A growth-induced dispersal model of non-motile cells calibrated against time-lapse microscopy

A thesis
Submitted towards the partial fullfilment of
BS-MS dual degree programme
by

SULTAN AHMED NAZIR

APRIL 2023

under the guidance of

DR. BRIAN P. INGALLS

UNIVERSITY OF WATERLOO

from June 2022 to March 2023

INDIAN INSTITUTE OF SCIENCE EDUCATION AND RESEARCH, PUNE

Certificate

This is to certify that this dissertation entitled *A growth-induced dispersal model of non-motile cells calibrated against time-lapse microscopy* submitted towards the partial fulfillment of the BS-MS degree at the Indian Institute of Science Education and Research, Pune represents original research carried out by Sultan Ahmed Nazir at University of Waterloo, under the supervision of Dr. Brian P. Ingalls, during academic year 2022 - 2023.

Dr. Brian P. Ingalls
Professor
University of Waterloo

Sultan Ahmed Nazir
20181010
BS-MS
IISER Pune

Declaration

I, hereby declare that the matter embodied in the report titled *A growth-induced dispersal model of non-motile cells calibrated against time-lapse microscopy* is the results of the investigations carried out by me at the University of Waterloo from the period 01-06-2022 to 31-03-2023 under the supervision of Dr. Brian P. Ingalls and the same has not been submitted elsewhere for any other degree.

Sultan Ahmed Nazir
20181010
BS-MS
IISER Pune

Acknowledgements

Foremost, I express my sincere gratitude to Brian for being a patient, kind, and resourceful mentor and an inspiration to the aspiring academic in me. I am incredibly grateful to Aaron for helping me navigate wet lab work and bringing a cheerful energy even when experiments were not necessarily working out. I would also like to thank other members of the Quantitative Cell Biology lab - Ati, Julien and Sara - for inspiring me with their research. I am also thankful to everybody else I shared the lab with for accommodating me with kindness. I thank Dr. Chaitanya Athale (IISER Pune), Dr. Nishad Matange (IISER Pune) and Dr. Matthew Scott (University of Waterloo) for taking out the time to engage in discussions that have helped me in different aspects of the research conducted for this thesis.

I am grateful to my home institution, IISER Pune, for literally being my home for the past five years and nurturing my interest in interdisciplinary research. I thank the INSPIRE-SHE program for financially supporting my academics and research during my undergraduate studies. I thank the Mitacs Globalink Research Internship program and the Department of Applied Mathematics at the University of Waterloo for assisting my stay in Canada financially and making necessary arrangements to provide a comfortable time during my research.

I am immensely grateful to my friends and family. Umma, Uppa and Ikka have been part and parcel of my achievements since birth. My friends have been a constant source of companionship and support during my time in IISER. Rosa is the cat who does not care for my undying affection and practically threatened me to mention her here. I am thankful to the members and allies of the 2SLGBTQIA+ communities in Waterloo, Kitchener and Toronto who were extremely welcoming and instilled a sense of community in me as I adapted to a new country.

Contributions

Contributor name	Contributor role
Sultan Nazir, Dr. Brian Ingalls, Aaron Yip	Conceptualization Ideas
Sultan Nazir, Dr. Brian Ingalls, Aaron Yip	Methodology
Sultan Nazir	Software
-	Validation
Sultan Nazir	Formal analysis
Sultan Nazir, Aaron Yip	Investigation
Dr. Brian Ingalls	Resources
Sultan Nazir, Aaron Yip	Data Curation
Sultan Nazir	Writing - original draft preparation
Dr. Brian Ingalls, Dr. Chaitanya Athale	Writing - review and editing
Sultan Nazir	Visualization
Dr. Brian Ingalls	Supervision
Dr. Brian Ingalls, Aaron Yip	Project administration
Mitacs GRI, Department of Applied Mathematics (University of Waterloo)	Funding acquisition

This contributor syntax is based on the Journal of Cell Science CRediT Taxonomy.

Abstract

One of the main techniques used in Synthetic Ecology is to perform experiments on simple synthetic communities under controlled environmental conditions to isolate and hence infer the mechanisms underlying the community dynamics. We commonly use isogenic strains grown in agar medium of known composition. For non-motile strains that interact with each other through diffusible molecules in the environment, spatial structure plays an important role in predicting community dynamics. Continuum models have proven to be one of the best choices of spatially explicit models to study spatiotemporal dynamics at the scale of diffusing molecules. However, there have not been many efforts towards calibrating such models in order to gain statistically reliable predictions. In this thesis, I extend a continuum model of density-dependent diffusion by augmenting it with terms that capture growth-induced dispersal. Using a reproducible protocol of growing bacterial cultures on agarose pads for imaging under a microscope, I gather time-series data of an *E. coli* monoculture. Following a pipeline of dynamic model calibration, I estimate the model parameters and infer avenues for improving the model formulation and the experimental design. Furthermore, I present a multispecies decomposition of the model and apply it to compare model predictions against preliminary data from a co-culture of antibiotic-resistant strains displaying cross-protection mutualism.

Contents

List of Figures

List of Tables

List of Abbreviations

ODE	Ordinary Differential Equation
PDE	Partial Differential Equation
ABM	Agent-based model
RD	Reaction-Diffusion
EPS	Extracellular polymeric substance
MLE	Maximum Likelihood Estimate
SA	Simulated Annealing
PL	Profile Likelihood

Chapter 1

Background

1.1 Introdution

1.1.1 Synthetic Biology

It is not news that microbes have played an essential role in the history of humankind. They have been important players in crucial advancements in medicine, agriculture, and manufacturing. Synthetic biology emerged as a field of research aiming to investigate cellular mechanisms and synthesize tools to design "devices" that perform desired tasks. These devices are living organisms with modified gene expression patterns. We can achieve this by inducing changes at the transcriptional, translational, or post-translational level events. The resulting phenotypes may be successful in accomplishing tasks that benefit humanity. For example, they can result in environmental changes through bioremediation (Cases and Lorenzo 2005), production of sustainable sources of energy (Savage, Way, and Silver 2008), or treating life-threatening diseases like cancer (Anderson et al. 2006).

Some of this research has adopted systems-level approaches to create circuits of genes that can allow a more diverse set of functions with greater efficiency than single mutations. This has been called "the second wave of synthetic biology" (Purnick and Weiss 2009). However, it has posed multiple challenges. Researchers explore simple genetic recipes for engineering organisms. Despite being able to design and characterize the elements or modules of these circuits, living cells often do not work like parts put together in a device and produce unpredictable circuitry (Kwok 2010). A more recent approach is to rethink these devices with cell-level elements. In their paper, Kuhn et al. (2010) presents the idea of a genome-level characterization of microbial strains to synthesize "whole-cell biocatalysts". Despite not fully understanding the complexity within cells, systems biology offers tools to study metabolic networks where whole cells are used as the nodes and catalysts. However, designing such population-level bioprocesses is also met with multiple challenges.

Biological systems, in reality, are much more complex than a typical synthetic design can foresee. In most cases, these complexities are ignored, sometimes leading to unpre-

dictable circuitry at the population level. Hence there is a need for increasing biological realism in these designs. Studies have shown that microbial communities consisting of strains with desired functions generally will not always be capable of collectively performing the desired task. This may be due to unanticipated ecological and evolutionary processes at play, causing the instability of the consortium. Designs need to explicitly incorporate the interactions between populations and the environment, and the physical structures that contribute to the ecological and evolutionary impact on them (Escalante et al. 2015).

1.1.2 Synthetic Ecology

More recently, synthetic microbial ecology emerged as a field of research with two main objectives. The first is a shared goal with synthetic biology: to "enable, control and optimize a desired biotransformation". However, synthetic ecology deals with engineering microbial assemblages in contrast to individual genotypes in synthetic biology. The second objective, unique to synthetic ecology, is to investigate the ecological and evolutionary feedbacks that emerge by interactions of a microbial genotype with another and the environment within these assemblages (Dolinšek, Goldschmidt, and Johnson 2016). There are many methods in our knowledge and currently in development to achieve the first goal (Zomorrodi and Segrè 2016). Many of these biotransformation events happen in complex environments like the human body. It becomes increasingly difficult to predict the outcomes of these assemblages in such scenarios, and it depicts the importance of working towards the second objective. The typical first step is to perform controlled experiments with simpler communities (Dolinšek, Goldschmidt, and Johnson 2016). Such experiments contribute to a detailed characterization of ecological interactions, which may not be possible in a more complex setting even if these interactions are present. Quite often, inter-specific interactions alone will lead to instability of the community. For example, a fast-growing strain consuming a shared resource may competitively exclude a slow-growing strain. Positive interactions can also lead to instability, such as the phenomenon of "tragedy of the commons". Theoretical studies have shown that cooperative interactions are sensitive to the emergence of "cheaters", and successful co-existence is seen only for a narrow range of parameters (Leenheer, Schuster, and Smith 2019).

Although interactions alone may sometimes predict instability, we know from experiments that various physical properties can act as ways to stabilize these communities. Dolinšek, Goldschmidt, and Johnson (2016) present temporal heterogeneity and spatial structure as two such properties. Some experiments attempting to study this aspect of community dynamics try to reproduce naturally occurring spatial structures like stream mesocosms (Cardinale 2011), but a more commonly used apparatus for the study of spatiotemporal dynamics of microbes is the agar plate.

1.1.3 Expanding communities on agar plates

Agar was suggested to Robert Koch in 1881 as a solid media alternative to gelatin for isolating bacteria (Sandle 2010). Agar is transparent, resistant to bacterial enzymes, and could be used for inoculation at higher temperatures than gelatin. Years later, it is still widely used by microbiologists to culture bacteria. As a solid or semi-solid surface that can be prepared with a known nutrient composition, it is suitable to study the spatiotemporal dynamics of microbial communities on different kinds of interfaces: between solid, semi-solid, liquid, or air. When we begin growing cells in a small region of the agar plate, they eventually grow and exhaust the nutrients available in that locality. If the cells are motile, they may follow the nutrient gradient by chemotaxis and invade nearby regions. For non-motile cells, if the nutrient concentration is high enough, the locality will get crowded with cells that continue growing and exerting pressure on nearby cells. Due to this pressure, the cells will be forced to move outwards and invade nutrient-rich regions of the plate. Our discussions will be limited to non-motile microbes as the spatiotemporal dynamics of motile cells is much more complex and beyond the scope of this thesis.

Various range expansion experiments on agar surfaces revealed community dynamics that could not be predicted in a well-mixed setup. Momeni, Waite, and Shou (2013) found that self-organization-induced spatial heterogeneity promoted cooperation over cheating. Harcombe (2010) found the emergence of cooperation due to spatial structure allowing preferential growth of cooperators. Hallatschek et al. (2007) demonstrated that genetic drift at the edge of the expanding front of a colony could potentially act against natural selection and allow the persistence of non-beneficial mutations by segregation in space. Van Dyken et al. (2013) found that colony growth promotes cooperation at the frontier by a “survival of the fastest” mechanism. Müller et al. (2014) found an antagonistic interaction between mutualism, which promotes the mixing of strains, and genetic drift, which promotes the demixing of strains, in an expanding colony.

As agar pads proved to be an essential apparatus for studying various ecological processes, there came a need to study the spatiotemporal dynamics of colony growth itself, an underlying mechanism for some of the patterns observed in the studies above. For instance, a paper by Hol et al. (2015) aimed to look at the community dynamics for different “experimental implementations of spatial structure” by testing range expansion with different methods of preparation of agar. They concluded that spatial structure alone might not explain the resulting patterns because it may be specific to the experimental apparatus. Tronnolone et al. (2018), through a combination of experimental and modelling methods, explored limited nutrient availability as a cause of non-uniform colony growth in bacteria and yeast. Although diffusion-limited growth (DLG) was discovered to be a possible mechanism in bacteria, the finding in yeast was attributed to pseudohyphal growth instead.

1.2 Modelling spatiotemporal dynamics

It has been over three decades since the necessity of an explicit exploration of spatial effects has been recognized in theoretical ecology (Kareiva 1994). Traditionally, spatial effects have been considered a minor contingency that may be ignored. In the following years, studies developed mathematical models that included spatial effects. They were, in general, derivatives of non-spatial models that made implicit assumptions about spatial effects while ignoring local-scale interactions, or "Spatially Implicit Models (SIMs)" (DeAngelis and Yurek 2017). The reason was ease in deriving simpler or useful theoretical results from non-spatial models. Eventually, it was recognized that fine-scale interactions and the resulting feedbacks are fundamental mechanisms that can give insights into various ecological phenomena. Keeping this in mind, we also know that detailed models pose the problem of the propagation of prediction uncertainty as pointed out by Peters et al. (2004) in their review on non-spatial, spatially implicit, and spatially explicit models. From a comparative statistical analysis, they infer that the choice of spatial detail should be a careful decision that depends on the spatial scales of the phenomenon we are interested in.

We will first look at some possible mechanisms by which spatial structure can influence community dynamics in microbial assemblages compared to its well-mixed counterpart. Assume that species A interacts with species B through a diffusible molecule C. The following are different ways in which bioprocesses in species A can influence the abundance and distribution of species B:

- **A eliminates C**: Let us say that A can produce an enzyme that degrades C upon contact. Then, A acts as a sink for C in the environment. By diffusion, a gradient of concentration of C is created. If C is a toxin that decreases the growth rate of B at a certain concentration, then B grows faster when close to A and slower when far from A. B then accumulates in the region surrounding A (See Figure 1.1 for an illustrative description of the range on interaction due to such a mechanism). If C is a nutrient that increases the growth rate of B at a certain concentration, then B grows slower when close to A and faster when far from A. Then, there forms a region around A that lacks cells of B.

- **A produces C**: Let us say that A can secrete C into the environment. By diffusion, a gradient of concentration of C is created. If C is a toxin, there forms a region around A that lacks cells of B. If C a nutrient, B accumulates in the region surrounding A.

In both cases, if they are well-mixed systems, the concentration of C is brought to a constant in space. Depending on the effect of its concentration on the growth of B, the density of B either increases or decreases everywhere uniformly. It has been studied that the range of interaction is influenced by the effective diffusion rate (cell crowding can reduce the diffusion rate) of the molecule mediating the interaction, the rate of secretion (or elimination), and the rate of uptake (Dal Co et al. 2020). The interaction range can strongly influence the spatial organization of species, which is an important determining factor of the community function and dynamics.

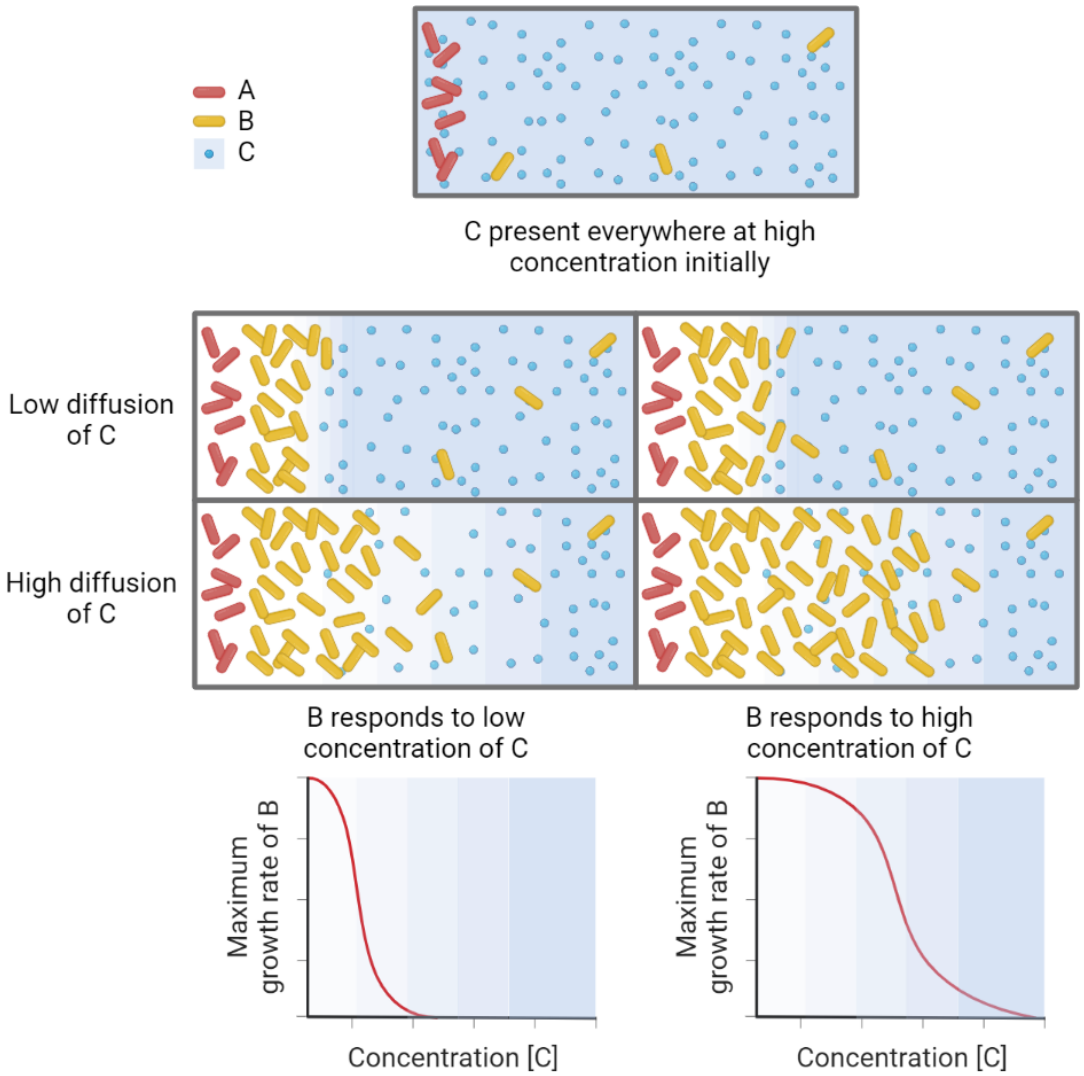


Figure 1.1: An illustration of how diffusion rate and uptake rate can together influence the range of interaction.

Here I have taken an example of one-way mutualism where species A benefits species B by eliminating molecules of C that lowers the growth rate of B. In all the cases, I assume that A eliminates C at a constant rate. Observe that the interaction range is smaller when the diffusion rate is smaller and B is highly sensitive to C (growth rate drops even at small concentrations of C). Meanwhile, the interaction range is larger when the diffusion rate is larger and B is only sensitive to high concentrations of C. **Created with BioRender.com**

The above cases include the fundamental mechanisms of most of the known interspecific interactions in non-motile microbes (Tan, Zuniga, and Zengler 2015), but there are exceptions. Some interactions may occur by direct contact between cells. For example, some bacteria are known to inject antibacterial proteins into other cells via secretion systems upon contact (Klein, Ahmad, and Whitney 2020). It must be noted that we have discussed a straightforward system of two species and one molecule. In reality, we are interested in much more complex systems, possibly with multiple species that exchange a large number of molecules. Although the underlying mechanism of each diffusion-mediated interaction can be summarized as above, the emergent properties become increasingly difficult to predict with each additional species or molecule. Therefore, we fail to learn the underlying mechanisms by simply observing the abundances and distributions of species, which are often the only observable quantities. We do not know the composition and gradients of molecules in the environment, which are often impossible to measure. This is where mathematical modelling becomes useful. Models offer theoretical insights into the relationship between the invisible components, such as the concentrations of molecules, and the observable quantities of the system, thereby augmenting the experimental observations (Tronnolone et al. 2018).

In the following section, I will review the different classes of mathematical models commonly used to study the spatiotemporal dynamics of microbial communities.

