2. Illustration of the Bayesian predictive and parameter inference processes. Following the arrows (1) to (2) we move from a prior parameter model (left, black) to associated predictive distribution (right, black) via the process and measurement models. Following the arrows (3) to (4) we condition on observed data to obtain a posterior parameter model (left, blue) and associated predictive distribution (right, blue). Our structural assumptions mean that the information gained is represented in updates of the parameter model while the process and measurement models maintain the same form. Modified from [42], which was based on [53].

interpretation of the 'evidence' that a set of measurements provides about parameter values within a fixed model structure. There is still a surprising amount of controversy and debate about fundamental principles and definitions of (statistical) evidence, however (see e.g. [56, 57]).

Following our conditional modelling approach, here we adopt the simple - yet generally applicable - principle of evidence based on conditional probability: if we observe b and $p(a|b) > p(a)$ then we have evidence for a. A 'gold-standard' theory of statistical evidence starting from this premise has been developed and defended recently by Evans in a series of papers (summarised in [36]). Besides simplicity, a nice feature of this approach that we will use is that it can be applied both to prior and posterior predictive distribution comparisons such as $p(\mathbf{y}|\mathbf{y_0}) \stackrel{?}{>} p(\mathbf{y})$, as well as to prior and posterior parameter distribution comparisons such as $p(\mathbf{k}|\mathbf{y_0}) \stackrel{?}{>} p(\mathbf{k})$. In practice, we emphasise the visual comparison of various prior and posterior distributions ([18] advocates a similar 'movie strategy' for the interpretation of statistical evidence and inference procedures).

Differential equation solvers

In all sections other than the later section in which we include higher-order spatial effects, we solve the differential equation model using the *PyCLAW* [58, 59] Python interface to the *CLAWPACK* [60] set of solvers for hyperbolic PDEs. We adapted a Riemann solver for the colour equation available from the Riemann solver repository (https://github.com/clawpack/riemann). For testing the inclusion of higher-order spatial effects (thus changing the class of our equations from hyperbolic to parabolic) we used the Python finite-volume solver *FiPy* [61].

Jupyter Notebook and summary of main packages used

Our code is available in the form of a Jupyter (previosuly IPython) Notebook (http://ipython.org/notebook.html) in the Supplementary Information. To run these we used the Anaconda distribution of Python (https://store.continuum.io/cshop/anaconda) which is a (free) distribution bundling a

number of scientific Python tools. Any additional Python packages and instructions which may be required are listed at the beginning of our Jupyter Notebook.

Results

Parameter inference under homeostatic (healthy) conditions

Fig 3 illustrates the process of updating from (realisations from) the prior (pre-data) distributions of the proliferation and velocity fields to (realisations from) their posterior (post-data) distributions. The left-hand side of the figure shows simulations from the prior distribution for proliferation field (top) and realisations from the induced distribution for the velocity field (bottom), respectively. The right-hand side shows the corresponding simulations after the prior parameter distribution has been updated to a posterior parameter distribution. The prior-to-posterior parameter estimation was carried out using the MCMC sampling approach described above with $t = 120$ min (2 h) as an initial condition and $t = 360$ min (6 h) and 600 min (10 h) as given data. The initial condition for the underlying labelled fraction was determined by fitting a smoothing spline to the data. The prior distribution for the proliferation field shown in Fig 3 incorporated a weak mean trend in net proliferation rates, rising from the crypt base to the mid crypt before falling exponentially to zero over the last few parameter regions post-crypt end, and a parameter correlation length of 1. These assumptions can be relaxed/varied with little effect, though typically a non-zero parameter correlation length and a shut-off in proliferation after the crypt end produce more stable (well-identified) estimates. As mentioned, the code is available for use and so these assumptions are able to be varied by future researchers.

Parameter inference for blocked proliferation conditions

Fig 4 is the same as Fig 3 described in the previous section, but this time under treatment by Ara-C. The previous results from the baseline case are shown in grey, while the new results under Ara-C treatment are shown in blue. Here 1140 min (19 h post IdU labelling, 2 h post Ara-C treatment) was used as the initial condition and

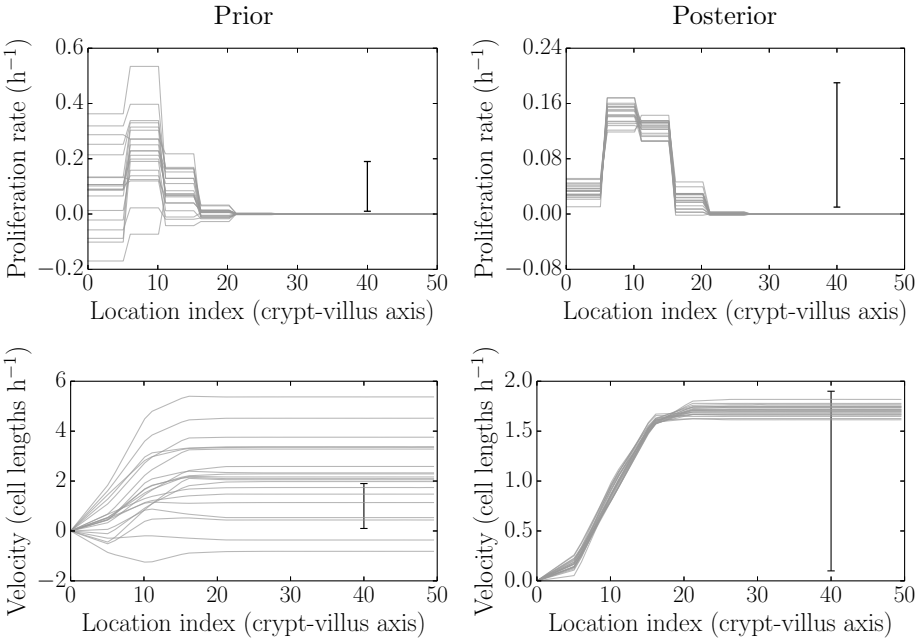

Figure 3. Simulated realisations from the prior (left) and posterior (right) distributions for proliferation profiles (top) and velocities (bottom). After data are obtained the posterior distributions are much more tightly-constrained, and are picking out biologically plausible results (see main text).