1.2.1 Classes of models for spatial study

Non-spatial models are typically modelled using Ordinary Differential Equations (ODEs). These models assume that the constituents of the community (the cells and the substrate) are well-mixed and hence interact by random collisions. This assumption helps in using mass action laws in describing the interactions. The changes in species biomass and substrate concentrations are modelled using Michaelis-Menten's equation for enzyme kinetics. Spatial processes such as immigration and emigration can also be incorporated in ODE models under simplistic assumptions (Wade et al. 2016). This section will discuss the two commonly used classes of spatially explicit models (SEMs). These are continuum models and agent-based models (ABMs).

Continuum models: These models are defined using Partial Differential Equations (PDEs). Here, changes in species densities and substrate concentrations are defined at a local scale smaller than the spatial scale of the system. The applicability comes from the fact that in the limit that the relative size of the local scale goes to zero, biomass and substrate behave like a continuous mass instead of discrete particles. For modelling microbial biomass, we assume that the size of this local scale is much larger than the size of a cell in order to avoid random fluctuations at smaller scales. PDEs define deterministic local dynamics defined using real numbers that indicate averaged population dynamics (Mattei et al. 2018). Stochasticity may also be explicitly modelled at these scales using Stochastic Partial Differential Equations (SPDEs). However, we will not go into the details of SPDEs. In summary, continuum approximation works when the scale of local interaction is larger than that of the cell but much smaller than that of the entire

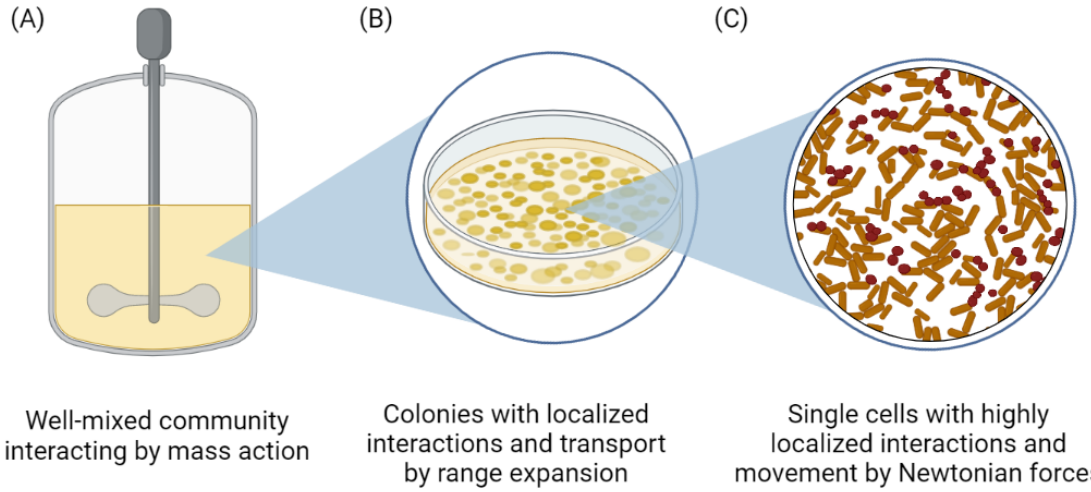


Figure 1.2: The choice of class of mathematical models largely depends on the spatial resolution of the system we are interested in.

(A) Large well-mixed communities are modelled using ODEs which does not require spatial structure. (B) Communities growing on surfaces only interact with species in a small neighbourhood. These often comprise a large number of cells and are studied using apparatus like agar plates. A continuum model is a suitable choice in this case. (C) We may be interested in community dynamics at a higher resolution where we distinguish individual cells. These systems comprise relatively small number of cells such that stochasticity and fine-scale laws of motion cannot be omitted. These systems may be studied using apparatus like microfluidic devices. An agent based model would be the suitable choice in this case. **Created with BioRender.com**

community.

The interactions between species and substrate at the local scale are modelled using the functions used in its well-mixed counterparts (ODEs). In other words, we assume that the system is well-mixed only locally. The major contrasting feature from ODE models is the presence of transport terms. These are additive terms in the rate equations for the local dynamics that define the rate of change of species biomass and substrate concentrations due to movement in space. This is usually modelled using spatial derivatives. Physically, this implies that the direction and magnitude of instantaneous movement depend on the gradient of densities or concentrations of some spatially explicit quantities. In general, if $X(x,t)$ is a finite dimensional vector where $X_i(x,t)$ is used to denote the biomass density of the i^{th} species or the substance concentration of a particular molecule in a very small locality around the point x at time t, then

$$\frac{\partial X_i}{\partial t} = \mu(X(x,t)) + \text{transport}$$

where $\mu(X(x,t))$ denotes the "reaction term" that comprises the growth or death of the species i (production or depletion for molecules) due to the abundance of species or compounds in the locality. Going by the well-mixed assumption at these small scales, we typically use Michaelis-Menten's equation to model uptake of substrate (Monod kinetics) and the law of mass action to model contact-dependent interactions between species.

Diffusion is the dispersal of suspended molecules due to random collisions that occur by virtue of the kinetic energy they possess. In this case, the transport term can be defined by the second order spatial derivative (or Laplacian) of its concentration. Therefore, the rate equation for molecules in the environment takes the form

$$\frac{\partial X_i}{\partial t} = \mu(X(x,t)) + D_i \nabla^2 X_i(x,t)$$

This is the general form of a commonly found group of continuum models known as reaction-diffusion (RD) models. The different approaches in modelling the transport term for biomass density is reviewed in Section 1.3.

Agent based models: These models explicitly define individual cells (or agents), allowing us to study spatial heterogeneity at a single-cell resolution. Although these models deal with high spatial resolutions, the assumptions used often depend on the extent of the biological reality one desires to capture. Some studies may have broader goals such that the exact physics at the level of a cell might be insignificant. For example, agent-based lattice models restrict the positions of cells to lattice points, and movement is restricted by the geometry of the lattice (Alamino 2020). Some other studies try to replicate realistic spatiotemporal dynamics of bacterial cells. Many modelling tools have been developed on this front. Simbiotics is one such tool that allows modelling multicellular systems in a 3-dimensional framework (Naylor et al. 2017). Based on the level of abstraction they wish to achieve, these ABMs have libraries of "submodels" that define various physical, chemical and biological processes. Although the cells are modelled as discrete entities, ABMs assume a continuous architecture for concentrations of molecules in the environment,

which are implemented using finite difference approximations as stand-ins for PDE models. In each discrete time step, the submodels are used to update the state variables of the cells, their positions and velocities, and the state variables of the environment. Some examples of such submodels are listed below:

- Uptake of nutrients from the cell neighbourhood using Monod kinetics.
- Motion of cells using Newtonian dynamics. If i is the index of the cell, the position and velocity of i are updated using the following ODEs:

$$\frac{d\vec{p}_i(t)}{dt} = \vec{v}_i(t)$$

$$\frac{d\vec{v}_i(t)}{dt} = \frac{\vec{F}_i(t)}{m_i}$$

  where $\vec{F}_i^T$ is the sum of all forces acting on i due to cell-cell collisions, viscous drag etc., and m_i is the mass of i.
- Biological processes such as cell growth as increase in the length of the cell, and cell division as a probabilistic decision that depends on its length.

Comparing the two approaches

1. **Computational complexity:** Although ABMs keep track of more information on community dynamics than continuum models can, they are computationally much more expensive than continuum models when we deal with large population sizes. Moreover, tracking individual cells at such scales becomes unnecessary to study population or community dynamics (Byrne and Drasdo 2009). Since ABMs require solving a system of ODEs for each cell, the computation time increases with every added cell. Continuum models may be solved analytically in some cases, but more often, they are simulated numerically. Numerical simulation of continuum models is easy to implement by discretizing space and time and rewriting derivatives by their finite difference approximations. The state of the system at every time step can be computed by explicit methods like the Forward Euler Method or implicit methods like the Backward Euler Method. Although explicit methods are computationally easier to implement, implicit methods ensure more stable discrete approximations. In either case, the computation time scales with the resolution of space and time rather than the population size as seen in ABMs.
2. **Relation to physical laws:** As described earlier, ABMs comprise highly parametrized physical laws at the single-cell scale. Laws of conservation of mass, momentum and energy are satisfied at the level of interaction between two cells. The resulting multitude of parameters are often difficult to estimate using data that is not resolved to such microscopic scales (Peters et al. 2004). Continuum models can implement

macroscopic laws from the field of continuum mechanics. For example, for a continuum of cell biomass $\rho(\vec{x}, t)$, the law of conservation of mass can be written as:

$$\frac{\partial \rho(\vec{x}, t)}{\partial t} + \nabla \cdot (\rho \vec{v}) = g(\rho)\rho$$

where $\vec{v}(\vec{x}, t)$ is the velocity and g is the rate of growth of cell density due to different biological processes.

3. **Access to numerical and analytical tools:** Differential equations have been widely studied by mathematicians. Describing the system using a system of PDEs allows us to access methods of numerical analysis and other analytical results in order to obtain useful inferences. ABMs on the other hand are often very specific implementations for which there may not exist useful mathematical results.

4. **Inferring mechanism:** Due to the multitude of complex intercellular interactions modelled in ABMs, it is a difficult task to deduce mechanisms of emergent patterns (Gorochowski 2016). In contrast, PDEs typically have only a handful of parameters that allow for an explicit parameter study of the spatiotemporal patterns.

For these reasons, we prefer using continuum models to study the spatiotemporal dynamics of microbial colonies. However, continuum models may pose a problem of mathematical complexity. Properties like stochasticity and anisotropy can be easily defined in ABMs using submodels as mentioned earlier. Although possible, incorporating such properties in a continuum model may not be a straightforward task. They can make the system mathematically very complex to work with and less accessible to the larger scientific community.

There has not been many efforts in bridging “microscopic” and “macroscopic” models. Recently, there has been a growing attention towards the need to derive large-scale continuum models from the microscopic rules used in ABMs (Lötstedt 2021). This is important in order to prove the validity of continuum models by checking for the retention of cell-level physical laws at the macroscopic scale. They will also help draw meaningful explanations for the parameters of continuum models in terms of cell-level parameters.

1.3 Modelling colony growth

Classically, transport of non-motile bacterial cells has been modelled as Fickian diffusion in the context of studying spatiotemporal patterns of cell populations (Murray 2003). This assumes that cells simply move down the gradients in cell densities (which as mentioned earlier, is not exactly the case). Further, if local cell dynamics is modelled using logistic growth, the resultant model is known as Kolmogorov–Petrovsky–Piskunov equation or Fisher-KPP equation (Kolmogorov, Petrovskii, and Piscounov 1937). It is mathematically written as

$$\frac{\partial \rho(\vec{x}, t)}{\partial t} = D\nabla^2 \rho + \lambda \rho(1 - \rho)$$

where ρ ia non-dimensionalized measure of cell density and λ is a scalar constant. Although there have been efforts to estimate a value for the diffusion constant for bacterial cells, it has been empirically studied that Fickian diffusion does not explain the mechanism of colony growth for the most part. In 1994, Barker and Grimson presented an extended Eden model which is a lattice-based model where colony growth is localized at the boundary and occurs through stochastic growth (Barker and Grimson 1994). This model helped explain roughness observed at the edge of bacterial colonies. Eden-like models are still widely used to study stochastic effects like "gene surfing" during range expansion of microbial communities (Hallatschek et al. 2007).

One of the earliest efforts at developing a continuum model for realistic colony growth was also by Grimson and Barker (1994). They claimed that there are two kinds of cell growth occurring within bacterial colonies: "local growth" in response to space available locally and "non-local growth" in response to difference in cell density with the neighbourhood. They claimed that expansion of bacterial colony is attributed to non-local growth at the edge of the colony where there is a steeper drop in cell density. Mathematically, the model is expressed by a RD equation with an additional term for non-local growth. If local growth is modelled as logistic growth, the PDE is written as

$$\frac{\partial \rho(\vec{x}, t)}{\partial t} = D\nabla^2\rho + \lambda\rho(1-\rho) + \Lambda(\nabla\rho)^2$$

where Λ is a scalar constant.

These continuum models failed to capture some experimental observations by Fujikawa and Matsushita (1991). They hypothesized that nutrient availability limited to the edge of the colony is the mechanistic cause of colony expansion. This was supported by observations of colonies growing towards higher nutrient concentrations, and repulsive growth of colonies growing close to each other. These findings indicated that nutrient diffusion is an important factor that can interact with cell growth, and was suspected to be a major driver for the speed of radial and vertical growth of a microbial colony. Diffusion limited growth (DLG) is used to denote colony expansion characterized by fast growing edge with a relatively slower rate of nutrient diffusion. It is known to affect the resultant colony morphology, for instance, by the formation of branches (Pipe and Grimson 2008). Mimura, Sakaguchi, and Matsushita (2000) defined a system of RD equations that explicitly modelled nutrient diffusion. They were able to capture the roughness of colony edges with an assumption of Fickian diffusion of cells. However, they incorporated a stochastic fluctuation in the diffusion coefficient to describe the branching phenomenon. They were able to conclude that DLG could explain the empirically observed diversity in colony morphology. Tronnolone et al. (2018) added to this with a parameter study to check for the conditions that allowed for directed growth of colonies in response to gradients in nutrient concentrations. They found that the occurrence of directed growth depended on the nutrient concentration and the diffusion constants.

In the last decade, there has been a shift in how we understand colony expansion. Warren et al. (2019) presented a 3-dimensional ABM of bacterial cells growing on a solid agar surface. They found that although the vertical growth of the colony was limited by nutrient availability, radial expansion was strongly determined by the mechanical forces

acting between the cells due to cell growth, and not nutrient diffusion. Continuum models have started looking at growth-induced pressure as a mechanistic cause of colony expansion. Cao et al. (2021) presented a "repulsion expansion model" that assumed a radially symmetric colony and looked at the dynamics of the radial distribution of cell density. The outward velocity of cells at every radial point in the colony depended on the integral of total cell growth rates from the center of the colony till the radius. Using this model, they were able to reproduce some empirically observed spatial patterns of gene expression in cell populations.

Farrell et al. (2013) developed a continuum model that too looked a the mechanical forces between cells as the driver for colony expansion. They used the Hertzian theory of elastic contact to model cell pressure caused by overlapping cells (overlap was modelled as an excess in cell density beyond a pre-defined upper bound for permissible cell density). If we assume that cells move down the gradient in pressure, we can use Darcy's law to relate pressure with velocity as

$$\vec{v} = -k_\mu \nabla p$$

where k_μ is a constant and p denotes pressure due to mechanical forces between cells. Substituting this in the equation for the conservation of mass in a continuum, we get the PDE relating the rate of change of cell density to cell pressure.

$$\frac{\partial \rho(\vec{x}, t)}{\partial t} - k_\mu \nabla \cdot (\rho \nabla p) = g(\rho)\rho \tag{1.1}$$

Lötstedt (2021) used discrete models and a detailed description of mechanical forces at the microscopic level to derive expressions for macroscopic pressure. Pressure can then be substituted in Equation 1.1 to obtain a continuum model. This was followed by a systematic comparison between predicted dynamics of the discrete models and the PDEs. Studies like these promise to relate macroscopic dynamics to microscopic parameters which gives us a detailed mechanistic view of colony expansion.

To summarize, with a blend of theoretical and experimental research, we have learned over the past few decades that the expansion of a colony of non-motile cells is attributed to the diffusion of nutrients and the mechanical forces between cells due to growth. A continuum model that explicitly incorporates nutrient diffusion and growth pressure is essential for marrying these findings and understanding colony growth better.

1.4 Objectives of study

This thesis consists three objectives. Firstly, I borrow a continuum model derived by Khassehkhan, Hillen, and Eberl (2009) that incorporates density-dependent diffusion used to study biofilms and re-derive it with modified assumptions to obtain a more mechanistically accurate growth-induced dispersal. This model can describe the spatiotemporal dynamics of non-motile cells growing on a surface (Discussed in Chapter 2). Given

the above discussion, such a predictive model is vital in avoiding expensive and time-consuming experiments and giving insights into the mechanistic basis for empirically observed phenomena.

Secondly, I take a step forward in the current approaches to model colony growth by demonstrating a quantitative pipeline to compare model predictions with experimental data. Model calibration is the systematic and iterative method of improving model ideation and estimating parameter values to fit the model predictions to observed data (Balsa-Canto, Alonso, and Banga 2010), which ensures the model predictions closely resemble reality. Each iteration gives us insights into the ability to estimate the model parameters by employing uncertainty analyses such as identifiability analysis (Discussed in Chapter 3 and 4). In doing so, we can identify redundancy in parameters (Kreutz et al. 2013) that helps in model selection and insufficiency in the experimental data (Villaverde et al. 2021; Atkinson and Donev 1992) that helps in better experimental design. I borrow an experimental protocol to obtain time-lapse data of single strain cultures (monocultures) of *E. coli* grown on agarose pads (Young et al. 2012). Following image processing, I demonstrate the use of the model calibration pipeline against the obtained microscopy data to estimate the parameters of the continuum model.

There have been few efforts toward calibrating PDE models. Only ABMs have been calibrated in the context of bacterial populations using high-resolution microscopy data (Yip et al. 2022). Calibration of ABMs becomes computationally expensive when dealing with populations the size of a colony or more. In the discussions accompanying model calibration, I highlight our challenges in achieving a fully calibrated PDE model and point out alternative experimental designs to enhance model calibration.

Thirdly, I present a multispecies version of the continuum model. This is achieved by a species decomposition method used by Rahman, Sudarsan, and Eberl (2015) to obtain a mixed-culture version of the biofilm model derived by Khassehkhan, Hillen, and Eberl (2009). Hence we obtain a generalized model for the spatiotemporal dynamics of a multispecies community of non-motile cells grown on a surface. I use a simple synthetic co-culture of two isogenic strains of *E. coli*, each with an inserted plasmid for resistance against a particular antibiotic. Therefore, they have a cross-protection mutualism interaction when grown on agar with both antibiotics. By qualitatively comparing the multispecies model predictions with timelapse images, I demonstrate that the presented model has the potential to predict the spatiotemporal dynamics of bacterial communities where species interact through diffusible molecules (Discussed in Chapter 5).

Chapter 2

Materials and Methods I: Continuum model

The model proposed in this thesis is an extension of the biofilm model developed by Eberl and colleagues over the past two decades (Khassehkhan, Hillen, and Eberl 2009; Eberl, Parker, and Van Loosdrecht 2001). In order to better understand the premise of this group of continuum models, I first present a brief overview of their biofilm model and then modify it to be better suited for explaining colony growth on a solid surface.

2.1 A continuum model of biofilm growth

They begin by defining a semi-discrete continuous time model inspired by Cellular Automata (CA) models of biofilm growth. This framework allows to define probabilistic rules of transitioning in space. The biological rationale that has inspired this model is that bacterial cells in a biofilm secrete and are embedded in extra-cellular polymeric substance (EPS) that immobilizes them. In the presence of nutrients, these cells divide and accumulate locally. When no space is available locally, the cells disperse into the neighbourhood and propagate the biofilm. A CA-like model permits the usage of phenomenological migration rules at a microscopic scale. Further, in the limit of fine grid refinement, we obtain PDEs that allow us to look at population-level dynamics.

Assumption 2.1.1: Cell movement in the biofilm model

In their model, Eberl, Parker, and Van Loosdrecht (2001) proposed the assumption that the movement of cells from a locality to the neighbourhood depends on the following two factors:

1. Availability of space in the neighbourhood to accommodate more cells.
2. Non-availability of space in the locality to accommodate growing cells.

Note that this assumption may be used for modelling any general population of non-

motile cells. In order to meet these conditions, they introduced two density-dependent functions, p and q, that satisfy the following properties.

Assumption 2.1.2: Properties of q and p

q and p are functions of local cell density ρ such that

1. they are bounded. Specifically, $q, p \in [0, 1]$
2. q is minimum and p is maximum in the absence of cells, i.e. $q(0) = 0$ and $p(0) = 1$.
3. q is monotonically increasing and p is monotonically decreasing, i.e. $\frac{dq}{d\rho} \geq 0$ and $\frac{dp}{d\rho} \leq 0$ everywhere.

Mechanistically, q (which depends on local cell density) defines the tendency of the cells to move outwards, and p (which depends on the neighboring cell density) defines the tendency for the neighbouring region to accept these cells. Using these functions, they derived the following continuum model to describe the rate of change of cell density:

$$\frac{\partial \rho}{\partial t} = \alpha_0 \nabla \cdot \left(D(\rho) \nabla \rho \right) + g(\rho)\rho \tag{2.1}$$

$$D(\rho) = pq + \rho \left(p \frac{dq}{d\rho} - q \frac{dp}{d\rho} \right) \tag{2.2}$$

where $g(\rho)$ is the per unit growth rate of cell density, and α_0 is a contant that determines the maximum dispersal rate of cells and has the dimensions of diffusion constant. This is a RD-like model with a density-dependent diffusion coefficient. Refer to Appendix A for the derivation.

The next step is to specify the functions q and p. To achieve this, they pose an "inverse problem" of deriving q and p using certain assumptions on $D(\rho)$ which can capture realistic biofilm growth dynamics. It must be noted that their aim is to build a 3-dimensional model that can explain biofilm propagation. This means that there needs to be a finite upper limit on the local density because cells cannot be packed beyond a certain density, and cells are immobilized in the EPS until the density is close to this limit. Keeping this in mind, they lay down certain constraints on cell diffusion as given in Assumption 2.1.3.

Assumption 2.1.3: Diffusion in a biofilm

If $D(\rho)$ denotes the density-dependent diffusion coefficient of cells and ρ_0 is the upper limit on cell density,

1. (Degeneracy) For $\rho << \rho_0$, diffusion coefficient is close to zero.
2. (Singularity) As ρ approaches ρ_0, $D(\rho)$ approaches infinity, i.e. very fast diffusion happens close to the maximum cell density.

However, we are not working in the realm of biofilms. Assumption 2.1.3 will not apply to the transport of cells in the absence of a material like EPS in the environment. More specifically, the degeneracy assumption at low cell density requires immobilization of cells. In regular colonies, cells can disperse by mechanical forces even at low densities. The singularity assumption at high density emerges due to working in three dimensions where there is a strong physical limit to cell density. However, in a two dimensional colony growing on agar surface, there is freedom to grow into the agar medium due to cells stacking on top of each other (which we will refer to as vertical growth of the colony). Since we do not explicitly consider this third dimension of growth, we will model this using a softer upper limit on cell density that is attributed to limited diffusion of nutrients in the vertical direction, i.e. perpendicular to the surface of the solid medium. Furthermore, this model ignores the effect of nutrient diffusion which may work for biofilms where only the external cells are exposed to nutrients due to limited diffusion in the EPS. Nutrient diffusion, however, is popularly hypothesized to be an important player in colony expansion as seen in the discussion in Section 1.3, and we will explicitly model this effect.

2.2 Repurposed model with growth-induced dispersal

I borrow the framework constructed in the previous section to derive a new continuum model that can realistically capture the spatiotemporal dynamics of bacterial growth on agar surface. As mentioned in Section 1.3, we have learned from previous modelling studies that mechanical forces between cells that emerge due to cell growth is presumably the major cause of colony expansion. However, it is still unclear if these forces primarily emerge from the middle or the edge of the colony. Nutrient diffusion is also hypothesized to drive colony expansion (Warren et al. 2019). Since nutrient availability is directly responsible for cell growth rate (assuming that it is the only limiting condition on growth), the diffusion rate of the nutrient could potentially indicate the region of the colony that exerts the most mechanical force for expansion. Clearly, we need a model that explicitly incorporates growth-induced dispersal of cells, and the interaction with nutrients diffusing in the environment.

In order to achieve such a model, I augment the assumptions on cell movement stated by the biofilm model with an additional statement.

Assumption 2.2.1: Cell movement in the growth-induced dispersal model

The movement of cells from a locality to the neighbourhood depends on the following factors:

1. Availability of space in the neighbourhood to accommodate more cells.
2. Non-availability of space in the locality to accommodate growing cells.
3. Local growth rate of cells which depends on the local availability of nutrients.

Before stating the model, I make one more assumption. In the biolfilm model, cell density was bounded above by incorporating very fast diffusion. Here we do not assume any property in the transport term that can limit the local accumulation of cells. I instead introduce a function $r(\rho)$ in the reaction term, and it has the following properties:

1. $r \in [0, 1]$
2. r is maximum in the absence of cells, i.e. $r(0) = 1$.
3. r is monotonically decreasing, i.e. $\frac{dr}{d\rho} \leq 0$ everywhere.

By multiplying this function with the growth term, there is a decrease in the net growth rate with increasing cell density. As mentioned earlier, this could be attributed to limited diffusion of nutrients in the vertical direction that can reduce the effective number of growing cells.

Under these assumptions, the spatiotemporal dynamics of non-motile cells growing on agar surface can be described by the following PDE:

$$\frac{\partial \rho}{\partial t} = \delta_0(p\nabla^2(Gq) - Gq\nabla^2 p) + Gr \tag{2.3}$$

where $G(\rho, S) = g(S)\rho$ and S is the nutrient concentration in agar. Note that δ_0 is different from α_0 defined in Equation 2.1. While α_0 set a constant maximum velocity of the flow of cells, in contrast, δ_0 does not have dimensions in time. The maximum velocity of the flow of cells here is determined by its product with the growth rate g. Therefore, the maximum velocity of cells is directly proportional to the local growth rate in the new model. Refer to Appendix A for more details and derivation.

Equations 2.1 and 2.2 can be rearranged to rewrite the biofilm model as:

$$\frac{\partial \rho}{\partial t} = \alpha_0(p\nabla^2(\rho q) - \rho q\nabla^2 p) + g(\rho)\rho$$

We can observe that the transport terms in the growth-induced dispersal model and the biofilm model are identical during exponential growth, which is presumably the case in the abundance of nutrients. Under these conditions, the transport terms differ only due to the different formulations of the functions q and p. It is also interesting to note that $q(\rho) = H(\rho)$ where H is the Heaviside function and $p(\rho) = 1$ (dispersal is independent

of neighboring cell density) are some extreme formulations satisfying the properties in Assumption 2.1.2. In this case, it is easy to check that we obtain Fickian Diffusion of cells.

The rate of change of nutrient concentration is determined by the following RD model:

$$\frac{\partial S}{\partial t} = \delta_s \nabla^2 S - \frac{Gr}{Y} \tag{2.4}$$

where Y is the yield of cell biomass per unit mass of the nutrient, and δ_s is the diffusion constant of the nutrient in agar.

The model is incomplete without defining the functions g, r, q and p. I define g using Monod's kinetics which is popularly use to model nutrient uptake. It uses a simplified assumption that the cell biomass is proportional to the amount of enzymes that convert the nutrient substrate into more cell biomass, which enables us to estimate nutrient uptake using enzyme kinetics.

$$g(S) = \mu \frac{S}{K_s + S}$$

where μ is the maximum rate of per unit growth of cell density. It is achieved when the nutrient concentration is high. K_s is the concentration of the nutrient at half-maximal rate of uptake per unit biomass.

r, q and p here are, in principle, the functions that give a macroscopic description of the mechanical forces acting at the level of cells. Ideally, we require a bottom-up approach in deriving them, and parametrizations utilizing physically meaningful parameters borrowed from cell-level physical laws. Although an important task in this line of work, it is beyond the scope of this thesis. Here I use Hill equations, which are widely used in Systems Biology, to define these functions. They agree with all the assumed properties of r, q and p: boundedness, monotonicity and differentiability. They also have very few parameters which makes them easier to work with in order to demonstrate model calibration in the next section. In all of the equations, I set the Hill coefficient to 1.

$$r(\rho) = \frac{R_{1/2}}{R_{1/2} + \rho}$$

r is the density-dependent function that implicitly limits the vertical growth of the colony. $R_{1/2}$ is the cell density at which local growth rate is halved. For large values of $R_{1/2}$, cells can continue to utilize locally available nutrients till exhaustion. At low values of $R_{1/2}$, accumulation of cells by growth slows down even in an abundance of nutrients. However, this does not affect the growth-induced migration of cells which can make space available locally for more growth.

$$q(\rho) = \frac{\rho}{Q_{1/2} + \rho}$$

q is the density-dependent function that confers a non-linear density dependence to growth pressure. A high growth rate alone might not exert enough pressure for migration unless there is a high cell density for translating this force. Growth pressure is equal to

Variable	Meaning	Dimensions
$\rho(x, y, t)$	Cell density	$[ML^{-3}]$
$S(x, y, t)$	Nutrient concentration	$[ML^{-3}]$
Parameter	**Meaning**	**Dimensions**
δ_0	Transport term coefficient of cells	$[L^2]$
δ_s	Diffusion coefficient of nutrient	$[L^2T]$
μ	Intrinsic growth rate	$[T^{-1}]$
K_s	Nutrient concentration at half maximal growth	$[ML^{-3}]$
Y	Yield of cell biomass per unit mass of nutrient	dimensionless
$R_{1/2}$	Cell density at half maximal local growth	dimensionless
$Q_{1/2}$	Cell density at half maximal growth pressure	dimensionless
$P_{1/2}$	Neighbouring cell density at half maximal cell flow rate	dimensionless
Function	**Meaning**	**Dimensions**
$g(S)$	Maximum cell growth rate at nutrient concentration S	$[T^{-1}]$
$r(\rho)$	Reduction factor in vertical growth rate at cell density ρ	dimensionless
$q(\rho)$	Inability to accommodate growth at cell density ρ	dimensionless
$p(\rho)$	Inability to accommodate inflow of cells at cell density ρ	dimensionless

Table 2.1: List of state variables, parameters and functions in the growth-induced dispersal model of colony growth.

the growth rate multiplied by q. For a fixed growth rate, $Q_{1/2}$ is the cell density at half maximal growth pressure. In other words, q determines how well cells can be packed locally before exerting enough pressure to move outwards. For instance, if cells stack on top of each other easily, we may have a high value of $Q_{1/2}$, but if cells in a monolayer would push cells out of the locality much before starting to stack up, we may have a low value of $Q_{1/2}$.

$$p(\rho) = \frac{P_{1/2}}{P_{1/2} + \rho}$$

p is the density-dependent function that describes the tendency for cells in the neighbourhood to oppose the movement of cells. For a fixed growth pressure, $P_{1/2}$ is the cell density in the neighbourhood at half maximal rate of flow of cells.

To observe the predicted dynamics, this model is numerically simulated using the

Forward Euler method and the finite difference approximations of the derivatives. For this project, I work within the spatial and temporal resolution and ranges of parameter values that permit stability with the Forward Euler method. Implicit methods like the Backward Euler method are recommended for a more exhaustive exploration of the model. I also assume that the boundaries of the 2-dimensional field are reflecting for both biomass and nutrients. This allows us to use the Neumann boundary condition where the derivatives at the boundaries (components of the gradients normal to the boundaries) are set to zero.

2.3 A multispecies decomposition of the model

As discussed in Chapter 1, we are typically interested in models that can be applied to co-cultures of more than one species of microbes to study the spatiotemporal dynamics resulting from their interactions. Experiments in this regard are commonly conducted on "isogenic mutants" (Bengtsson-Palme 2020). These are strains derived from the same parent with specific genetic alterations. They allow us to conduct controlled experiments where the emergent dynamics can be attributed to interactions via the modified genes within these strains and not due to interactions that may appear due to genetic differences that are unaccounted for. In mathematical terms, to model such a system, we may assume that the strains differ only in their roles in biochemical processes but are identical in terms of mechanical effects. In such a case, in a RD framework, we assume that the differences between strains appear only in their corresponding reaction terms and have identical functions in the diffusion terms.

Here I take the generalized continuum model for colony growth expressed by Equations 2.3 and 2.4, and linearly decompose it into PDEs for individual strains. In other words, I claim that Equation 2.3 describes the rate of change of total cell biomass of comprising all the constituent species, and we express the rate of change of cell biomass of individual species as PDEs that sum up to Equation 2.3. Henceforth, let k indicate an index given to a particular strain, where $k \in 1, 2, 3...N$ for a community of N strains. Let $\rho_k(x, y, t)$ denote the cell density of strain k and the total cell density be given by $\rho = \sum_{k=1}^{N} \rho_k$. Then, the rate of change of cell density of strain k is can be given by

$$\frac{\partial \rho_k}{\partial t} = \delta_0 \left(p \nabla^2 \left(\frac{G q \rho_k}{\rho} \right) - \frac{G q \rho_k}{\rho} \nabla^2 p \right) + G_k r \tag{2.5}$$

where G_k is the growth rate of strain k and $G = \sum_{k=1}^{N} G_k$. Refer to Appendix B for the derivation.

Chapter 3

Materials and Methods II: Model calibration

3.1 Data collection

In this section, I will explain the protocol followed to obtain the time-lapse images of the cell density distribution, which has been used to calibrate the continuum model.

3.1.1 Preparation of monoculture on an agarose pad

The size of a typical bacterial colony ranges between a few millimetres to a few centimetres in diameter. To obtain clear images of a single expanding colony, I use the help of an inverted microscope equipped with a 10x objective (see Section 3.1.2 for details) which requires me to localize the growing monoculture as close to the objective as possible. For this, I inoculate a spot of growing bacteria in the middle of an agarose pad (a thin slice of solidified agarose, as opposed to agar plates commonly used to grow colonies). The procedure is an application of the method of agarose pad preparation presented by Young et al. (2012). I use antibiotic resistance as the selectable marker in the medium for growing the monoculture. The exact procedure I followed is described below and illustrated in Figure 3.1.

1. Slowly heat and melt 150 mg of low-melt agarose in 10.0 mL of 0.4% M9 Glucose medium in a falcon tube. Vortex to ensure that the agarose is mixed uniformly. Allow the agarose to cool down to a temperature that is not cold enough to begin solidifying but not hot enough to denature antibiotics. The rule of thumb is to be able to touch the falcon tube for at least 10 seconds. At this temperature, add the antibiotic to which the strain of bacterium has resistance. Vortex again to ensure uniform distribution of the antibiotic. Here, I added 10.0 μL of Kanamycin (at a stock concentration of 50 mg/mL).

2. Place a cover slip (22-mm^2) on Parafilm spread on a flat surface. Pipette 1.0 mL of the agarose solution onto it under sterile conditions.

3. Immediately, vertically drop another coverslip of identical dimensions onto the agarose solution. This was suggested by Young et al. (2012) as a method to ensure the formation of a flat and even agarose pad due to the pressure applied by the coverslip on top and the surface tension at the edges of the coverslips that keeps the agarose solution intact and sandwiched between them.

4. Allow the agarose to solidify between the coverslips for about an hour at room temperature. Afterward, carefully slide a coverslip to expose one of the flat slides of the agarose pad.

5. Pipette 1 μL of bacteria suspended in liquid 0.4% M9 Glucose (with the antibiotic selector) during the exponential growth stage. The strain I used is *E. coli* MG1655 with an inserted plasmid (pEB1-mEYFP) which has a Kanamycin resistance gene. The growth stage is confirmed using OD_{600} readings of the suspension in a spectrophotometer. Here, the cells are harvested at an optical density of OD_{600} = 0.5.

6. Allow the spot to dry for 15 minutes so the bacteria are no longer freely suspended in the droplet. Meanwhile, using a sterile scalpel, cut out the edges of the pad surrounding the spot to get a smaller pad that comfortably fits into the glass bottom of a cover slip dish. One can fit two pads of roughly 1-cm^2 size on the glass bottom. I have used the MatTek 50 mm Dish (No. 1.5 Coverslip, 30 mm Glass Diameter, Uncoated).

7. Gently flip the pad onto the coverslip dish so the cells face the glass bottom. Ensure the pad is flat against the glass bottom without any air bubbles between them.

8. The monoculture is then observed on an inverted microscope.

If this procedure is followed correctly, one should see a roughly circular spot of cells on the same focal plane. We will also observe a dense ring of cells along the edge of the spot. This is likely formed due to the "coffee ring effect", according to which colloidal particles settle close to the edge of a droplet during evaporation (Deegan et al. 1997). Young et al. (2012) has presented a troubleshooting table listing some possible problems (and corresponding solutions) that one may encounter in following this protocol.

3.1.2 Time-lapse microscopy

I used the ZEISS Axio Observer, an inverted microscope with automated components, to image the colony. Since we are interested in the colony-level dynamics, I used an objective lens of 10x magnification (EC Plan-Neofluar 10x/0.30 M27). A microscope stage incubator was placed to ensure the cells grew at 37^0C. The microscope was controlled using the Zen PRO software.

Using Zen PRO, I first searched for the cells on the pad and manually focused on them using coarse adjustment followed by fine adjustment. After this, I fix the following settings:

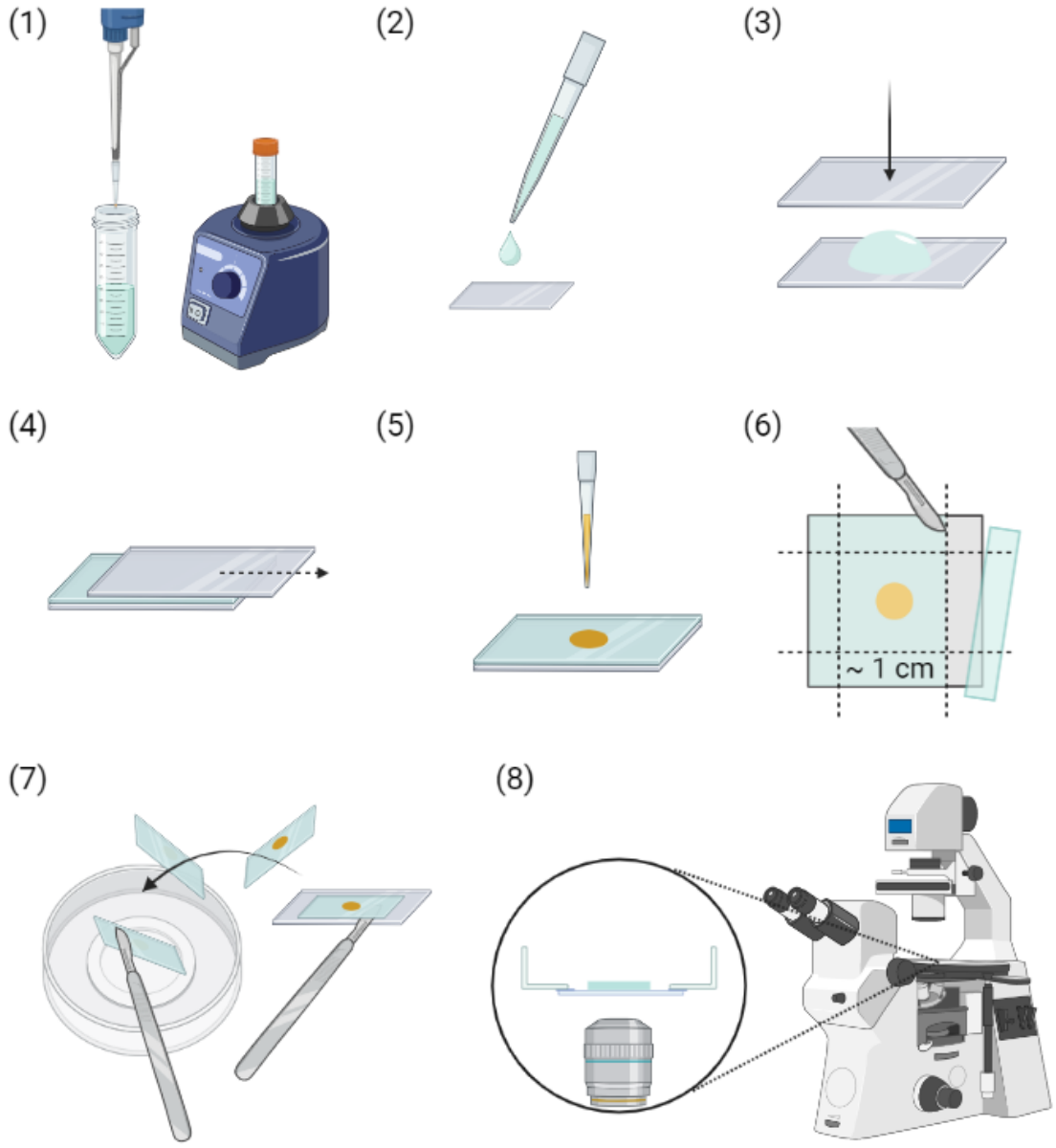


Figure 3.1: Protocol for the preparation of a monoculture on an agarose pad in order to capture time-lapse images on an inverted microscope.