1500 min (25 h post IdU labelling, 8 h post Ara-C treatment) used for fitting. The intermediate time 1260 min (21 h post IdU labelling, 4 h post Ara-C treatment) and later time 1620 min (27 h post IdU labelling, 10 h post Ara-C treatment) were used as out-of-sample comparisons (see later).

As can be seen, there is a clear inhibition of proliferation and an even clearer effect on the migration (growth) velocity. The underlying parameter results are clearly more variable than those in the baseline case. This may indicate, for example, greater parameter underdetermination and/or inconsistency of the model. This is not surprising as we expect all the proliferation parameters to be reduced to similar (low) values and hence the parameters become less distinguishable.

To add additional stability to the results we can attempt to reduce underdetermination in the parameters by increasing the parameter correlation length and inducing an effectively more 'lumped' representation of the parameter field (since values tend to stick together more). Doing this removed the more extreme negative net proliferation in the posterior profile, however it still allowed for small amounts of

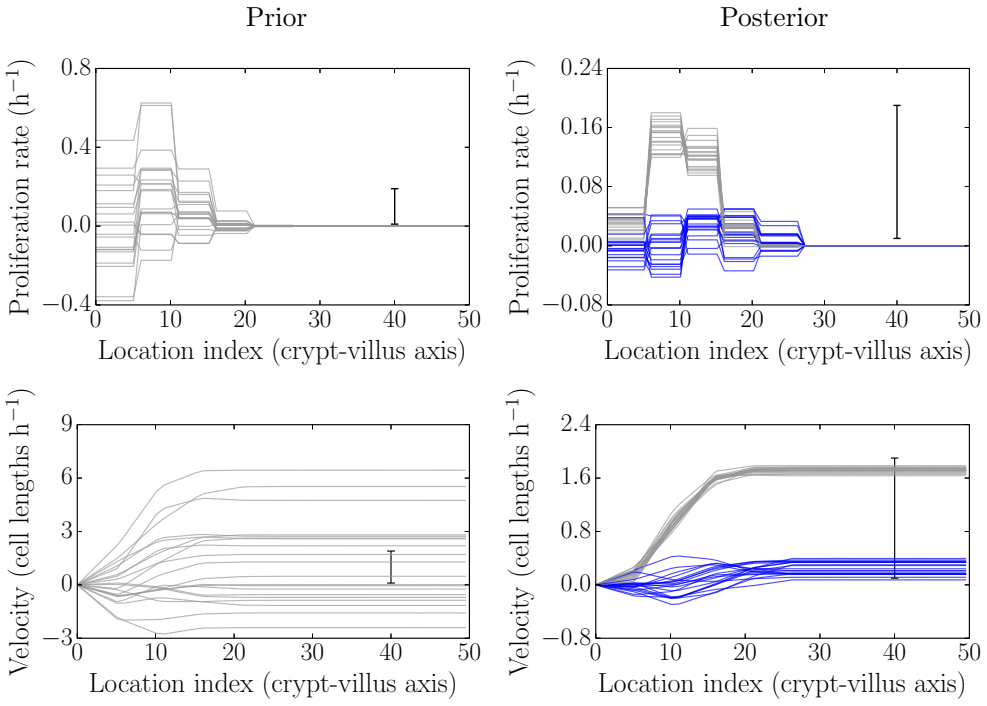

Figure 4. Simulated realisations from the prior (left) and posterior (right) distributions for proliferation profiles (top) and velocities (bottom) under Ara-C treatment (blue) as compared to no treatment (grey). The velocities are reduced to near zero, as are the proliferation rates, though the latter are noisier.

negative net proliferation/velocity (the available Jupyter notebook can be used to explore various prior assumptions).

Again, the need to introduce more stability is likely due to some combination of the limitations of resolution, a consequence of trying to fit the data too closely, or an indication of model inadequacies. In particular, under inhibited-profileration conditions the effective number of parameters would be expected to be reduced. When fitting the full model, with largely independent parameters for each region, it is to be expected that some additional regularisation would be required for greater stability.

Parameter inference for recovering proliferation conditions

Ara-C is metabolised between 10-12 h post-treatment. The two times considered here, 1620 min and 2520 min, correspond to 10 h and 25 h post Ara-C-treatment, respectively, i.e to the end of the effect and after the resumption of proliferation. Hence, to check for the recovery of proliferation, we fitted the model using 1620 min as the initial condition and 2520 min as the final time.

Fig 5 is the same as Fig 3 and Fig 4 described in the previous sections, but this time after/during recovering from treatment by Ara-C. The previous results from the baseline case are shown in grey, while the new results following recovery from Ara-C treatment are shown in blue. Here 1620 min (27 h post IdU labelling, 10 h post Ara-C treatment) was used as the initial condition and 2520 min (42 h post IdU labelling, 25 h post Ara-C treatment) used for fitting. We did not make additional out-of-sample comparisons in this case, though in-sample posterior predictive checks were still carried out (see later).

Here the proliferation and velocity profiles indicate that proliferation has resumed, as expected. The rates of proliferation appear to be lower than under fully healthy conditions, however, perhaps due to incomplete recovery (the initial condition being right at the beginning of the recovery period). The timing of the recovery of proliferation and the well-identified proliferation and velocity profiles inferred give no indication that any other mechanism is required to account for these data, however.