The numbering of each step corresponds to the steps described in Section 3.1.1.
Created with BioRender.com

- An outline of the rectangular field that needs to be imaged at every time point: This requires stitching multiple images together for a wide field. Here, I obtained a field of dimensions 3.7 mm x 4.6 mm with a roughly circular spot of cells in the middle with a diameter of approximately 2.3 mm and an image resolution of 0.908 μm per pixel.

- Auto-focus with a range of 12 μm in the z-axis for the lens to parse during each shot.

- At every time-point, a frame is captured using light from each specified channel immediately after the other. The channels that I used were:

  Phase: White light is shined behind the cells, and the light scattered by the contents within the cells is amplified through the interference of this phase-shifted light with the background light. This is the principle of phase contrast microscopy (Martin 1947). A more significant fraction of light will be scattered if there are more phase shifts in the path of the light; therefore, a thicker layer of cells will produce a more extensive interference and, consequently, higher intensity in the Phase channel (See Figure 3.2 (A)). Lighting settings: 50% intensity from TL LED Lamp with 24 ms exposure time.

  Cy3: This channel emits a laser with an excitation wavelength of 548 nm and an emission wavelength of 561 nm. mEYFP is a green/yellow fluorescent protein synthesized by the *E. coli* strain. It has excitation and emission wavelengths that fall within the filtered ranges of wavelengths that are shined on the specimen and captured by the objective, respectively. By capturing this emitted light, the recorded intensity measures the localization and density of mEYFP in the absence of other excitable compounds that fall in the excitation range of the Cy3 channel (See Figure 3.2 (B)). Lighting settings: 50% intensity from X-Cite 120LED Lamp with 300 ms exposure time.

- Time-series settings were specified to capture an image at an interval of one hour for a total duration of 22 hours.

3.1.3 Image processing and spatiotemporal resolution

The first step of image processing is finding the "best fit" of intensity values. This is done automatically on Zen PRO where the intensities are normalized to baseline. By doing so, we eliminate any background fluorescence in the Cy3 channel and the bit of unscattered light in the Phase channel and maximize the contrast of images.

As mentioned in Section 1.4, one of the primary goals of this study is to compare the model predictions with experimental observations quantitatively. In order to accomplish this, I processed the raw data to obtain resolutions identical to the numerical simulations.

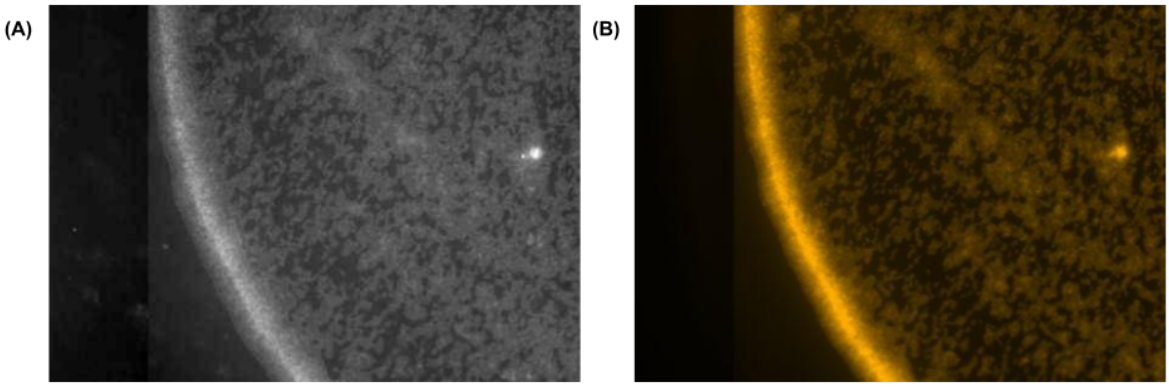


Figure 3.2: Phase and Fluorescence channel

(A) Image taken using the Phase channel (image brightness corrected for better visualization) shows higher intensity in regions with a thicker layer of cells. (B) Image taken in the Cy3 channel shows higher intensity in regions with higher density of mEYFP.

- **Temporal resolution:** The resolution in time is limited by the experimental data. The microscope takes a few minutes to capture and stitch images in all the channels. The time interval between images is set to one hour to prevent any effects of overexposure of cells to irradiation of light. However, the numerical simulation of the model can achieve a temporal resolution of the order of a few seconds. In a RD model, the stability criterion for the Forward Euler implementation demands a small time interval (where the upper bound is determined by the fastest diffusion rate in the model) Jeong (2018). Although we will simulate small time intervals, we will extract the spatial data every hour of simulation time. For all the simulations shown in this thesis, the time interval is set to Δt = 1s, and spatial data is extracted at every 3,600$^{\text{th}}$ time-step (1 hour of simulation time) for model calibration.
- **Spatial resolution:** Using the 10x objective, we obtain images of high spatial resolution of a colony. As specified in Section 3.1.2, the length of each pixel at this resolution is 0.908 μm, which is smaller than the typical length of an *E. coli* cell. However, we do not require a cell-level description of changes to observe continuum dynamics. By the principle of continuum mechanics, in a discrete space approximation used for the Forward Euler implementation of the PDE model, we require each discrete point in space to describe averaged local population dynamics. In order to achieve this, I use an averaging algorithm (See Figure 3.3) to resolve the images to pixels of size 45.4 μm. To match with experimental data, the distance between discrete points in space used in the simulations is also set to Δx = 45.4 μm.

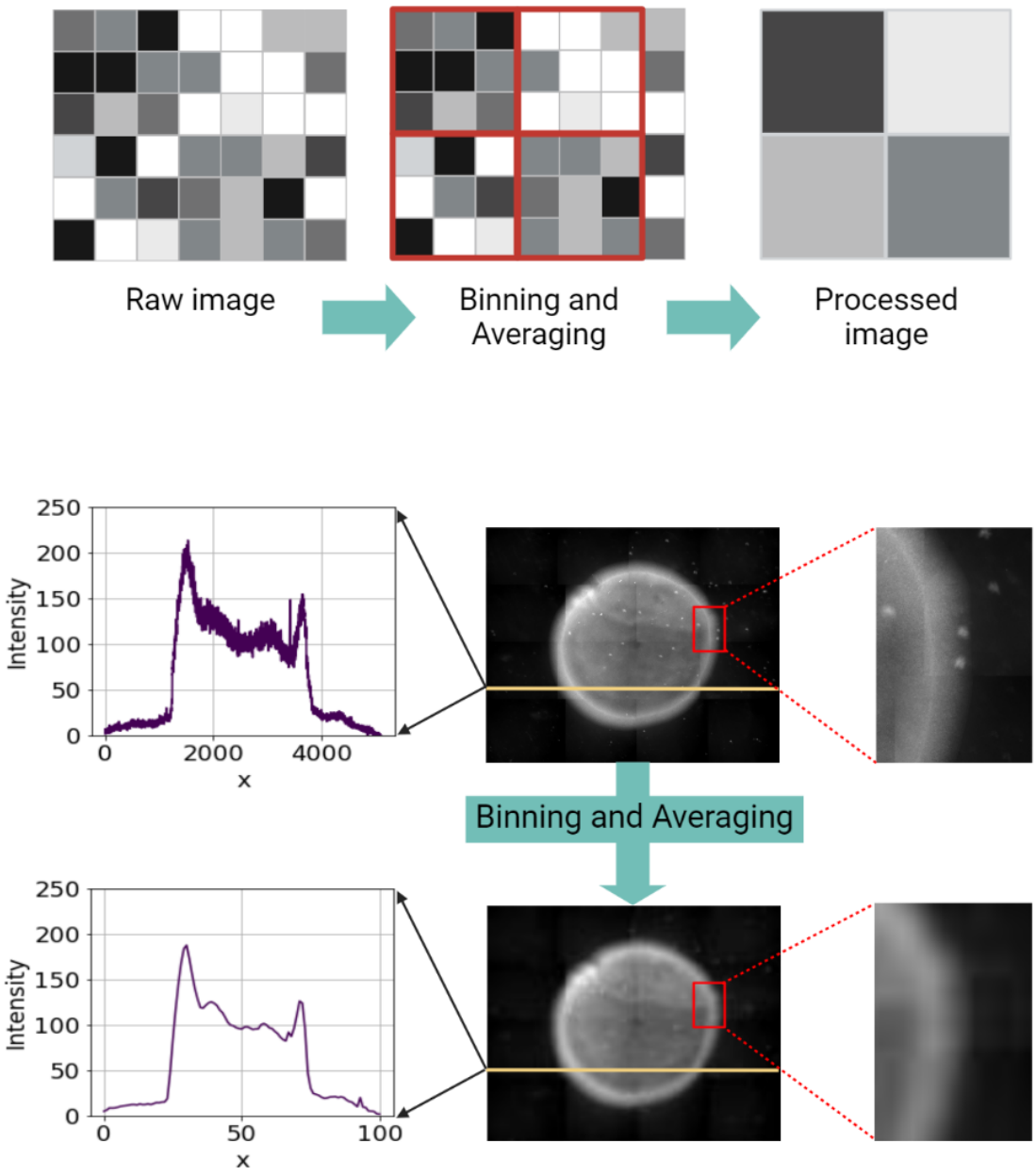


Figure 3.3: A binning and averaging method of image processing to reduce the spatial resolution of images.

(Top) A diagrammatic representation of the algorithm: The pixels of the image are binned into square grids of equal size. Each of these squares is assigned the average intensity value of the collection of pixels it contains in order to obtain an image of lower resolution. (Bottom) Application of the algorithm is shown for a phase contrast image of a bacterial monoculture. The plot on the left-hand side shows the intensity values along the yellow line. The image on the right-hand side is a zoomed-in section of the image. Observe that the discontinuities are diminished in the processed image.

3.2 Model calibration

Model calibration is a multi-step process that may often be iterated before achieving a fully reliable predictive model (Balsa-Canto, Alonso, and Banga 2010). A well-studied set of tools and supporting theories for calibrating linear models exists. These methods (such as least-squares estimation) are popularly used for qualitatively summarizing more complex non-linear models, but they are inaccurate in drawing statistical inferences (Myung 2003). Villaverde et al. (2021) has presented a detailed pipeline for model calibration in the context of ODEs. The steps described in the process are not specific to ODEs and generally applicable to PDE models. However, the implementations of these methods are still largely unavailable and yet to be characterized just as well for PDEs. This can be attributed to the non-availability of statistical results for dealing with multiple variables.

In this project, I aim to calibrate the PDE model presented in Chapter 2 using the microscopy data obtained using the apparatus described earlier in this chapter. In this section, I will follow the protocol of dynamic model calibration proposed by Villaverde et al. (2021) and describe the mode of application of each step for the model and data we have. Due to the lack of standardized methods for PDEs and the lack of replicates in the data obtained, which limits our ability to estimate error accurately, I will rely on some ad hoc results using stand-in estimates of error during the process.

3.2.1 Objective function

In order to match model predictions with experimental data, we need some means of comparing both. Qualitatively, we may do this by eye by comparing relative sizes or colours of observable quantities (or graphical representations of these quantities such that differences are easily noticeable). However, a qualitative comparison may not capture the nuances in differences, can be tedious when dealing with more than a few degrees of freedom, and is prone to biases. Instead, we quantify the deviation of the model prediction from the data (or "goodness of fit"). We can then use computers to rapidly compare this quantity for different parameter combinations and automate the minimization of the deviation (maximizing the goodness of fit). The objective function uses features of the experimental data utilizing an error model to return a quantity that summarizes the deviation of model prediction from the experimental data for a particular parameter vector.

The likelihood function is a commonly used objective function that tells us the probability of obtaining the observed experimental data provided that we know the values of the parameters (Myung 2003). The goal is to obtain the parameter vector that maximizes this probability, a process known as Maximum Likelihood Estimation. The corresponding parameter value is the Maximum Likelihood Estimate (henceforth referred to as MLE). During the process, the likelihood function is often log-transformed, which is easier to compute and converges faster (Hass et al. 2019). The resulting function is called the log-likelihood.

In the context of an ODE model, let $\mathcal{D}$ denote the data such that the value of the i^{th} observable at time t_k is denoted by $\mathcal{D}_i(t_k)$. Let the predicted value of the i^{th} observable at time t_k for a parameter vector θ be denoted by $\mathcal{X}_i(t_k, \theta)$. It is a common practice to assume that the deviation of the experimental observation from the true value at each time-point is given by an independent normally distributed random variable with mean zero and standard deviation $\sigma_i(t_k)$. In mathematical terms, we assume that

$$\mathcal{D}_i(t_k) = \mathcal{X}_i(t_k, \hat{\theta}) + \epsilon_i(t_k), \quad \epsilon_i(t_k) \sim \mathcal{N}(0, \sigma_i(t_k))$$

where $\theta = \hat{\theta}$ is the "true" parameter vector at which the model captures the system behaviour accurately. In such a case, the objective function is a weighted least sqaures function given by

$$\mathcal{O}(\mathcal{D}|\theta) = \sum_i \sum_k \frac{(\mathcal{D}_i(t_k) - \mathcal{X}_i(t_k, \theta))^2}{\sigma_i^2(t_k)}$$

and the MLE of the parameter is obtained using the relation

$$\mathcal{O}(\mathcal{D}|\hat{\theta}) = \min_{\theta} \mathcal{O}(\mathcal{D}|\theta)$$

If it is possible to obtain a dataset where data is taken for the same points in space over all time points, we can treat each spatial point as an observable. In this case, under the assumption of independent normally distributed error, we can rewrite the objective function as

$$\mathcal{O}(\mathcal{D}|\theta) = \sum_i \sum_j \sum_k \frac{(\mathcal{D}_i(x_j, t_k) - \mathcal{X}_i(x_j, t_k, \theta))^2}{\sigma_i^2(x_j, t_k)} \tag{3.1}$$

where i runs over all observables, j runs over all spatial points, and k runs over all time points. Refer to Appendix C for the derivation.

Generally, the assumption of independent normally distributed error may not always apply. If it does, it may be tedious to estimate the standard deviation due to the need for more replicates.

Error Model

- **Correlated error:** One needs to be careful about correlations between measurement deviations before using independent distributions to model them (Villaverde et al. 2021). For instance, Equation 3.1 assumes independent error at every point in space and time when dealing with spatiotemporal data. However, there may be sources of error that vary independently across spatial points but add constant deviations to observables at every time point at a particular spatial point. Such deviations are correlated in time at every spatial point. In the Phase channel time-lapse images obtained for the monoculture, we observe scattering due to the non-uniform composition of the agarose, scratches on the agarose surface during preparation, and disturbances on the surface of the glass bottom due to repeated usage. These

sources of scattering stay fixed in space over time and are, therefore, correlated in time. To accommodate this observation, I will relax the assumption of independently distributed error over time and introduce a source of correlation in an ad-hoc objective function in the following text. In reality, the unwanted scattering also has a certain degree of correlation in space as the corresponding pixels can be seen to be aggregated. However, for simplicity, I will assume that the error is independent of spatial coordinates.

- **Proxy for standard deviation:** In this project, I use a single time-lapse of the monoculture for calibration; hence, there is a severe lack of replicates. Conventionally, replicates are used to estimate the standard deviation of error. We can expect the standard deviation of the error to depend on factors like the absolute cell density, the phase of growth of cells, and the position relative to the colony. Therefore, assuming a fixed standard deviation could potentially overestimate or underestimate the confidence in data at certain points in space and time. There are two suggested alternatives for these circumstances (Hass et al. 2019). (1) One can introduce standard deviations as model parameters. Since we cannot assume a constant standard deviation, we would have a large number of model parameters, making the calibration process tedious and unreliable. (2) One can formulate a dependence of standard deviation on model parameters or state variables. This method would be appropriate in capturing error due to stochastic growth where we can expect spatial and temporal auto-correlation. However, since we are using the error model to estimate the parameters, introducing such a dependence could introduce biases in the calibration process. Therefore, due to the lack of a better alternative, I use the standard deviation of all intensity values averaged in the algorithm shown in Figure 3.3 as a proxy for the standard deviation in the intensity at that point in space and time. The following text denotes this value by $\sigma(x, y, t)$ for spacial coordinates (x, y) and time t.

Using the error model mentioned above, I present Assumptions 3.1 to derive an objective function. For the reasons which are also described above, these assumptions are far from reality. Nevertheless, this is an early step towards introducing correlated error in the context of calibrating against biological data (Villaverde et al. 2021).

Assumption 3.2.1: Error model

Let the cell density, measured as the phase contrast intensity, be the only observable (or output channel). Assume the following ideation of the sources of error in the observed cell density.

1. There is independent normally distributed additive error due to light scattered by disturbances in the agarose gel and on the cover slip at every spatial point. The deviation from true value due to this error is conserved across time-points at every spatial point.
2. In addition, there is independent normally distributed additive error at every point in space and time that can capture all the other sources of error.

Partition of pixels and quantification of phase contrast intensity from non-cell sources

To quantify the distribution of phase contrast intensity due to scattered light from unwanted disturbances in the field, I use the fluorescence channel image at a particular time-point as a mask to separate pixels with cells from pixels without cells in the phase contrast image at the same time-point. The threshold of fluorescence intensity to create a mask is determined, as a rule of thumb, to be 10% of the maximum fluorescence intensity in that frame to account for the effect of dispersion of the emitted light to the neighbouring cell-free pixels.

In mathematical terms, we have essentially split the data into sets of intensity measures from pixels with cells and pixels without cells. If Ω denotes the set of all the points (pixels) in space and time available in the time-lapse data, i.e.,

$$\Omega = \{1, 2, ...L_x\} \times \{1, 2, ...L_y\} \times \{1, 2, ...t_f\}$$

where L_x and L_y are the number of pixels along the X and Y axes of the field, and t_f is the final time-point.

The set of all pixels with cells is defined as

$$\Omega_{\mathsf{C}} = \{(x, y, t) \quad | \mathcal{I}_{\mathsf{Cy3}}(x, y, t) \geq 0.1 \max_{(x,y)} \mathcal{I}_{\mathsf{Cy3}}(x, y, t)\}$$

and the set of all pixels without cells is defined as

$$\Omega_{\mathsf{NC}} = \Omega \setminus \Omega_{\mathsf{C}}$$

The phase contrast intensity values in the pixels in Ω_{NC} corresponding to the initial time-point are used to estimate the mean and standard deviation of the distribution of phase contrast intensity that can be attributed to unwanted disturbances. The initial time-point is chosen because cell density is the lowest and cells cover fewer pixels, which allows us to clearly observe the background scattering. See Figure 3.4 for a summary

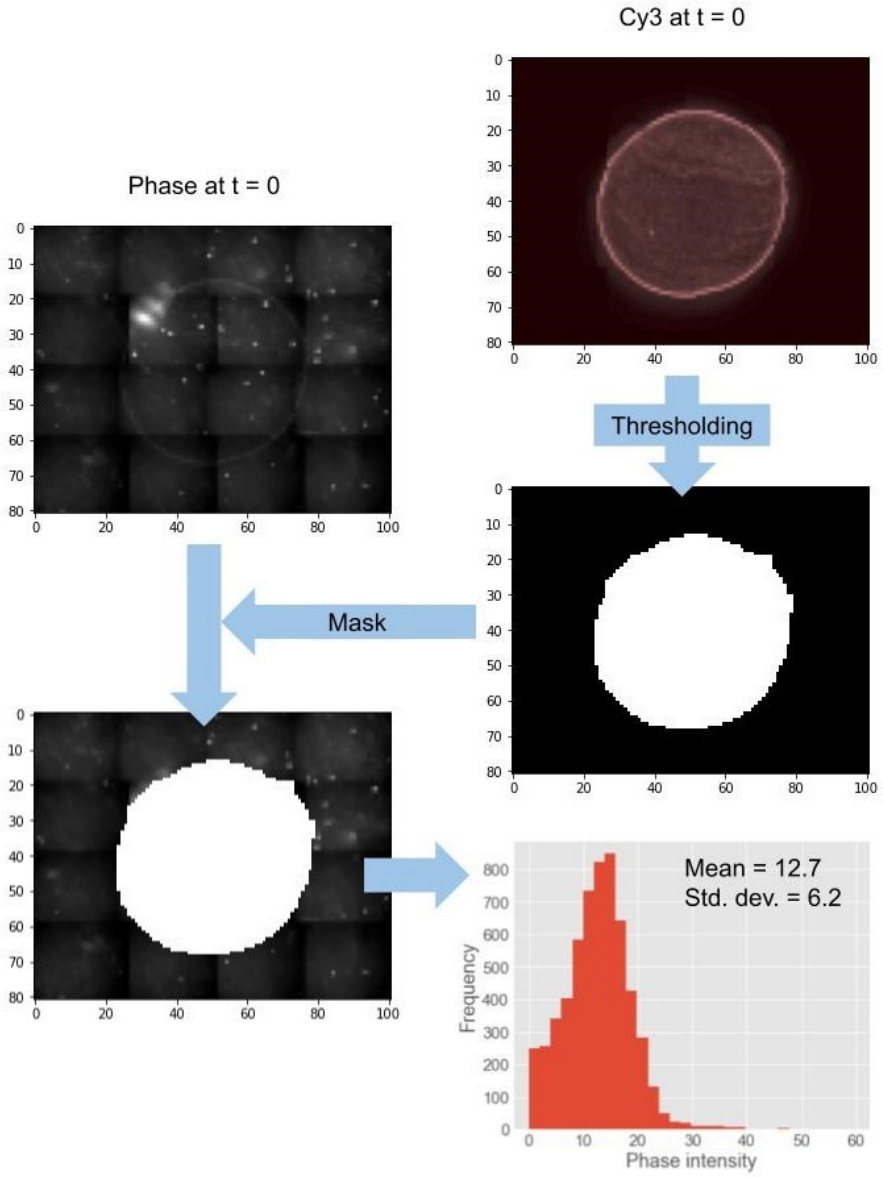


Figure 3.4: Quantification of phase contrast intensity from non-cell sources.
(Brightness of the images shown in the figure has been adjusted for better visualization)

of this process and the obtained statistics. It can be observed that the distribution of intensities due to background scattering is better explained by a log-normal distribution, but I assume a normal distribution to derive a simpler mathematical result for the likelihood function (See Appendix C). Let μ_{BG} and σ_{BG} be the mean and standard deviation respectively.

Before calibrating the model, let us subtract all the intensity values in the dataset by the number μ_{BG}. We have the error distribution centered around zero now, and let $\mathcal{I}$ denote the dataset of the resulting Phase channel intensities. Note that this is different from the background subtraction performed during Image processing, where, in principle, the intensities due to unscattered light is removed from the data. Here, we are subtracting the average intensity due to light scattered by disturbances that are not cells.

Furthermore, we use only the pixels in Ω_{C} to compute the objective function as the remaining pixels contain no information of the cell density. Consider the following definition:

$$\Omega_{\text{C}}(a,b) = \{(x,y,t) \in \Omega_{\text{C}} \quad |(x,y) = (a,b)\}$$

Then, using the assumptions of the error model and the partition of pixels, we obtain the following objective function:

$$\mathcal{O}(\mathcal{I}|\theta) = \sum_{x=1}^{L_x}\sum_{y=1}^{L_y}\left[\sum_{\omega\in\Omega_{\text{C}}(x,y)}\left(\frac{(\mathcal{I}(\omega)-\rho(\omega,\theta))^2}{\sigma^2(\omega)}\right) - \frac{\mu_{\text{err}}^2(x,y,\theta)}{\sigma_{\text{err}}^2(x,y)}\right] \tag{3.2}$$

$$\sigma_{\text{err}}(x,y) = \left(\frac{1}{\sigma_{\text{BG}}^2} + \sum_{\omega\in\Omega_{\text{C}}(x,y)}\frac{1}{\sigma^2(\omega)}\right)^{-1/2} \tag{3.3}$$

$$\mu_{\text{err}}(x,y,\theta) = \sigma_{\text{err}}^2(x,y)\sum_{\omega\in\Omega_{\text{C}}(x,y)}\left(\frac{\mathcal{I}(\omega)-\rho(\omega,\theta)}{\sigma^2(\omega)}\right) \tag{3.4}$$

where $\rho(\omega,\theta)$ is the intensity predicted by the model at pixel ω for the parameter vector θ. Observe that in the limit that σ_{BG} tends to zero (i.e., there is no temporally correlated error), we retain Equation 3.1 (sum of squares). Refer to Appendix C for the derivation of Equations 3.1-3.4.

Note 3.2.1: Initialization of simulation

I compare the data with the simulation for every parameter vector such that the initial cell density is identical to the initial cell density recorded in the time-lapse. However, in the Phase contrast data, the initial cell density distribution is convoluted by undesired sources of scattering. If the simulations are initialized with this image, these disturbances will also be interpreted as cells, resulting in an erroneous prediction of dynamics. Therefore, I initialize the simulation using the initial image captured by the fluorescence channel, where all the light captured is attributed to light emitted from the cells.

- Fluorescence channel data is used only for initialization and not for comparison at subsequent time-points. This is because, during growth, the maturation of GFP-derived fluorescent proteins such as mEYFP, which is used here, has been known to be limited by oxygen availability (Drepper et al. 2010). Therefore, the intensity of emission fails to capture the cell density. However, at the initial time-point, it is safe to assume that nutrients like oxygen are uniformly distributed in space and therefore do not hamper the estimation of cell density using fluorescence intensity.
- We do not know the conversion factor for Fluorescence (Cy3) and Phase intensities. Therefore, I introduce a new parameter κ such that the initial cell density (measured as phase contrast intensity) is the Fluorescence intensity scaled by κ.

$$\rho(x, y, t = 0, \theta) = \kappa \cdot \mathcal{I}_{\text{Cy3}}(x, y, t = 0)$$

3.2.2 Parameter estimation

To obtain the MLE of the parameter vector, we need to minimize the objective function (Equation 3.2). Parameter optimization is the process of estimating the parameter values by minimizing the objective function. For a large number of parameters, an exhaustive exploration of the parameter space is not feasible due to the computation time required. Non-linear models typically produce more than one minimum in the objective function landscape. The lowest minimum in the parameter space is called the global minimum, and every other minimum is called a local minimum. Hence, we require optimization algorithms that search a much smaller subset of the parameter space and can fetch the parameter estimate corresponding to the global minimum of the objective function. Optimization methods such as gradient descent which are used to find minima for linear models, if applied to non-linear models, pose the threat of convergence at local minima. To this day, we do not have a fast and standardized method of finding the global minimum for a generalized non-linear model (Villaverde et al. 2021; Ashyraliyev et al. 2009). We rely on stochastic processes like random walks in the parameter space and multiple repetitions of the optimization process with randomized initializations to increase our confidence in the obtained global minimum. Global searches are typically followed by

Parameter	Meaning	Value used
T_0	Initial (and highest) temperature	5x10^8
d_0	Initial (and largest) upper bound on step size	1.0
d_{min}	Smallest upper bound on step size (Asymptote)	0.1
N_p	Number of parameters	9
α	Coefficient of exponential decay of temperature	0.9
β	Coefficient of exponential decay of step size	0.95
i_{max}	Maximum number of steps	1500

Table 3.1: List of simulated annealing parameters.

local searches to increase the precision of estimation.

In this project, I adopt a stochastic optimization method, Simulated Annealing (SA), introduced by Kirkpatrick (1984). SA involves a random walk in the parameter space while computing the objective function at every move. Accepting every move is a probabilistic decision that depends on the "trade-off" in the objective function value relative to an abstract notion of "temperature". The implementation of SA requires the following ingredients (Refer to Figure 3.5 for a detailed algorithm):

1. A parameterized model: We have a system of two PDEs (Equations 2.3 and 2.4) that consists of 9 parameters (Listed in Table 2.1, in addition to κ introduced in Note 3.2.1).

2. A random generator of moves in the parameter space: At each step of the annealing process, I make a move (sampled from a uniform distribution of fixed length) independently along every log-transformed parameter axis.

3. A quantitative objective function: The objective function defined in Equation 3.2 is computed for the newly generated parameter vector and compared to the preceding parameter vector. The move is accepted or declined using the Metropolis algorithm (Kirkpatrick 1984). According to the Metropolis algorithm, every move that lowers the objective function is accepted. Every move that increases the objective function is accepted with a probability defined by the Boltzmann function that depends on the temperature. At a very low temperature, there is almost no chance of accepting an "uphill" move, so the process becomes a local optimization method. At a high temperature, the search process can move uphill and escape local minima.

4. An annealing or cooling schedule and the duration of the process: Over time, I decrease the temperature used in the Metropolis algorithm, which allows for the

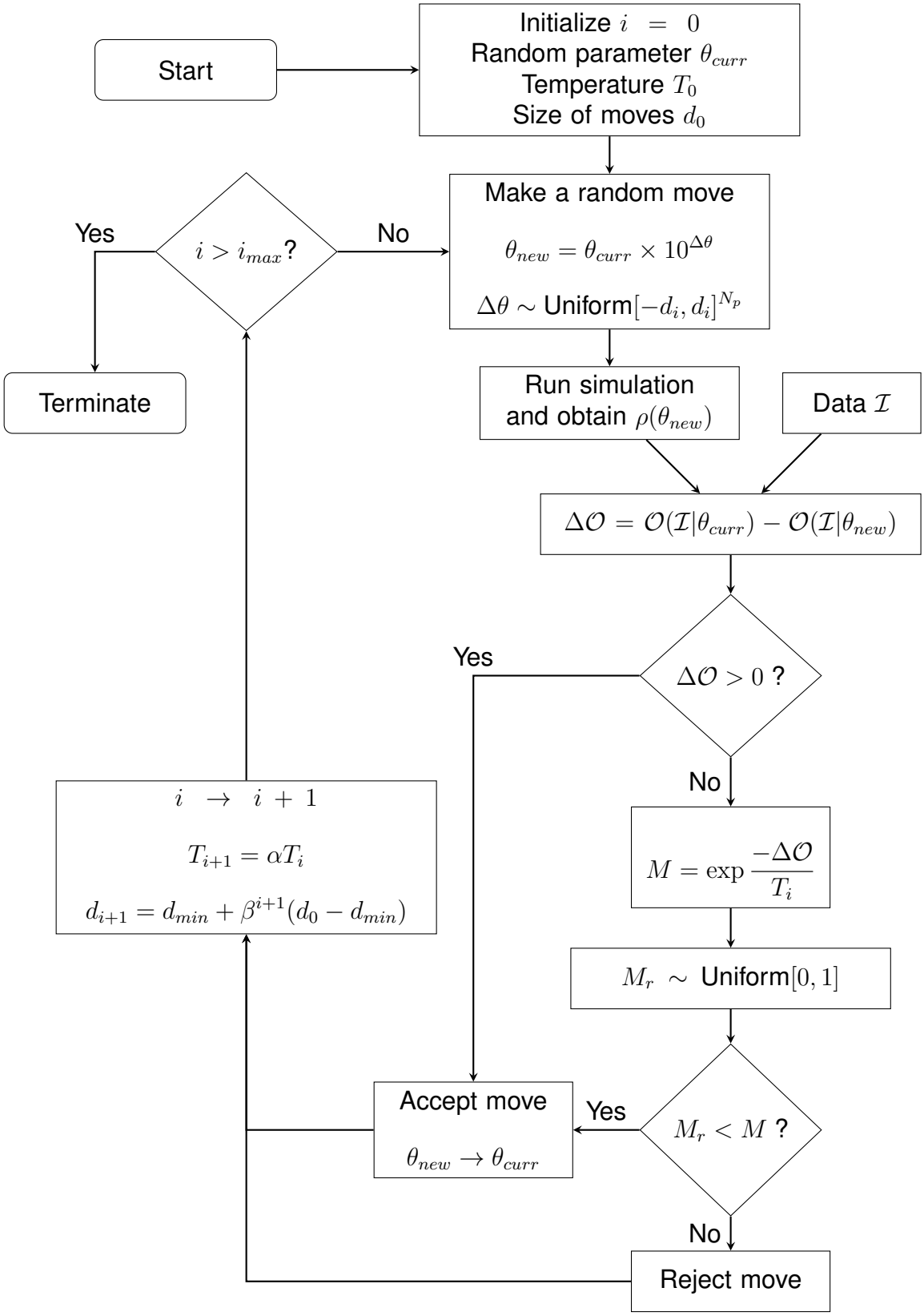


Figure 3.5: Flowchart summarizing the algorithm used for simulated annealing

transition from a global search to a local search. I adopt the commonly used exponential schedule (Ingber 1993) for the lack of a standardized cooling schedule. I also gradually decreased the size of the moves made over time to increase the precision of the search. The algorithm is terminated after a fixed number of steps.

3.2.3 Identifiability analysis

Identifiability refers to the ability of the model and experimental data to determine the parameter values. A model parameter is said to be identifiable if we obtain a finite confidence interval for its estimation. There are two kinds of parameter identifiability (Wieland et al. 2021):

- **Structural identifiability** deals with properties of the model structure itself which renders some parameters indeterminable for a given set of observables independent of the specific values of experimental data. A parameter is considered structurally non-identifiable if varying its value does not change the predicted model dynamics, i.e., the values of the observables (or output channels). A reparametrization of the model typically resolves the presence of structural non-identifiability.
- **Practical identifiability** deals with the inability to determine parameters due to the lack of information in the experimental data. For a structurally identifiable parameter and a given set of experimental data, practical identifiability looks at the finiteness of the confidence interval in determining the parameter value. Practical non-identifiability is typically resolved by modifying the experimental design, for example, to provide more information or to expand the range of experimental conditions.

A rich theory exists for the identifiability analysis of linear ODEs, such as using Fisher information matrix (FIM). More recent developments are being made to develop mathematical tools for the identifiability analysis of non-linear ODE models (Wieland et al. 2021). We still lack *a priori* methods to determine structural identifiability in PDE models. However, a successful method for identifiability analysis using experimental data is the profile likelihood (Murphy and Van Der Vaart 2000). It applies to calibrating any mathematical model where a statistically motivated objective function can be defined.

Profile Likelihood

The profile likelihood, henceforth referred to as PL, is a function of one of the components of the parameter vector. It returns the maximum value of the likelihood function for a fixed value of the parameter of interest. It is defined as the following:

$$\mathsf{PL}_j(p) = \max_{\theta|\theta_j=p} \mathsf{LL}(\mathcal{I}|\theta)$$

where $\mathsf{LL}(\mathcal{I}|\theta)$ denotes the log-transformed likelihood of obtaining the dataset $\mathcal{I}$ for the parameter vector θ. Let π be a placeholder function for the PL that ignores all the terms

independent of the parameter vector. In this case, we get the following relation between the PL and the objective function:

$$\pi_j(p) = -2\text{PL}_j(p) + \text{ constant} = \min_{\theta|\theta_j=p} \mathcal{O}(\mathcal{I}|\theta) \tag{3.5}$$

This is an *a posteriori* method that uses the N_p-dimensional likelihood function and condenses it into a 1-dimensional function that summarizes the multidimensional landscape. In addition, it is possible to infer parameter identifiability (both structural and practical) from the shape of this new function (Raue et al. 2009).

1. If π_j stays flat against the parameter axis, the parameter is said to be structurally non-identifiable. In this case, we can infer that changing the parameter value does not change the minimum attainable value of the objective function. This may be due to redundancy in parameterization or because a change in the parameter of interest does not affect the predicted dynamics.

2. If π_j rises along a direction of the MLE but does not rise above a statistically significant threshold due to attaining some lower asymptote, the parameter is said to be practically non-identifiable. This signifies that although a particular parameter value maximizes the likelihood, it has an infinite confidence interval. In theory, practical non-identifiability can be resolved by increasing the range of experimental conditions or replicates of data.

3. If π_j rises above a statistically significant threshold on both sides of the MLE, the parameter is said to be identifiable. In this case, we can estimate a finite confidence interval for the parameter.

The confidence interval threshold is determined using the error model of the experimental data. Under the commonly used assumption of independent normally distributed error, the theory claims that the confidence interval threshold in the PL is determined by the chi-squared distribution (Kreutz et al. 2013; Feder 1968). For a general error model, an alternate "bootstrapping" method is suggested (Xun et al. 2013). Here, an ensemble of simulated data is produced using the predicted trajectory at MLE and augmenting error sampled using the error model. The objective function is optimized for each realization of the simulated data, and the minimum is recorded. The 95% confidence interval is determined by the 95$^{\text{th}}$ percentile of the obtained distribution of optimized objective function.

PL is, however, generally a computationally expensive method as it requires multiple rounds of re-optimization of the likelihood. Each optimization further requires replicates to ensure confidence, as SA does not guarantee precision. However, a parallel processing system can produce these replicates and reduce computing time. In this regard, I use clusters of the Advanced Research Computing facilities offered by the Digital Research Alliance of Canada to parallelize the codes.

The Python scripts used run these algorithms are available on my GitHub repository.

Chapter 4

Results

In this chapter, we will first see the results from the numerical simulations of the model where one can see the variation in the morphology of density profiles affected by the dispersal parameters and limited nutrient diffusion. Following this, we will see the implementation of SA to estimate the model parameters. Finally, we will plot the PLs of all the parameters and infer their identifiability.

4.1 Simulations of the growth-induced dispersal model

Being a novel formulation, let us look at some preliminary results of the continuum model presented in Chapter 2 to confirm that the model displays expected behaviour. For this, I run simulations of the 1-dimensional version of the model starting with a point inoculation at the center. The numerical simulations are carried out at a spatial resolution of $\Delta x = 25\mu m$ and temporal resolution of $\Delta t = 1s$.

1. $Q_{1/2}$ and $P_{1/2}$ affects the colony morphology.