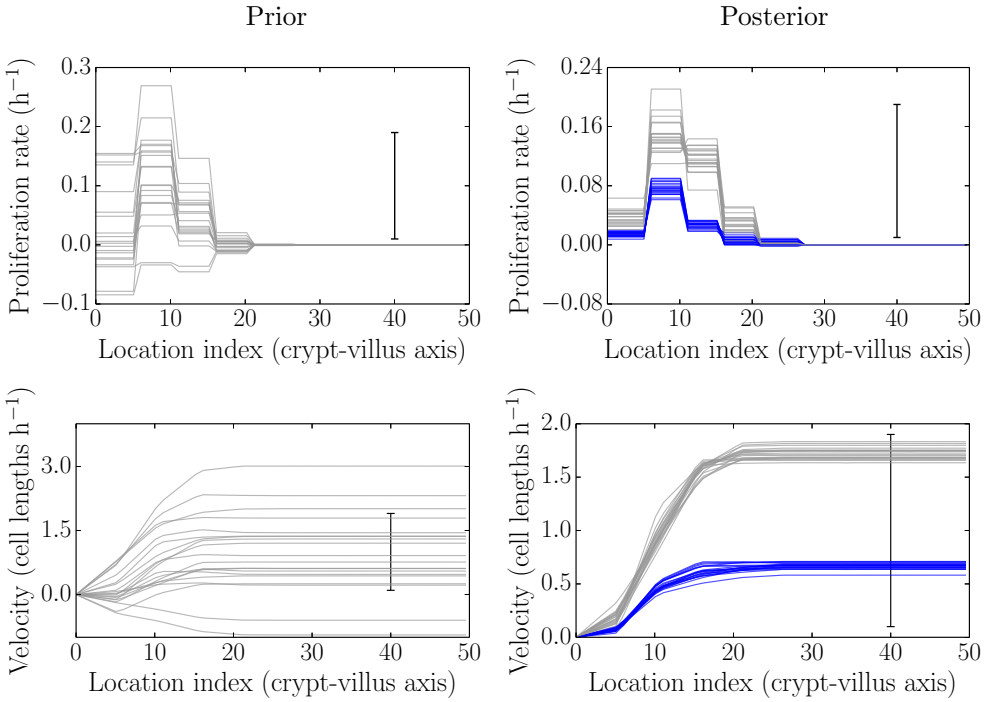

Figure 5. Simulated realisations from the prior (left) and posterior (right) distributions for proliferation profiles (top) and velocities (bottom) after recovery from Ara-C treatment (blue) as compared to no treatment (grey). The velocities and proliferation rates show a clear recovery towards healthy conditions, though not to the full level.

Predictive checks under homeostatic (healthy) conditions

Fig 6 illustrates simulations from the predictive distributions corresponding to the prior and posterior parameter distributions of Fig 3. This enables a first self-consistency check - i.e. can the model re-simulate data similar to the data to which it was fitted [17, 62]. If this is the case then we can (provisionally) trust the parameter estimates in the previous figure; if this was not the case then the parameter estimates would be unreliable, no matter how well-determined they seem. Here the model appears to adequately replicate the data used for fitting.

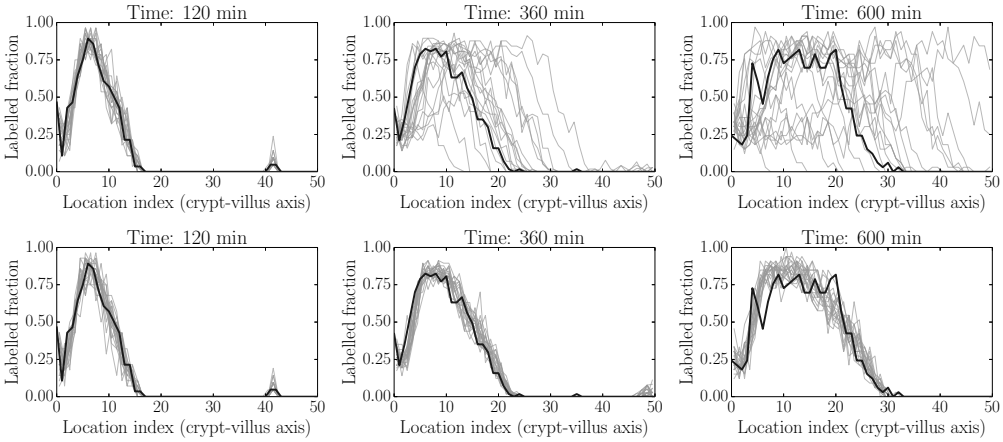

Figure 6. Simulated realisations from prior (top) and posterior (bottom) predictive distributions (grey) for label data at fitted times (120 min, 360 min and 600 min i.e. 2 h, 6 h and 10 h). Actual data are indicated by black lines. Again the posterior distributions are much more constrained than the prior distributions, representing the gain in information from collecting (and fitting to) experimental data. The first profile in each panel is held as a constant initial condition in this example.

Fig 7 and Fig 8 illustrate two additional ways of visualising replicated datasets. The former visualises the label profile along the crypt-villus axis at the future unfitted/out-of-sample time 1080 min (18 h), while the latter visualises both fitted (120 min/2 h, 360 min/6 h and 600 min/10 h) and unfitted/out-of-sample (1080 min/18 h) predictions plotted in the characteristic plane (t, x) in which the slopes along lines of constant colour should be inversely proportional to the migration velocities at that point, due to the (hyperbolic) nature of our 'colour' equation (see e.g. [43]). We have interpolated between the dotted grid lines. These figures, in combination with Fig 6 above, indicate that the model is capable of reliably reproducing the data to which it was fitted, as well as predicting key features of

unfitted datasets such as the rate of movement of the front. On the other hand, there is clearly a greater misfit with the predicted rather than fitted data. In order to locate the possible source of misfit we considered various model residuals and error terms - see 'Locating model misfit' below.

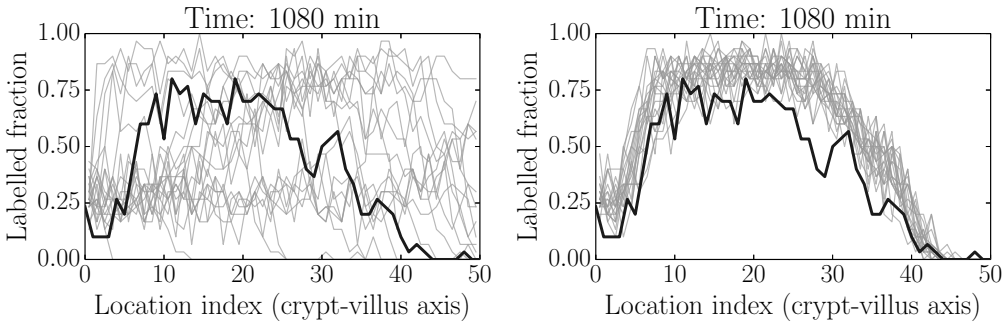

Figure 7. Simulated realisations from prior (left) and posterior (right) predictive distributions (grey) for label data at the unfitted (out-of-sample) time 1080 min (18 h). Actual data are indicated by black lines. The model appears to give reasonable predictions capturing the main effects, but there is also clearly some misfit to be explored further.

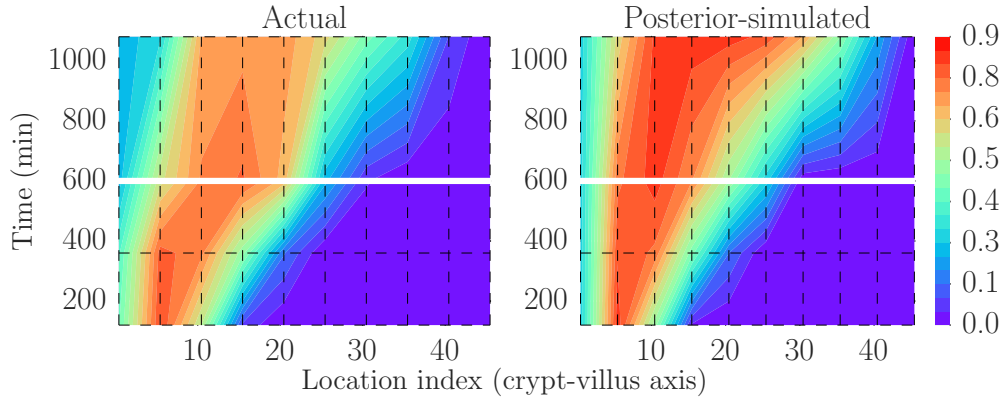

Figure 8. Actual (smoothed) data (left, black box) and one replication based on the model (right; plotting the latent/measurement-error-free process) as visualised in the characteristic plane. This has been discretised and interpolated between the dotted lines to facilitate fair but coarse-grained comparisons. The model structure implies that there should be lines of constant colour tracing out curves with slopes inversely proportional to the migration velocities at that point. The model again captures a number of these key qualitative features, but also fits less well for the out-of-sample (above the horizontal gap at 600 min/10 h) data. There is little variability in the latent model process and so only one replication is shown.

Predictive checks under blocked proliferation conditions

Here 1140 min (19 h; post IdU labelling) was used as the initial condition and 1500 min (25 h) used for fitting. 1260 min (21 h) and 1620 min (27 h) were used as out-of-sample (non-fitted) comparisons. Fig 9 is analogous to Fig 6 in the healthy case. In general all of the features up to 1620 min (27 h) in Fig 9, and for both fitted and predicted times, appear to be reasonably well captured. The fit at 1620 min is generally good, but perhaps worse than the other cases. This could be due to both errors in longer-time predictions and to the beginning of proliferation recovery. We explore both longer-time misfits and recovering proliferation conditions in what follows.

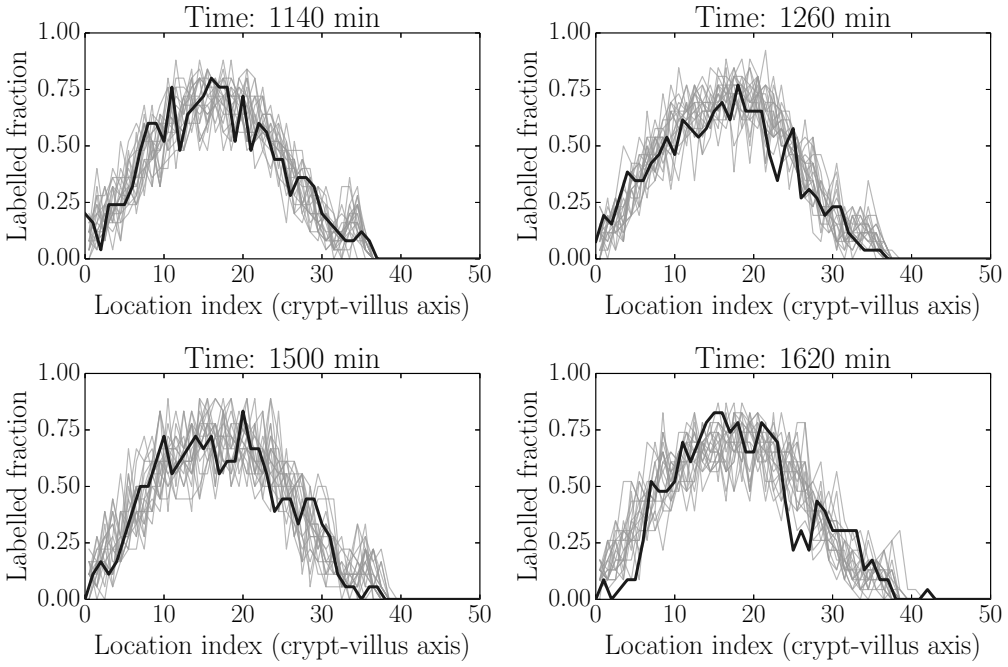

Figure 9. Simulated realisations from posterior predictive distributions (grey) for label data at 1140 min (initial condition), 1500 min (fitted time) and at two out-of-sample/unfitted times (1260 and 1620 min). The posterior distributions appear to adequately capture the actual label data (black).

Predictive checks under recovering proliferation conditions

As discussed above, Ara-C is metabolised between 10-12 h post-treatment. The two times considered here, 1620 min and 2520 min, correspond to 10 h and 25 h post

Ara-C-treatment, respectively, i.e to the end of the effect and after the resumption of proliferation.