In Figure 4.1, I have fixed the values of all parameters except $Q_{1/2}$ and $P_{1/2}$. In the four panels, I have plotted the cell density in 1-dimensional space at different time-points ($t = 0, 15, 30, 45, 60$ hours) for different combinations of values of $Q_{1/2}$ and $P_{1/2}$. In each case, each of their values is either high (=1000) or low (=10). We observe the following:

- Lower values of $Q_{1/2}$ results in a larger colony. According to the model, this is the case where growth pressure acts strongly even at low cell densities. Therefore the growth-induced dispersal model predicts greater dispersal for lower $Q_{1/2}$.
- $P_{1/2}$ qualitatively has a less drastic effect on colony expansion. However, we can see a subtle effect on the morphology of the colony. Lower $P_{1/2}$ means that the neighbourhood exerts greater resistance to dispersal even at small cell densities. This results in a tapered ends in the colony, compared to higher values of $P_{1/2}$. It needs to be checked statistically whether the effect of $P_{1/2}$ (and hence the function p) is significant.

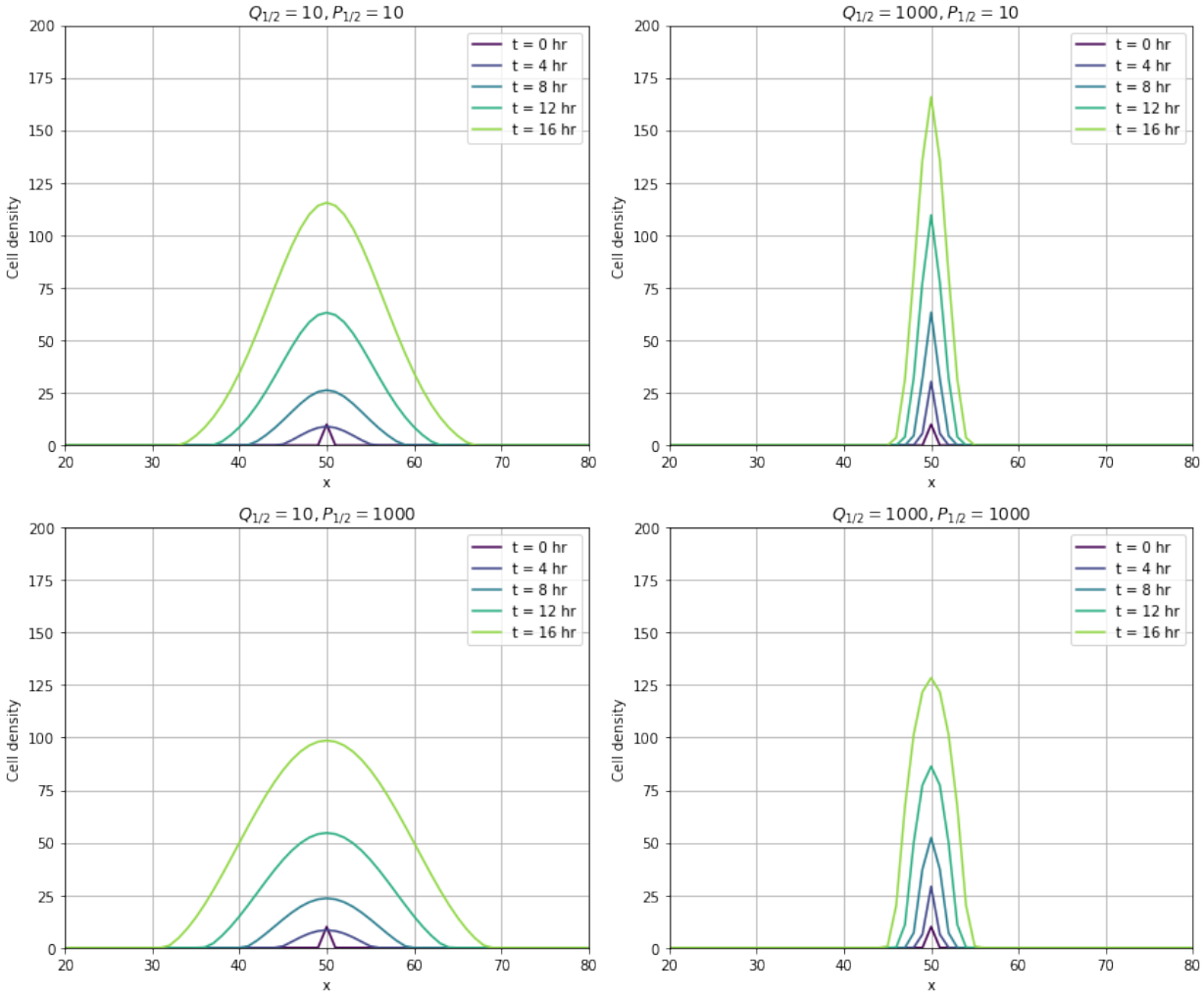


Figure 4.1: Change in colony morphology in response to $R_{1/2}$, $Q_{1/2}$ and $P_{1/2}$.

Initialization: $\rho(x, t = 0) = \begin{cases} 10 \text{ for } x = 50 \\ 0 \text{ elsewhere} \end{cases}$, $S(x, t = 0) = 2$ for all x.

Parameters: $\mu = 0.00022$, $R_{1/2} = 52$, $K_s = 1.46$, $\delta_0 = 2806$, $\delta_s = 58$.

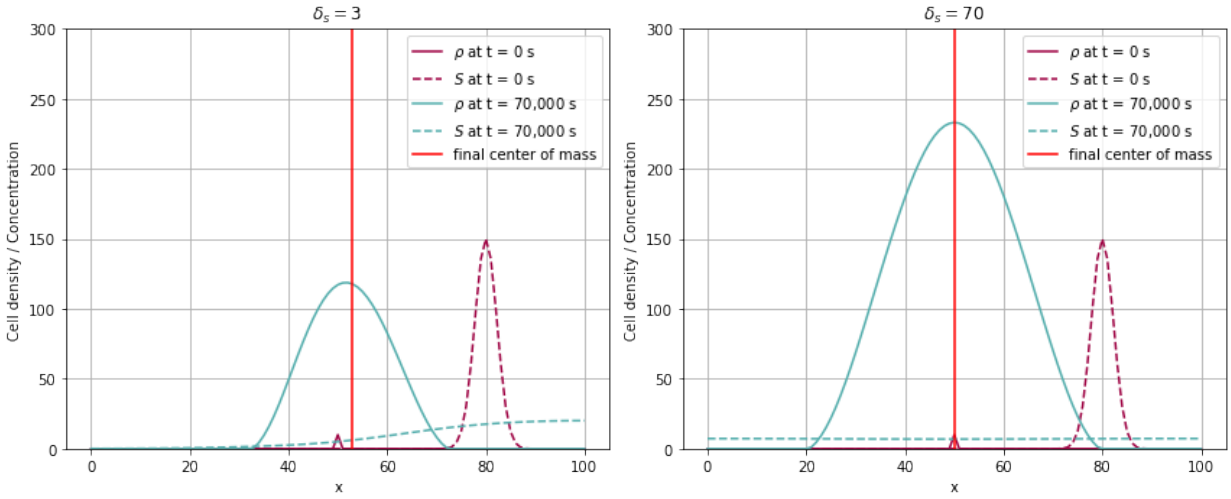


Figure 4.2: Effect of nutrient diffusion on colony growth.

Blue lines show the cell density distribution, Purple lines show the nutrient concentration, dotted lines show the initial state, and solid lines show the final state. Red line locates the center of mass of the colony.

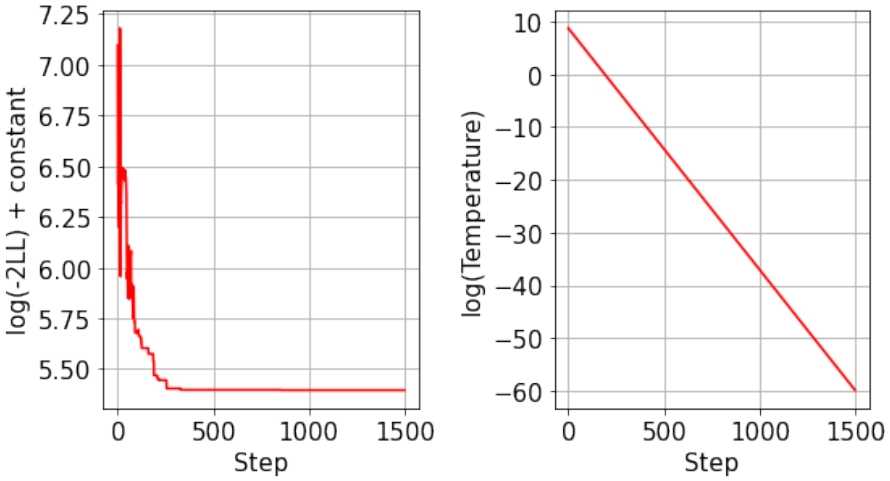
Initialization: $\rho(x, t = 0) = \begin{cases} 10 \text{ for } x = 50 \\ 0 \text{ elsewhere} \end{cases}$, $S(x, t = 0) = 150 \exp(\frac{-(x-80)^2}{10})$.

Parameters: $\mu = 0.00022$, $R_{1/2} = 52$, $Q_{1/2} = 10$, $P_{1/2} = 1000$, $K_s = 1.46$, $\delta_0 = 2806$.

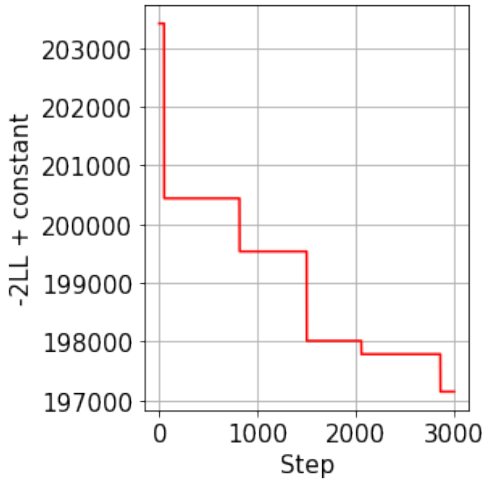
2. Diffusion of nutrients affects the colony morphology.

Fujikawa and Matsushita (1991) shows experimentally that the direction of gradient in nutrient concentration causes the colony to grow towards higher nutrient concentration. Such an effect can only be studied by a model explicitly looking at nutrient diffusion and uptake. Figure 4.2 displays two cases where all parameter values except the diffusion constant of the nutrient are kept constant. I initialize the nutrient concentration to be a Gaussian distribution where the mean lies towards the right of the colony. Comparing the two panels, we can see the that the center of mass of the colony moves towards right when the nutrient diffusion is low. Therefore, this model is able to capture one of the evidences of diffusion limited growth (DLG).

This is a preliminary case study where the choices of parameters are arbitrary. In the next section, I will aim to quantitatively determine the parameters using the materials and methods described in Chapter 3.

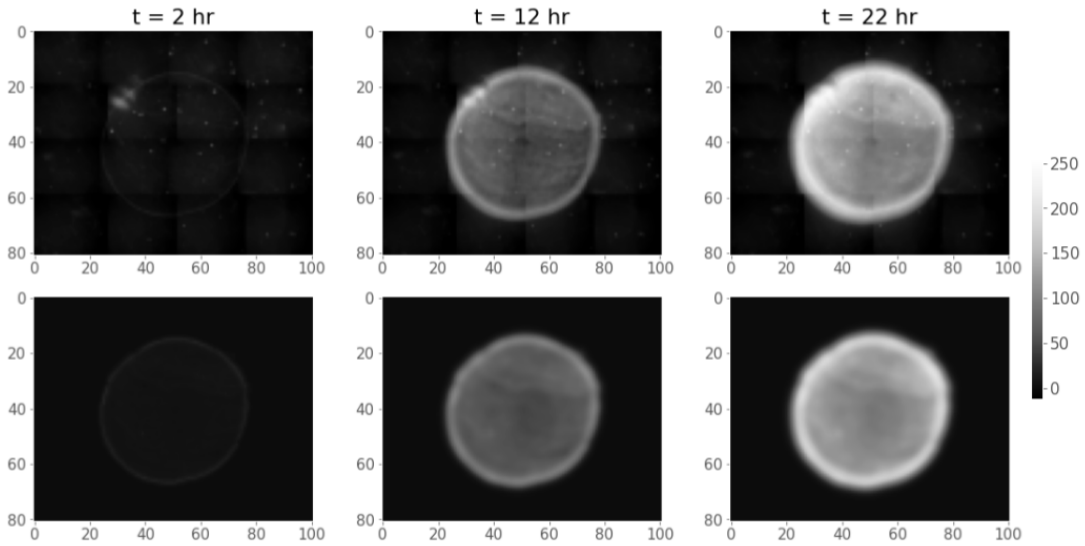


(a) Global optimization using an exponential cooling schedule.

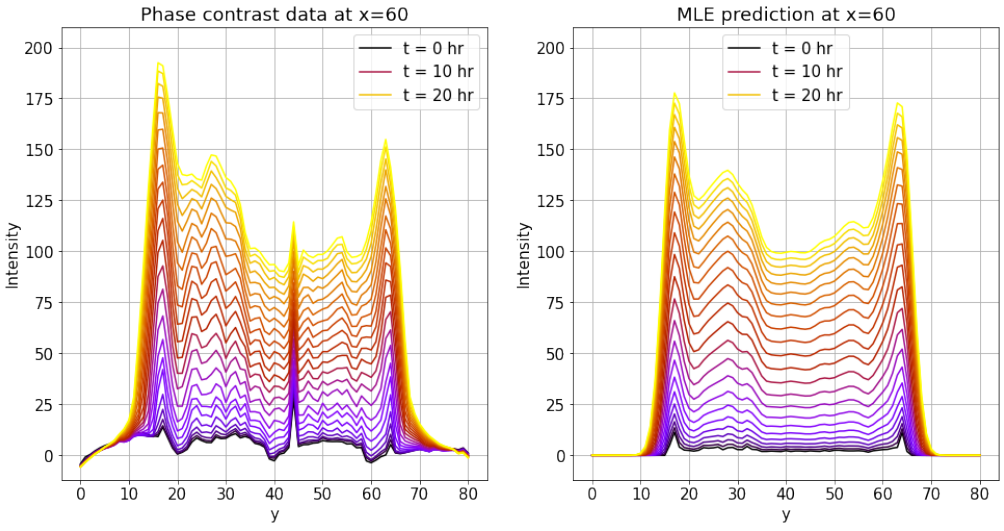


(b) Local optimization (Fine-tuning) using a fixed low temperature.

Figure 4.3: Optimization of objective function using simulated annealing



(Top) Images obtained in the Phase channel are shown at three different time-points. (Below) The intensities predicted by the model at MLE at the corresponding time-points.



The intensity (proxy for cell density) along the y-axis is plotted for a fixed x-coordinate. The colours represent different time-points, each captured at an interval of one hour. Notice the large peaks at the boundary of the culture formed due to the coffee ring effect (Deegan et al. 1997). We can also see noise in the phase contrast data due to background scattering.

Figure 4.4: Phase contrast intensity data v/s Intensity predicted at MLE.

	Initial guess	**After global optimization**	**After local optimization**
$\mathcal{O}(\mathcal{I}\|\theta) =$	1.240×10^7	2.034×10^5	1.946×10^5
κ	2.64	0.920	0.636
μ	1.91 x 10^{-4}	1.28 x 10^{-4}	1.40 x 10^{-3}
$R_{1/2}$	286	177	161
$Q_{1/2}$	61.0	13.7	7.31 x 10^{-2}
$P_{1/2}$	1360	624	1030
Y	78.7	47.9	55.2
K_s	1.75 x 10^{-4}	0.604	13.3
δ_0	733	1200	603
δ_s	164	77.4	80.6

Table 4.1: Parameter values attained during estimation

4.2 Parameter Estimation

4.2.1 Simulated Annealing

The parameter values are estimated by the minimization of the objective function (See Section 3.2.1 and Appendix C) using the process of SA (See Section 3.2.2 and Figure 3.5) SA being a stochastic process, does not ensure a precise estimation of the parameter. Moreover, there is also a non-zero probability that a single realization of the process can converge at a local minimum. In order to improve our efforts in searching for a global minimum, it is recommended to perform multiple replicates of the process and evaluate the final estimates (Ashyraliyev et al. 2009). Figure 4.3 (a) shows the reduction in the objective function in one of the replicates. The estimate corresponding to the lowest objective function is then fine-tuned by subsequent local optimization (See Figure 4.3 (b)). The parameter estimates obtained by this algorithm is shown in Table 4.1. The final values in Table 4.1 will be henceforth treated as the MLE ($\hat{\theta}$). The trajectory of cell density (measured as phase contrast intensity) predicted by the MLE has been visualized in Figure 4.4.

4.3 Identifiability Analysis

Now that we have the parameter estimates, the next task is to perform identifiability analysis. For this, I plot the PL for each parameter. By trial and error, an interval on the

log-scale around each estimate is determined. This interval is then divided into discrete points equally spaced on the log-scale in order to compute the PL. As mentioned earlier, we are required to re-optimize the objective function at every point in this interval in order to find the PL. I use SA for this sequential optimization process. For each step taken in either direction of the MLE, the value of the parameter of interest is kept fixed and SA is run over the remaining subset of the parameter space. The optimized objective function is recorded and the new estimate is used as the initial guess in the next step. Due to the stochastic nature of this process, I perform seven replicates and determine the PL by taking the minimum over these replicates at every point. The identifiability inferred from the shape of the PL for each parameter is summarized in Table 4.2.

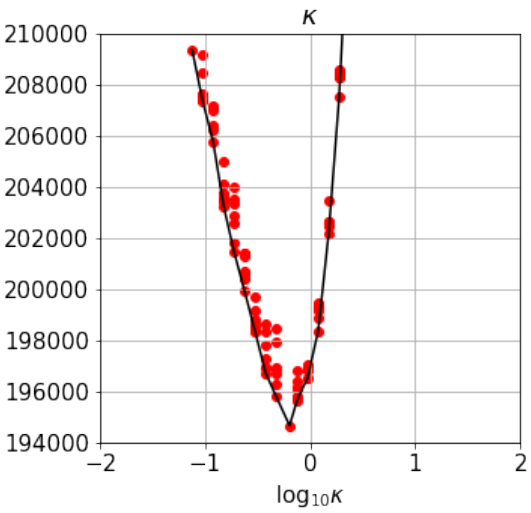
4.3.1 Identifiable Parameters

Four out of the nine parameters show characteristics of identifiability in their PLs (See Figure 4.5) In these plots, the objective function rises monotonically to both sides of the MLE. κ determines the initialization of cell density, $R_{1/2}$ determines the reduction in vertical growth rate in response to cell density, YY determines the overall vertical height that can be attained for a given initial nutrient concentration, and δ_s is responsible in determining the relative growth rate along the radius of the circular culture by controlling the access to nutrient. These four parameters are responsible for determining the vertical intensity profile which has been directly measured in the data set. Note that $R_{1/2}$ is identifiable because there is a density-dependent reduction in growth rate within the data set used for calibration. If the data was collected only for a short time-interval such that the cell density was too small for this effect, the model would predict uninhibited growth and $R_{1/2}$ will be practically non-identifiable. In mathematical terms, if the measured cell density was much smaller than the true value of $R_{1/2}$ (weak density-dependent inhibition of growth), then,

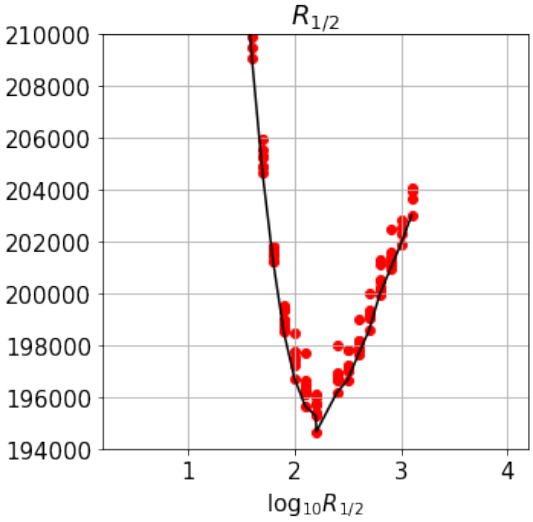
$$r(\rho) = \frac{R_{1/2}}{R_{1/2} + \rho} \simeq \frac{R_{1/2}}{R_{1/2}} = 1$$

In this range of cell densities, change in $R_{1/2}$ would not be reflected in the obtained data set, and hence would be a source of practical non-identifiability.