Again, as expected, the label has resumed movement in concert with the resumption in proliferation. The model appears to fit reasonably well

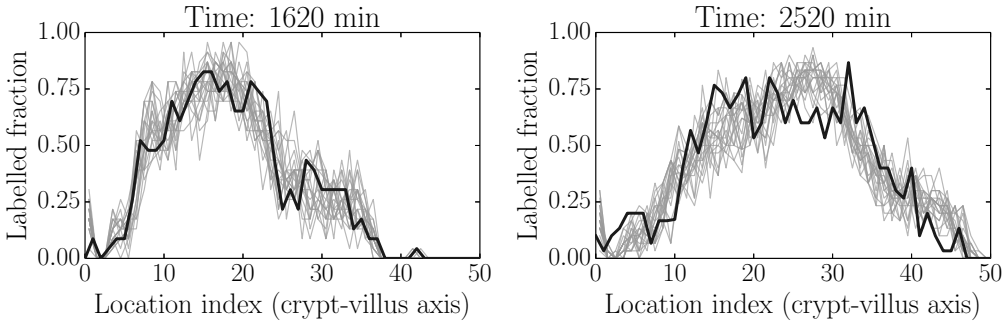

Figure 10. Simulated realisations from posterior predictive distributions (grey) for label data at 1620 min (initial condition) and 2520 min (fitted). These indicate that proliferation has resumed, consistent with the time taken to metabolise Ara-C - see the main text for more detail.

Summary so far

We conclude that, despite some minor misfit, especially for the longest-time out-of-sample simulations and for the post-treatment conditions, the model behaves essentially as desired under experimental perturbation and indicates that it is likely capturing the essential features of interest.

Next we consider how we might explore possible sources of misfit.

Locating model misfit

Here we consider how to unpick the contributions of the various model parts to the above re-simulated and out-of-sample datasets. We base this on assessing the model adequacy under baseline (healthy) conditions as we are more confident of the experimental effects under this scenario.

Fig 11 shows the following checks: measurement error as determined by subtracting a smoothed spline from the observed data (dark line) and comparing these to the results obtained by subtracting the process model from the simulated data (panels 1-4,

moving left-to-right and top-to-bottom, showing fitted - 120 min/2 h, 360 min/6 h and 600 min/10 h - and unfitted/out-of-sample - 1080 min/18 h - times). This presentation follows the noise-checking approach in [63], as well as the general recommendations given in [17, 62]. Reliable interpretation of these as 'true' measurement residuals depends on the validity of the normal approximation 3 since these expressions are not directly interpretable in terms of the discrete binomial model (see e.g. [17, 62]). These are also visualised in terms of the corresponding cumulative distributions in the middle panel (panel 5, following as above). Panels 6-9 show the differences between the underlying process model and the smoothed spline fitted to the data. As can be seen, the measurement model appears approximately valid at all times, while the process model appears to have non-zero error for the 1080 min sample. We consider this in more detail next.

Possible model improvement and robustness - higher-order spatial effects

As discussed in the process model section above, the presence of cellular structure in the epithelial tissue means that higher-order spatial effects may be present. Here we consider to what extent these may account for the minor misfit identified above.

To give an idea of the qualitative differences induced by including these higher-order terms, consider Fig 12 in which we compare to the (healthy) 1080 min (18 h) data in which we found some indication of a process model error. Note that while this model appears slightly better able to fit the qualitative features of the data, both models require similar modification of the parameter values to quantitatively improve the fit to our out-of-sample data (we have illustrated both for a case where parameter values are reduced by 20%). Thus the key (yet relatively small) difference between the model and out-of-sample data is likely due to an effect other than finite-cell sizes; for example it may be due to time-variation in parameter values due to circadian rhythms (we have assumed steady-state parameter values), label dilution or an unmodelled mixing phenomenon in the full two-dimensional case. We note however that these effects are small and appear to be important primilarily for predicting much further ahead in time than the fitted data and the steady-state parameter assumption is likely

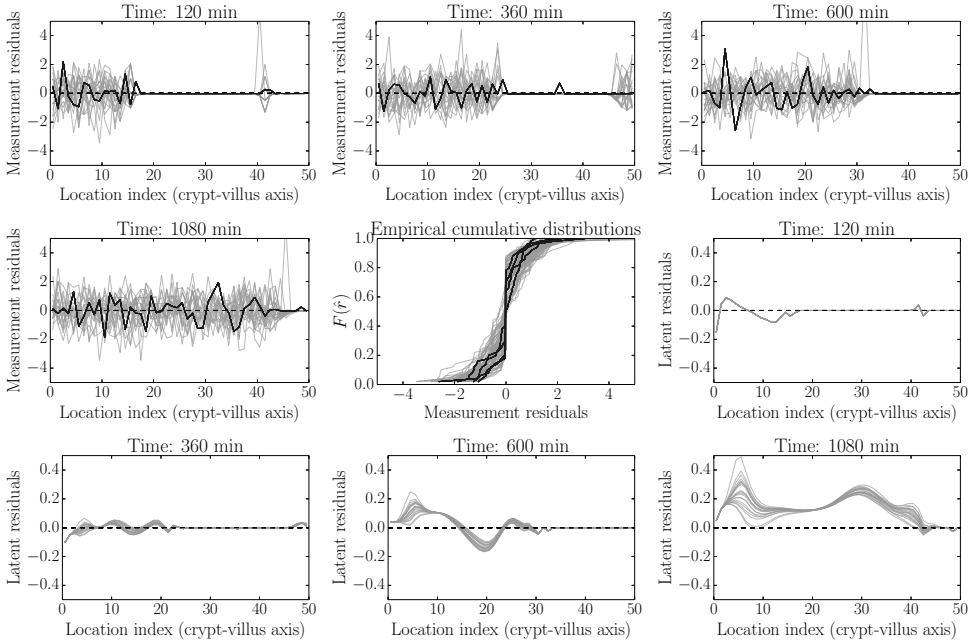

Figure 11. Model and data residual components. Panels 1-4, moving left-to-right and top-to-bottom, shows measurement error as determined by subtracting a smoothed spline from the observed data (dark line) and comparing this to the results obtained by subtracting the process model for fitted - 120, 360 and 600 mins - and unfitted/out-of-sample - 1080 min - times from the realised data (grey). These measurement error distributions are also visualised in terms of the corresponding cumulative distributions in the middle panel (panel 5, following as above. Black - actual data, grey - model simulations). Panels 6-9 show the differences between realisations of the underlying process model and the smoothed spline fitted to the data. As can be seen across panels, the measurement model appears approximately valid at all times, while the process model appears to have non-zero error for the 1080 min sample. This observation is discussed in the text.

valid for reasonable time intervals. This means that the more easily interpretable original model may be sufficient for many purposes.

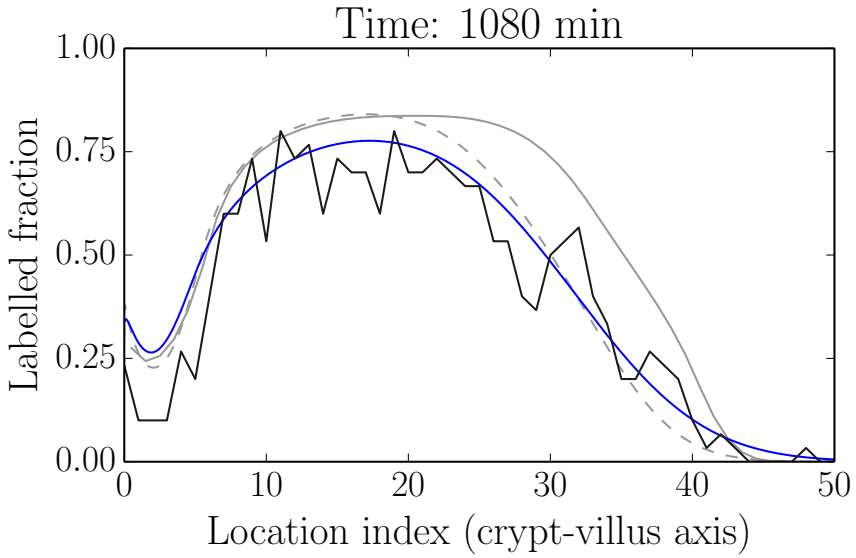

Figure 12. Comparison of the modified process model which includes higher-order spatial terms (blue) to the original model (grey, dashed), both at lowered proliferation rates (decreased 20%), which is required for a better fit to the data. The original model at the original fitted proliferation rates is also shown (grey, solid). Although the model with higher-order spatial terms gives a better qualitative fit to the data for the same proliferation rates, it is clear that the dominant cause of misfit is better attributed to (time) varying proliferation rates (in the context of the present set of models).

Discussion

Understanding the complicated dynamics of the intestinal epithelium requires an interdisciplinary approach involving experimental measurements, mathematical and computational modelling, and statistical quantification of uncertainties. While a diverse range of mathematical models have been constructed for epithelial cell and tissue dynamics (reviewed in [64–68]), from compartment models to individual-based models to continuum models, we lack consistent and reproducible frameworks for comparing models representing conjectured biological mechanisms both to each other and to experimental data (for an overview, see our review [42]). These shortcomings may explain why questions such as the connection between proliferation and migration

and its variation under experimental perturbations remain open, despite much investigation [8–14].

The aim of the present work was to acknowledge and confront these difficulties explicitly, and to present some initial constructive steps in establishing such a framework. To do this we carried out new experiments (described more fully in a companion paper [15]) aimed at determining how proliferation rates, tissue growth and cellular migration rates are related in the intestinal epithelium under healthy, damaged (Ara-C treated) and recovering conditions. We performed BrdU/IdU cell-labelling experiments under these respective conditions. In considering how to best process these data and interpret them using mathematical models, we then developed a probabilistic, hierarchical (conditional) framework.

Our hierarchical framework provides a best-practice approach for systematically modelling and understanding the uncertainties that have, in our view, prevented past studies in this area from providing reliable mechanistic conclusions - uncertainties in experimental measurement and treatment, difficult-to-compare mathematical models of underlying mechanisms, and unknown or unobserved parameters. Our approach was influenced by recognising the benefits that the hierarchical Bayesian approach has demonstrated in applications across a number of different disciplines (e.g. in environmental and geophysical science as in [22, 23]; ecological modelling as in [24, 25]; and in Bayesian statistical modelling and inverse problems more generally as in [17–21, 26]). We also note that a hierarchical approach can have significant benefits outside the Bayesian framework (see for example the 'extended likelihood' approach described in [27–29]).

The hierarchical approach has advantages not only in terms of providing a framework for combining disparate sources of uncertainty, but also as a framework for facilitating modelling derivations and relating discrete and continuous models. Though the resulting measurement, process and parameter models can or have all been derived by other means, as far as we are aware this particular perspective has not been systematically utilised in the same manner as considered here - at the very least it appears uncommon within the mathematical/systems/computational biology communities. Furthermore, in the main text we noted the connections of our

conditional, probabilistic approach for relating discrete and continuous models to similar procedures in the numerical analysis literature. This raises exciting connections to the developing field of probabilistic numerical methods and computing [69].

We also note the connection between the choice of a measurement model as required here (and/or process model error, and following e.g. [18–22, 70]), and the development of approximate sampling and parameter fitting procedures, which are particularly useful for analytically difficult models. A key concern of the latter is the appropriate choice of summary statistics for constructing a 'synthetic likelihood' [71] or similarly-modified posterior target for Approximate Bayesian Computation (ABC) [72–74]. This choice determines (implicitly or explicitly) in which ways a given model or set of models can be considered an 'adequate' representation of the data, which features are considered to be reproducible and what the associated 'noise' structure should be ([75] presents an alternative approach to characterising data features and model adequacy). These issues are crucial in deciding how to model the complexity of epithelial cell and tissue dynamics. An important next step, as described above, would be to bring more process model types and explicit measurement modelling into this framework and to evaluate and compare them under carefully modelled experimental conditions. Extensions incorporating other mechanical and/or cellular-level information (e.g. [11, 12]) would provide a natural next step.