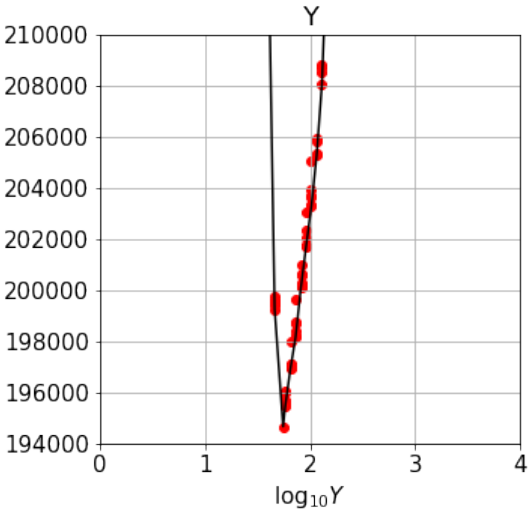
Although the shape of PL depicts identifiability, the statistical inference of finite confidence intervals is still in question. To estimate the confidence interval, I use the bootstrapping method described in the earlier chapter. Figure 4.6 shows the distribution of objective function, each value optimized for a randomly generated realization of data (simulated data) using the error model presented in Section 3.2.1. This distribution reflects the stochasticity in estimating the minimum objective function due to the stochasticity in the data. In theory, the 95$^{\text{th}}$ percentile of this distribution denotes the objective function value below which there is a 95% chance of finding $\mathcal{O}(\mathcal{I}|\hat{\theta})$ for a random data set $\mathcal{I}$ and its corresponding MLE $\hat{\theta}$. However, we observe that the minimum value of the objective function attained with the real data set (1.946 x 10^5) is much higher than values attained for the simulated data (in the range 1.61 x 10^5 - 1.63 x 10^5) and therefore cannot determine a finite confidence interval for these parameters. Statistically, this is a



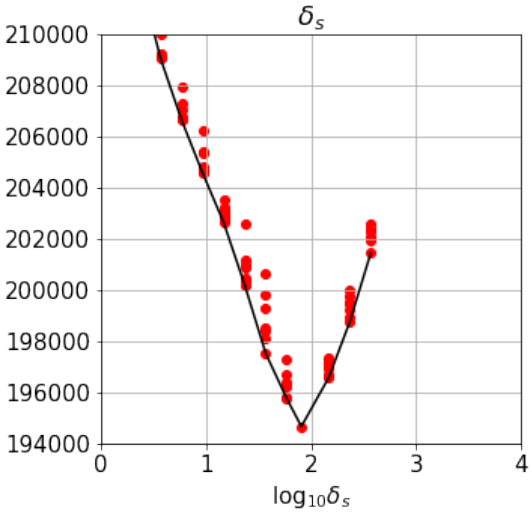
(a) PL of κ



(b) PL of $R_{1/2}$



(c) PL of Y



(d) PL of δ_s

Figure 4.5: Profile Likelihood of identifiable parameters

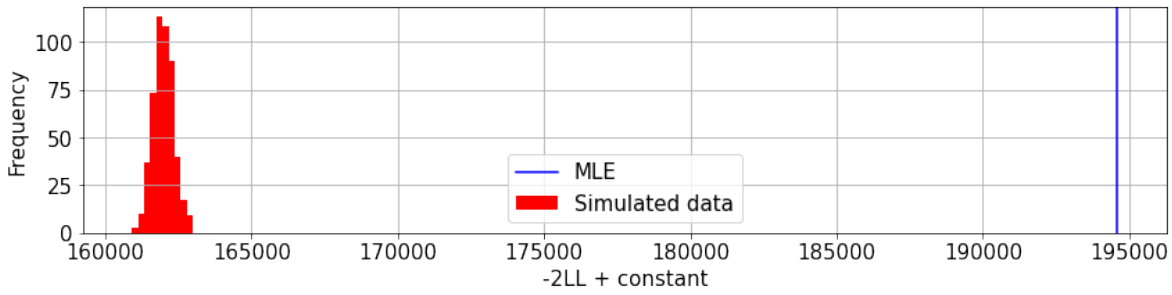


Figure 4.6: Frequency distribution of objective function optimized for simulated data

case of underfitting (Villaverde et al. 2021). We will discuss further implications of this observation in Chapter 6.

4.3.2 Non-identifiable parameters

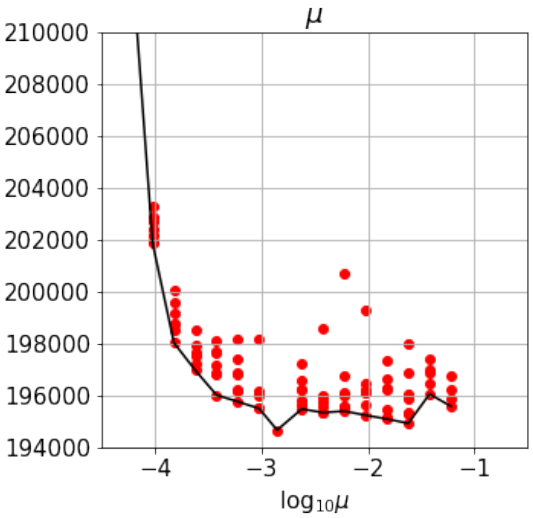
Although we do not directly measure μ and K_s, these are also parameters that have a strong control over the vertical profile of cell density. However, in Figure 4.7 (a) and (b), we can see that their PLs flatten out to the right (for higher values) which implies practical non-identifiability. In this flat region, we also observe a strong correlation between the values of μ and K_s (See Figure 4.7 (c)). We infer that any change in the objective function due to changing one of their values is compensated by a change in the value of the other parameter. However, it needs to be noted that this redundancy does not occur for all values of the parameters and hence these are structurally identifiable parameters. The MLE of K_s is of the order of 10s or more, while the maximum nutrient concentration is set to 1. Using this observation, we can approximate the nutrient uptake function as

$$g(S) = \mu \frac{S}{K_s + S} \simeq \frac{\mu}{K_s} S$$

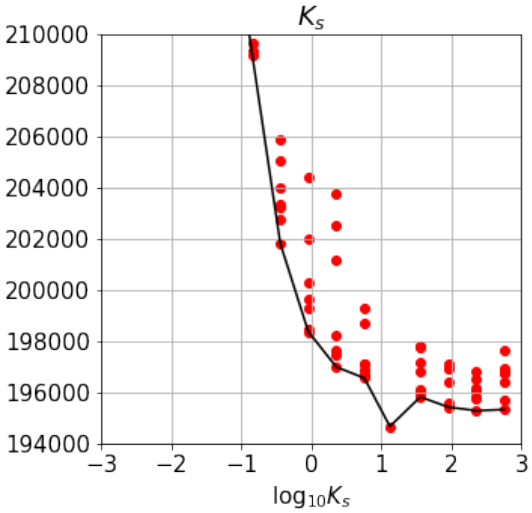
From this result, we infer that the parameters μ and K_s are practically non-identifiable when dealing with a nutrient-limited experimental apparatus i.e., $S_{t=0} << K_s$. Here, we learn that we require data collected at high nutrient concentrations to estimate these parameters individually. In a nutrient-limited apparatus, the parameter $\frac{\mu}{K_s}$ would be identifiable. This paves way for model reduction by assuming a linear dependence of growth on nutrient concentration for a dataset with limited nutrient availability. Refer to Section 6.1 for further discussion on reparametrization of the nutrient uptake term.

The remaining parameters, $Q_{1/2}$, $P_{1/2}$ and δ_0, contribute to determining the dispersal of cells. We can see in Figure 4.8 (a) that the PL of $P_{1/2}$ also flattens out for high values, displaying practical non-identifiability. Like K_s, $P_{1/2}$ is also a Monod function parameter that is estimated to be much larger than the variable that the function is dependent on, i.e. $\rho << P_{1/2}$. Therefore, we can apply the following approximation:

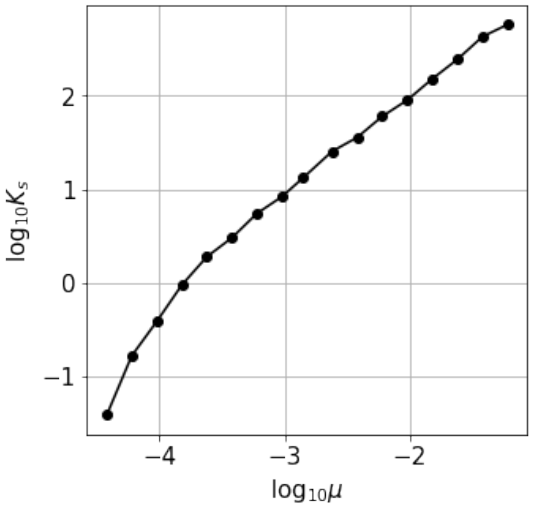
$$p(\rho) = \frac{P_{1/2}}{P_{1/2} + \rho} \simeq \frac{P_{1/2}}{P_{1/2}} = 1$$



(a) PL of μ

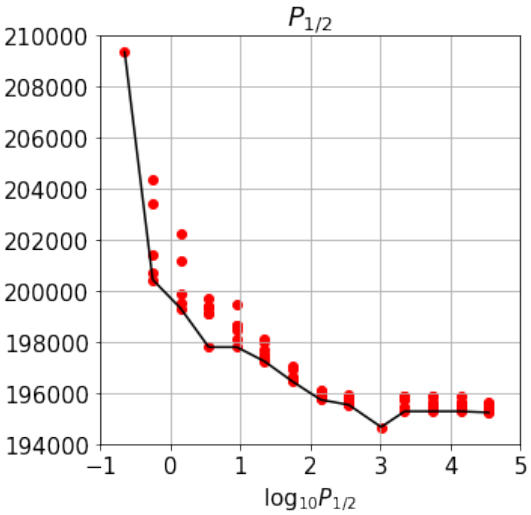


(b) PL of K_s

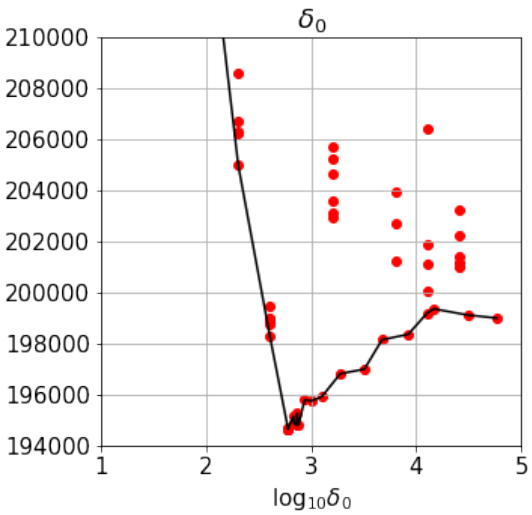


(c) μ and K_s show high correlation

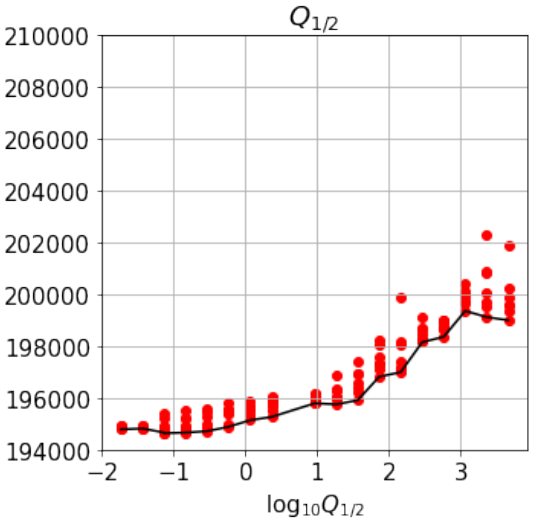
Figure 4.7: Profile Likelihood of redundant practically non-identifiable parameters



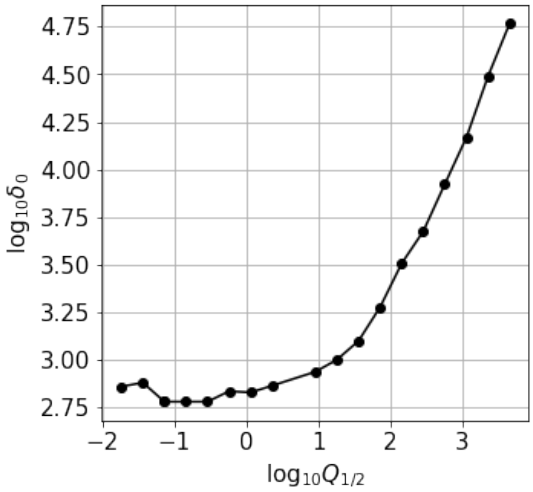
(a) PL of $P_{1/2}$



(b) PL of δ_0



(c) PL of $Q_{1/2}$



(d) δ_0 and $Q_{1/2}$ show high correlation at large values

Figure 4.8: Profile Likelihood of other non-identifiable parameters

However, unlike μ and K_s, this is not a case of redundancy because around the minimized value of the objective function, $P_{1/2}$ does not show significant correlation with any other parameter. $P_{1/2}$, within the range of density values we are dealing with, disappears from the model, indicating that the neighboring cell densities had no role in determining the dispersal of cells in the experimental apparatus used.

The PL of $Q_{1/2}$, shown in Figure 4.8 (c), attains a minimum at very small values and flattens out to the left, again, displaying practical non-identifiability. The rise of the curve to the right may not be statistically significant, and therefore it might not be possible to apply the approximation that $Q_{1/2} << \rho$ right away. At large values of $Q_{1/2}$, there is also a source of redundancy in parameters (See Figure 4.8 (d)). In this limit,

$$q(\rho) = \frac{\rho}{Q_{1/2} + \rho} \simeq \frac{\rho}{Q_{1/2}}$$

i.e., the growth pressure is scaled linearly by the local density. Using this approximation, we can re-write the transport term as

$$\delta_0(p\nabla^2(Gq) - Gq\nabla^2 p) \simeq \frac{\delta_0}{Q_{1/2}}(p\nabla^2 G - G\nabla^2 p)$$

The redundancy in the parameters δ_0 and $Q_{1/2}$, in this limit, predicts a flat PL. Since the MLE does not fall in this range of parameter values, the plateau is seen at a higher value of the objective function to the right of the MLE. The statistical significance of this rise will determine whether we can apply the limit of very small $Q_{1/2}$, in which case, the parameter disappears from the model.

$$q(\rho) = \frac{\rho}{Q_{1/2} + \rho} \simeq \frac{\rho}{\rho} = 1$$

For δ_0, we can see in Figure 4.8 (b) that the objective function rises monotonically to both sides of the MLE, However, due to the redundancy with $Q_{1/2}$, the PL reaches an asymptote to the right. The statistical significance of the rise in the objective function before saturating determines whether δ_0 is identifiable or practically non-identifiable. Refer to Section 6.1 for further discussion on reparametrization of the transport term.

Parameter	Shape of PL with respect to the MLE	Identifiability
κ	Monotonic rise on both sides	Identifiable*
μ	Monotonic rise to the left, flat to the right	Practically non-identifiable
$R_{1/2}$	Monotonic rise on both sides	Identifiable*
$Q_{1/2}$	Flat to the left, rise and plateau to the right	Practically non-identifiable
$P_{1/2}$	Monotonic rise to the left, flat to the right	Practically non-identifiable
Y	Monotonic rise on both sides	Identifiable*
K_s	Monotonic rise to the left, flat to the right	Practically non-identifiable
δ_0	Monotonic rise to the left, rise and plateau to the right	Identifiable or practically non-identifiable*
δ_s	Monotonic rise on both sides	Identifiable*

Table 4.2: Inference of identifiability from the shape of the Profile Likelihood
* Uncertainty estimation requires knowledge of the confidence interval threshold

Chapter 5

Case study: Antibiotic cross-protection mutualism

Antimicrobial resistance has recently emerged as a global health crisis due to the over-exposure of microbes to antibiotics (Morehead and Scarbrough 2018). The acquisition of resistance by some cells can also confer resistance to sensitive cells through conjugation and cross-protection. In the first case, it has been observed that bacteria can acquire resistance against an antibiotic by receiving plasmids carrying resistance genes through conjugation (Graf et al. 2019). In the latter case, the enzymatic inactivation of antibiotics in the environment by the resistant cells can protect the sensitive cells in its surrounding (Yurtsev, Conwill, and Gore 2016). When multiple cells with resistance against different antibiotics are present together, they can protect each other from the antibiotics in the environment, an effect that is classified as cross-protection mutualism. In this section, I will focus on cross-protection mutualism through enzymatic inactivation in a simple consortium of two isogenic strains, each with resistance against a different antibiotic. As a first speculative step towards understanding their spatiotemporal dynamics, I employ the multispecies version of the continuum model to simulate this community.

The two strains used in this study are derived from *E. coli* MG1655 and differ by means of an inserted plasmid. The plasmid in strain 1 has a pEB1 backbone, Kanamycin resistance as a selectable marker, and expresses a yellow fluorescent protein (mEYFP). The plasmid in strain 2 has a pEB1 backbone, Chloramphenicol resistance as a selectable marker, and expresses a cyan fluorescent protein (SCFP3A). Refer to Balleza, Kim, and Cluzel (2018) for more information on these plasmids and fluorescent proteins.

5.1 Data Collection

To obtain the time-lapse data for the co-culture with antibiotic-mediated cross-protection mutualism, I follow the same protocol followed for the monoculture as described in Section 3.1 with the following modifications:

- Each strain in grown separately in test tubes containing 0.4% M9 glucose and the

corresponding antibiotic selector (50μg/mL Kanamycin for strain 1, and 25μg/mL Chloramphenicol for strain 2). They are then individually harvested during the exponential phase. The cells are washed, centrifuged, re-suspended in the M9 medium without the antibiotics, and brought to a similar optical density (roughly OD_{600}=1). Equal volumes of the suspended cultures are then mixed to obtain a co-culture of similar optical density. 1μL of this mixture is spotted on the agarose pad.

- During the preparation of agarose pads, both antibiotics are added to the cooling agarose solution, obtaining concentrations of 5μg/mL for Kanamycin and 3.75μg/mL for Chloramphenicol.
- While imaging the co-culture, an additional fluorescence channel is used. The CFP channel shines and captures wavelengths of light that excite and is emitted by SCFP3A. This range of wavelengths does not interfere with the spectrum used by the Cy3 channel and vice-versa, allowing us to observe the two strains distinctly.

mEYFP and SCFP3A have significantly different maturation times and brightness (Balleza, Kim, and Cluzel 2018). In addition, it has been found that the light emitted per cell for GFP-derived fluorescent proteins depends on factors like dissolved oxygen and, consequently, the growth stage (Drepper et al. 2010). Since these effects are difficult to account for due to the lack of complete characterization, the fluorescence intensity measures are poor indicators of cell density. Hence, we do not have a measure of the relative abundance of each strain. I proceed with a qualitative comparison of the observations with simulations.