As a final methodological point, by making our code and data available, as well as leveraging already-available open-source scientific Python software, we open up our work to other researchers to build on.

The main results established using the above framework were

- An adequate description of intestinal epithelial dynamics is achievable using a model based on purely proliferation-driven growth
- This model is consistent with healthy, proliferation-inhibited (Ara-C-treated) and recovering conditions
- The measurement and process model errors can be reasonably distinguished and checked separately

- This checking indicates that much of the natural variability is directly attributable to the collection process and this process can be modelled in a simple manner
- Possible model errors can also be identified and proposed explanations incorporated and tested within our framework, and thus the proper interpretation of experimental procedures is aided by using an explicit mathematical model and its predictive simulations
- Including finite-cell-size effects gives a slightly better qualitative fit to experimental data, but the dominant sources of the long-time misfits are likely due to some other factor such as (relatively slowly) time-varying proliferation rates (e.g. due to circadian rhythms) or label dilution.

Acknowledgements

This work was funded by the BBSRC-UK, project numbers BB/K018256/1, BB/K017578/1, BB/K017144/1 and BB/J004529/1 and the EPSRC-UK, project number EP/I017909/1.

References

1. Wright NA, Alison M (1984) The biology of epithelial cell populations. Oxford University Press, USA

2. Radtke F, Clevers H (2005) Self-renewal and cancer of the gut: Two sides of a coin. Science 307:1904–1909

3. Reuss L (2010) Epithelial transport. Comprehensive physiology. John Wiley & Sons, Inc.

4. van der Flier LG, Clevers H (2009) Stem cells, self-renewal, and differentiation in the intestinal epithelium. Annu Rev Physiol 71:241–260

5. Turner JR (2009) Intestinal mucosal barrier function in health and disease. Nat

Rev Immunol 9:799–809

6. Marchiando AM, Graham WV, Turner JR (2010) Epithelial barriers in homeostasis and disease. Annu Rev Pathol 5:119–144

7. Barker N (2014) Adult intestinal stem cells: Critical drivers of epithelial homeostasis and regeneration. Nat Rev Mol Cell Biol 15:19–33

8. Kaur P, Potten CS (1986) Cell migration velocities in the crypts of the small intestine after cytotoxic insult are not dependent on mitotic activity. Cell Tissue Kinet 19:601–610

9. Kaur P, Potten CS (1986) Circadian variation in migration velocity in small intestinal epithelium. Cell Tissue Kinet 19:591–599

10. Tsubouchi S (1983) Theoretical implications for cell migration through the crypt and the villus of labelling studies conducted at each position within the crypt. Cell Tissue Kinet 16:441–456

11. Dunn S-J, Näthke IS, Osborne JM (2013) Computational models reveal a passive mechanism for cell migration in the crypt. PLoS One 8:e80516

12. Meineke FA, Potten CS, Loeffler M (2001) Cell migration and organization in the intestinal crypt using a lattice-free model. Cell Prolif 34:253–266

13. Loeffler M, Stein R, Wichmann HE, Potten CS, Kaur P, Chwalinski S (1986) Intestinal cell proliferation. I. A comprehensive model of steady-state proliferation in the crypt. Cell Tissue Kinet 19:627–645

14. Loeffler M, Potten CS, Paulus U, Glatzer J, Chwalinski S (1988) Intestinal crypt proliferation. II. Computer modelling of mitotic index data provides further evidence for lateral and vertical cell migration in the absence of mitotic activity. Cell Tissue Kinet 21:247–258

15. Parker A, Maclaren OJ, Fletcher AG, Muraro D, Kreuzaler PA, Byrne HM, Maini PK, Watson AJM, Pin C (2016) Cell proliferation within small intestinal crypts is the principal driving force for cell migration on villi. The FASEB Journal; published

ahead of print October 20, 2016, doi:10.1096/fj.201601002

16. Bernardo JM, Smith AFM (2009) Bayesian theory. John Wiley & Sons

17. Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB (2013) Bayesian data analysis, third edition. Taylor & Francis

18. Tarantola A (2005) Inverse problem theory and methods for model parameter estimation. SIAM

19. Berliner LM (1996) Hierarchical Bayesian time series models. In: Maximum entropy and Bayesian methods. Springer Netherlands, pp 15–22

20. Cressie N, Wikle CK (2011) Statistics for spatio-temporal data. John Wiley & Sons

21. Wikle CK (2015) Modern perspectives on statistics for spatio-temporal data. WIREs Comput Stat 7:86–98

22. Berliner LM (2003) Physical-statistical modeling in geophysics. J. Geophys. Res. 108:

23. Wikle CK (2003) Hierarchical models in environmental science. Int Stat Rev 71:181–199

24. Cressie N, Calder CA, Clark JS, Ver Hoef JM, Wikle CK (2009) Accounting for uncertainty in ecological analysis: The strengths and limitations of hierarchical statistical modeling. Ecol Appl 19:553–570

25. Ogle K (2009) Hierarchical Bayesian statistics: Merging experimental and modeling approaches in ecology. Ecol Appl 19:577–581

26. Blei DM (2014) Build, compute, critique, repeat: Data analysis with latent variable models. Annual Review of Statistics and Its Application 1:203–232

27. Pawitan Y (2001) In all likelihood: Statistical modelling and inference using likelihood. OUP Oxford

28. Pawitan Y, Lee Y (2016) Wallet game: Probability, likelihood and extended likelihood. Am Stat 0:1–7

29. Lee Y, Nelder JA, Pawitan Y (2006) Generalized linear models with random

effects: Unified analysis via h-likelihood. CRC Press

30. Box GEP (1976) Science and statistics. J Am Stat Assoc 71:791–799

31. Pearl J (2009) Causality. Cambridge University Press

32. Pearl J (2009) Causal inference in statistics: An overview. Stat Surv 3:96–146

33. Dawid PA (2010) Seeing and doing: The Pearlian synthesis. In: Rina Dechter, Hector Geffner, and Joseph Y. Halpern (ed) Heuristics, probability and causality: A tribute to judea pearl. College Publications London, pp 309–325

34. Dawid PA (2010) Beware of the DAG! NIPS Causality: Objectives and Assessment 6:59–86

35. Box GEP (1980) Sampling and Bayes' inference in scientific modelling and robustness. J R Stat Soc Ser A 143:383–430

36. Evans M (2015) Measuring statistical evidence using relative belief. CRC Press

37. Gelman A, Shalizi CR (2013) Philosophy and the practice of Bayesian statistics. Br J Math Stat Psychol 66:8–38

38. Hartig F, Calabrese JM, Reineking B, Wiegand T, Huth A (2011) Statistical inference for stochastic simulation models – theory and application. Ecol Lett 14:816–827

39. Wilkinson DJ (2011) Stochastic modelling for systems biology. CRC Press

40. Gutmann MU, Corander J (2015) Bayesian optimization for Likelihood-Free inference of Simulator-Based statistical models. arXiv preprint arXiv:1501.03291

41. Van Kampen NG (1992) Stochastic processes in physics and chemistry. Elsevier

42. Maclaren OJ, Byrne HM, Fletcher AG, Maini PK (2015) Models, measurement and inference in epithelial tissue dynamics. Math Biosci Eng 12:1321–1340

43. LeVeque RJ (2002) Finite volume methods for hyperbolic problems. Cambridge University Press

44. Askes H, Metrikine AV (2005) Higher-order continua derived from discrete media: Continualisation aspects and boundary conditions. Int J Solids Struct 42:187–202

45. Baker RE, Yates CA, Erban R (2010) From microscopic to macroscopic

descriptions of cell migration on growing domains. Bull Math Biol 72:719–762

46. Hywood JD, Hackett-Jones EJ, Landman KA (2013) Modeling biological tissue growth: Discrete to continuum representations. Phys Rev E Stat Nonlin Soft Matter Phys 88:032704

47. Fozard JA, Byrne HM, Jensen OE, King JR (2010) Continuum approximations of individual-based models for epithelial monolayers. Math Med Biol 27:39–74

48. Murray PJ, Edwards CM, Tindall MJ, Maini PK (2009) From a discrete to a continuum model of cell dynamics in one dimension. Phys Rev E Stat Nonlin Soft Matter Phys 80:031912

49. Dawid PA (1979) Conditional independence in statistical theory. J R Stat Soc Series B Stat Methodol 41:1–31

50. Woodward J (2003) Making things happen: A theory of causal explanation. Oxford University Press

51. Woodward J (1997) Explanation, invariance, and intervention. Philos Sci 64:S26–S41

52. Spirtes P, Glymour CN, Scheines R (2000) Causation, prediction, and search. MIT Press

53. Iglesias M, Stuart A (2014) Inverse problems and uncerrainty quantification. SIAM news (july/August)

54. Robert C, Casella G (2013) Monte carlo statistical methods. Springer Science & Business Media

55. Foreman-Mackey D, Hogg DW, Lang D, Goodman J (2013) Emcee: The MCMC hammer. Publications of the Astronomical Society of the Pacific

56. Mayo DG (2014) On the birnbaum argument for the strong likelihood principle. Stat Sci 29:227–239

57. Evans M (2014) Discussion of “on the birnbaum argument for the strong likelihood principle”. Stat Sci 29:242–246

58. Ketcheson DI, Mandli K, Ahmadia AJ, Alghamdi A, Luna MQ de, Parsani M,

Knepley MG, Emmett M (2012) PyClaw: Accessible, extensible, scalable tools for wave propagation problems. SIAM J Sci Comput 34:C210–C231

59. Mandli KT, Ketcheson DI, others (2014) PyClaw software.

60. Clawpack Development Team (2014) Clawpack software.

61. Guyer JE, Wheeler D, Warren JA (2009) FiPy: Partial differential equations with python. Comput. Sci. Eng.

62. Gelman A (2004) Exploratory data analysis for complex models. J Comput Graph Stat 13:755–779

63. Aguilar O, Allmaras M, Bangerth W, Tenorio L (2015) Statistics of parameter estimates: A concrete example. SIAM Rev 57:131–149

64. Johnston MD, Edwards CM, Bodmer WF, Maini PK, Chapman SJ (2007) Examples of mathematical modeling: Tales from the crypt. Cell Cycle 6:2106–2112

65. Carulli AJ, Samuelson LC, Schnell S (2014) Unraveling intestinal stem cell behavior with models of crypt dynamics. Integr Biol 6:243–257

66. Kershaw SK, Byrne HM, Gavaghan DJ, Osborne JM (2013) Colorectal cancer through simulation and experiment. IET Syst Biol 7:57–73

67. De Matteis G, Graudenzi A, Antoniotti M (2013) A review of spatial computational models for multi-cellular systems, with regard to intestinal crypts and colorectal cancer development. J Math Biol 66:1409–1462

68. Fletcher AG, Murray PJ, Maini PK (2015) Multiscale modelling of intestinal crypt organization and carcinogenesis. Math Models Methods Appl Sci 25:2563–2585

69. Hennig P, Osborne MA, Girolami M (2015) Probabilistic numerics and uncertainty in computations. Proc Math Phys Eng Sci 471:20150142

70. Mosegaard K, Tarantola A (2002) Probabilistic approach to inverse problems. International Geophysics Series 81:237–268

71. Wood SN (2010) Statistical inference for noisy nonlinear ecological dynamic systems. Nature 466:1102–1104

72. Marin J-M, Pudlo P, Robert CP, Ryder RJ (2012) Approximate Bayesian

computational methods. Stat Comput 22:1167–1180

73. Wilkinson RD (2013) Approximate Bayesian computation (ABC) gives exact results under the assumption of model error. Stat Appl Genet Mol Biol 12:129–141

74. Ratmann O, Andrieu C, Wiuf C, Richardson S (2009) Model criticism based on likelihood-free inference, with an application to protein network evolution. Proc Natl Acad Sci U S A 106:10576–10581

75. Davies PL (2014) Data analysis and approximate models: Model choice, Location-Scale, analysis of variance, nonparametric regression and image analysis. CRC Press