5.2 Model

The response of cell growth to the concentration of an antibiotic (antibiotic inhibition) is typically modelled using a Hill function with two parameters: the half-maximal inhibition concentration (IC_{50}), and the Hill coefficient (n) that measures the "dose sensitivity" (Regoes et al. 2004). Therefore, we can define the per unit density growth rate of a strain sensitive to an antibiotic A as

$$g(S, A) = \mu \left(\frac{S}{K_s + S} \right) \left(\frac{1}{1 + \left(\frac{A}{IC_{50}} \right)^n} \right)$$

Incorporating this new definition of growth rate, we can rewrite Equation 2.5. Let ρ_1 and ρ_2 denote the cell densities of strain 1 and strain 2 respectively. Let A_1 and A_2 denote the concentrations of Kanamycin and Chloramphenicol respectively. Then, we can describe

the dynamics of cell density as following:

$$\frac{\partial \rho_k}{\partial t} = \delta_0 \left(p\nabla^2 \left(\frac{Gq\rho_k}{\rho} \right) - \frac{Gq\rho_k}{\rho} \nabla^2 p \right) + G_k r \quad , k = 1, 2$$

$$G_1(\rho_1, S, A_2) = \rho_1 g_1(S, A_1) = \mu_1 \rho_1 \left(\frac{S}{K_s + S} \right) \left(\frac{1}{1 + \left(\frac{A_2}{IC_{50,2}} \right)^{n_2}} \right)$$

$$G_2(\rho_2, S, A_1) = \rho_2 g_2(S, A_2) = \mu_2 \rho_2 \left(\frac{S}{K_s + S} \right) \left(\frac{1}{1 + \left(\frac{A_1}{IC_{50,1}} \right)^{n_1}} \right)$$

$$G = G_1 + G_2$$

where r, q and p are functions of the total cell density $\rho = \rho_1 + \rho_2$.

To complete the model, we need to describe the spatiotemporal dynamics of the contents in the agarose medium. I use the same RD model for nutrient diffusion that was used for the monoculture.

$$\frac{\partial S}{\partial t} = \delta_s \nabla^2 S - \frac{Gr}{Y}$$

The change in antibiotic concentrations is also modelled using RD equations. The reaction term here represents the loss of antibiotics in the system due to the inactivation by the resistant strains. Since this process is known to occur within the cells, I use Michaelis-Menten kinetics for enzyme reaction, assuming that the density of the resistant strain is proportional to the concentration of the inactivating enzyme.

$$\frac{\partial A_1}{\partial t} = \delta_{A1} \nabla^2 A_1 - v_1 \frac{\rho_1 A_1}{K_{m,1} + A_1}$$

$$\frac{\partial A_2}{\partial t} = \delta_{A2} \nabla^2 A_2 - v_2 \frac{\rho_2 A_2}{K_{m,2} + A_2}$$

This formulation comes with an abundance of new parameters. During simulations, I will use the MLE values for parameters that are retained from the monoculture model calibration. For the newly introduced antibiotic-related parameters, it is not the goal of the project at this stage to estimate them and hence I will rely on some guess values. The parameters in the cross-protection model are summarized in Table 4.3.

5.3 Results

In the time-lapse images of the co-culture, we can observe the following phenomena (See Figure 5.1):

- For the first few hours, there is slow growth in strain 1 (yellow cells) and almost no observable growth in strain 2 (cyan cells). Then we see sudden, rapid and radial

Parameter	Meaning	Dimensions
δ_0	Transport term coefficient of cells	$[L^2]$
δ_s	Diffusion coefficient of nutrient	$[L^2T]$
δ_{A1}	Diffusion coefficient of Kanamycin	$[L^2T]$
δ_{A2}	Diffusion coefficient of Chloramphenicol	$[L^2T]$
μ_1	Intrinsic growth rate of strain 1	$[T^{-1}]$
μ_2	Intrinsic growth rate of strain 2	$[T^{-1}]$
K_s	Nutrient concentration at half maximal growth	$[ML^{-3}]$
Y	Yield of cell biomass per unit mass of nutrient	dimensionless
$IC_{50,1}$	Half maximal inhibition concentration of Kanamycin	$[ML^{-3}]$
$IC_{50,2}$	Half maximal inhibition concentration of Chloramphenicol	$[ML^{-3}]$
n_1	Hill coefficient of Kanamycin response	dimensionless
n_2	Hill coefficient of Chloramphenicol response	dimensionless
v_1	Maximum rate of inactivation of Kanamycin	$[M^{-1}L^3T^{-1}]$
v_2	Maximum rate of inactivation of Chloramphenicol	$[M^{-1}L^3T^{-1}]$
$K_{m,1}$	Half maximal inactivation concentration of Kanamycin	$[ML^{-3}]$
$K_{m,2}$	Half maximal inactivation concentration of Chloramphenicol	$[ML^{-3}]$
$R_{1/2}$	Total cell density at half maximal local growth	dimensionless
$Q_{1/2}$	Total cell density at half maximal growth pressure	dimensionless
$P_{1/2}$	Neighbouring cell density at half maximal cell flow rate	dimensionless

Table 5.1: List of parameters in the cross-protection model.

growth of the cyan cells at specific locations. Following these “explosions”, we see an increasing growth of the yellow cells, particularly around the regions where these explosions originated.

- The explosions occur first in regions with higher initial cell density and later in regions with lower cell densities. Therefore, there is a lag in the onset of the growth of cyan cells depending on the cell density.

Kanamycin and Chloramphenicol are ribosome-targeting antibiotics with two different mechanisms of action. Kanamycin is a bactericidal drug that irreversibly binds to the ribosome and induces mistranslation (Davis 1987), while Chloramphenicol is a bacteriostatic drug that binds reversibly and inhibits peptide bond formation (Harvey and Koch 1980). Due to this difference in mechanism, Kanamycin has been known to show a significantly higher magnitude of dose sensitivity than Chloramphenicol (Greulich et al. 2015). In light of this information, I suspect that the explosion of cyan cells (Kanamycin sensitive cells) could be attributed to the concentration of Kanamycin falling beneath a sharp dose threshold due to inactivation by the Kanamycin resistant cells. Therefore, I set the dose sensitivity parameters to $n_1 = 1.1$ and $n_2 = 10$ in the model. The remaining antibiotic-related parameters are set to some guess values based on trial and error in order to capture the experimental observation. Figure 5.2 shows the results from the simulations.

The initial condition is set to a point inoculum at the center of the frame with an equal abundance of both strains. For a more quantitative comparison with the time-lapse images, it is suggested to use the fluorescence images at the initial time-point to initialize the cell densities. The nutrient and antibiotics are uniformly distributed with all initial concentrations normalized to 1. Interestingly, the simulations can qualitatively capture the phenomena observed in the time-lapse images.

- For the first few hours, due to the initial concentrations of the antibiotics and the corresponding dose-response curves defined for the strains (See Figure 5.2 (a)), yellow cells increase in fractional abundance (See Figure 5.2 (b)). During this time, the yellow cells inactivate Kanamycin in its immediate environment. The concentration of Kanamycin, starting in the center, falls below $IC_{50,1}$ and initiates the rapid growth of cyan cells. Below this dose threshold, cyan cells almost attain their maximum growth rate, which causes the explosion phenomenon. Hence, we see a significant increase in cell density and a faster rate of colony expansion. Following this explosion, cyan cells inactivate Chloramphenicol, starting in the middle, allowing for the subsequent growth of yellow cells to higher densities.

- The simulations also predicted the delay in the growth of strain 2 due to lower initial cell densities. This can be attributed to the slower inactivation rate of Kanamycin by the limited amount of yellow cells and, consequently, the longer duration to cross the dose threshold. See Figures 5.2 (a) and (b).

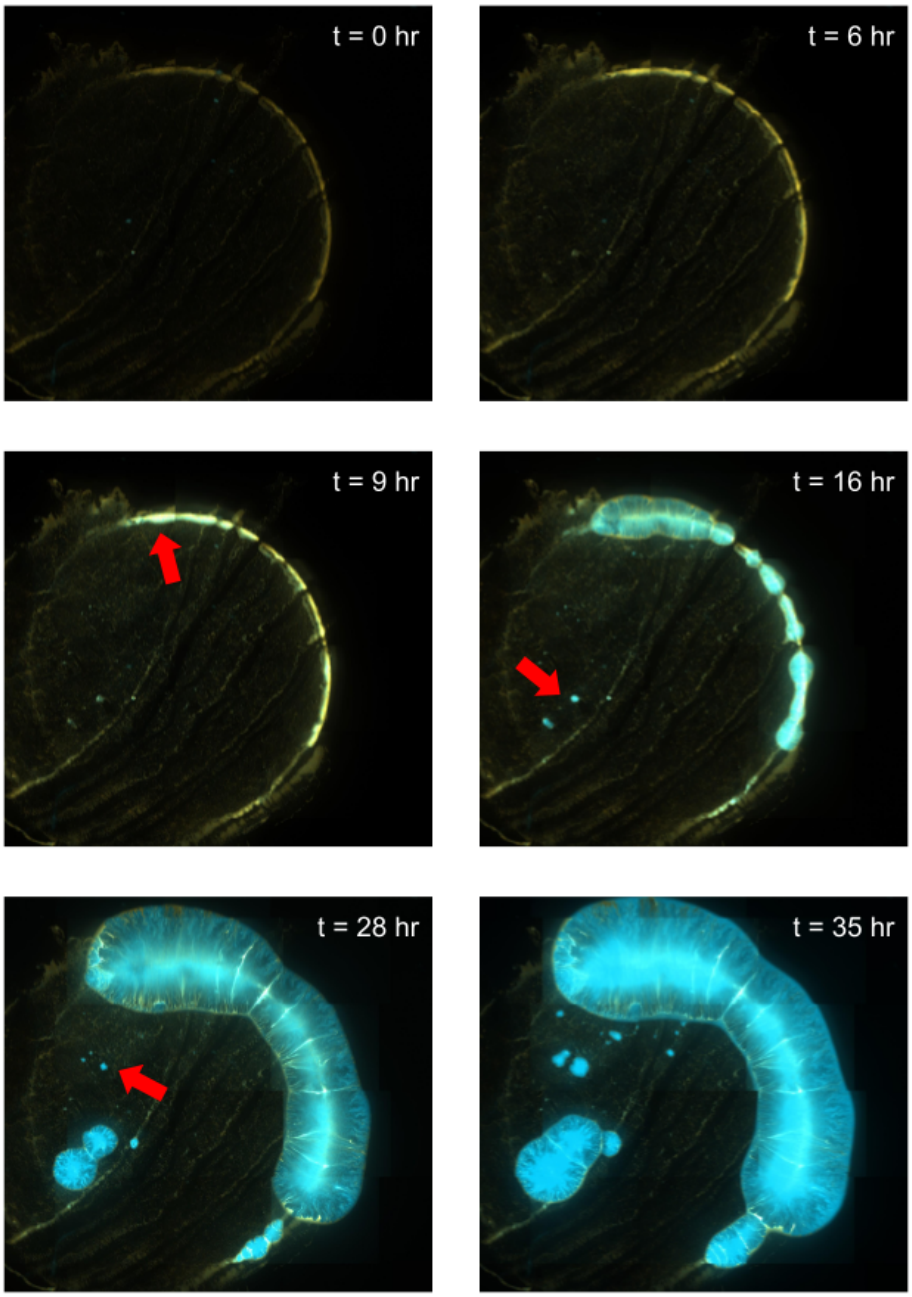
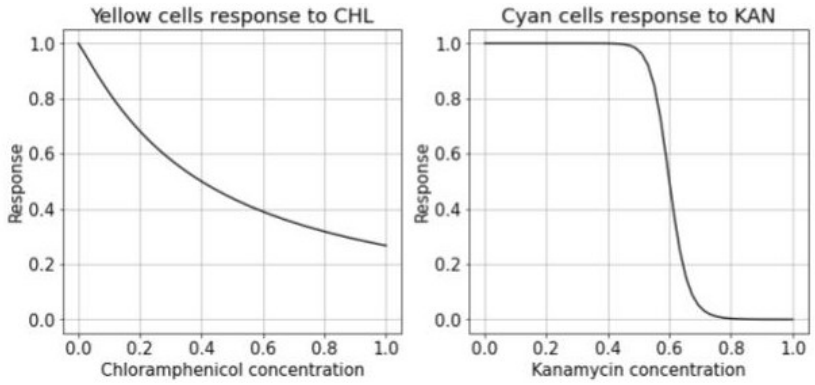
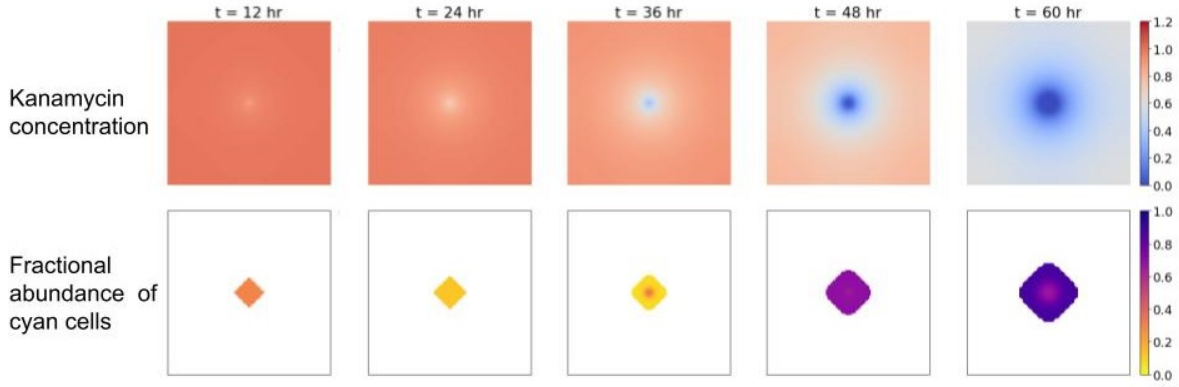


Figure 5.1: Time-lapse images of co-culture.

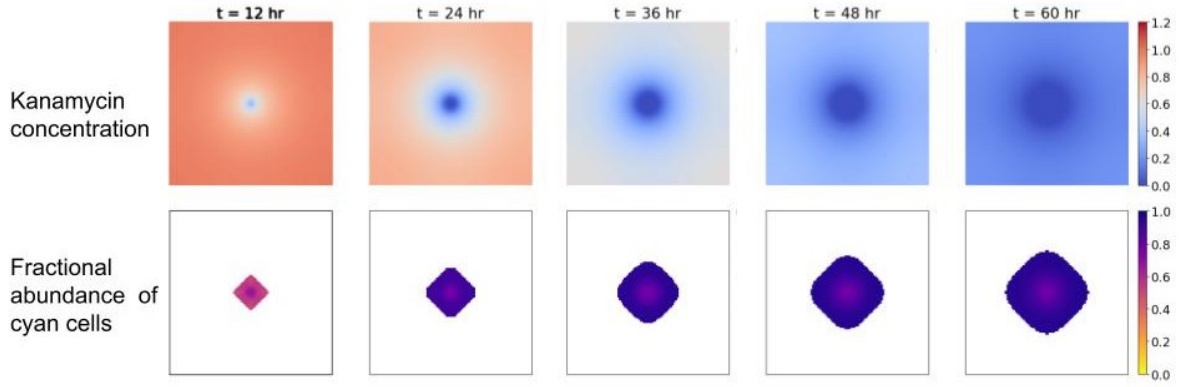
Notice the slow growth of yellow cells over the first six hours and the relative absence of cyan cells. At roughly the ninth hour, there is an onset of growth of cyan cells at the densest regions. Over time, cyan cells proliferate at various sites. There is also an increasing abundance of yellow cells at these sites following the growth of cyan cells. The red arrows point to some regions where the cyan cells are beginning to proliferate.



(a) Dose-response curves



(b) Initial density = 1.0



(c) Initial density = 10.0

Figure 5.2: Simulations of the cross-protection model.
Parameters: $IC_{50,1} = 0.6, IC_{50,2} = 0.4, n_1 = 10, n_2 = 1.1, v_1 = 0.01, v_2 = 0.0075, K_{m,1} = 10, K_{m,2} = 10$

Chapter 6

Discussion and Future work

In this project, I extended a continuum model of biofilms to include growth-induced dispersal such that it is capable of describing the spatiotemporal dynamics of non-motile cells growing on a surface. Following an experimental protocol for time-lapse microscopy of bacteria grown in agarose pads, I used the phase contrast data as a proxy measurement of cell density to calibrate the model. Using the dynamic model calibration pipeline, I obtained parameter estimates and inferred their identifiability. Model calibration is often iterative to improve the model formulation and the confidence in the estimated parameters at every iteration (Balsa-Canto, Alonso, and Banga 2010). Based on the results of this study, I infer possibilities of improving the growth-induced dispersal model and the experimental design to calibrate its parameters.

6.1 Reparametrization and Optimal Experimental Design

In a spatially explicit model with a handful of parameters interacting non-linearly, we generally lack the analytical tools to exhaustively foresee the effects of different combinations of parameters on the model predictions (Ashyraliyev et al. 2009). When numerically solving the model over the entire parameter space becomes computationally expensive, we rely on algorithms such as SA to estimate the parameters that best explain the experimental observations. We must also ensure that these estimates are reliable by employing identifiability analysis.

Identifiability analysis using PL reveals the scope for the improvement of parameter estimation in two ways:

- **Model reparametrization:** If, for data under a specific range of experimental conditions, there is practical non-identifiability of a parameter, we infer that this parameter has no significant effect on the predicted dynamics or that there exist other correlated parameters that exhibit redundancy, within this particular range. In such a case, there is a possibility of eliminating this parameter via model reduction and obtaining a simpler model. Such a simple model is sufficient to explain the true dynamics within the specified set of experimental conditions. Suppose the non-effect

is seen for all ranges of experimental conditions (which is practically impossible to observe and needs to be discovered analytically). In that case, the parameter is structurally non-identifiable, and the model is generally reducible irrespective of the experimental conditions.

- **Optimal Experimental Design:** Optimal Experimental Design (OED) theory comprises a well-studied set of statistical tools used to infer better experimental design to achieve a particular goal such as parameter estimation (Atkinson and Donev 1992). Here, we use PL analysis to infer possible alterations in the experimental design that can help us estimate the model parameters with greater certainty. Non-identifiability can also call for changes in the experimental design. For instance, the practical non-identifiability of a parameter could be attributed to a limited range of experimental conditions, in which case one can explore a broader range of experimental conditions to improve the identifiability of the parameter. Structural non-identifiability could be attributed to a limited number of observables, in which case, one can increase the number of observed quantities to make the parameter structurally identifiable. Practical non-identifiability is sometimes also attributed to the uncertainty in the observations. Minimizing the sources of error or minimizing the measurement uncertainty by increasing the number of replicates can lower the confidence interval threshold relative to the minimum of the PL, making the parameter identifiable.

Let us first discuss possible reparametrizations in the growth-induced dispersal model in light of the results shown in Section 4.3. The following are some ways to remove or introduce parameters to obtain a concise model with optimized parametrization:

1. **Nutrient uptake:** In the range of experimental conditions with limited nutrient availability, μ and K_s display redundancy. Since the MLE falls within this parameter range, we infer that a nutrient-limited system better explains nutrient uptake dynamics. Therefore, for the bacterial strain, the media, and the apparatus described in Section 3.1.1, we can predict the cell densities just as well with a new parameter $\mu_s = \mu/K_s$ which is the intrinsic growth rate per unit concentration of the nutrient. Then, the growth rate per unit cell density is

$$g(S) = \mu_s S$$

2. **Density-dependence of cell movement:** We estimate a very large value of $P_{1/2}$ for the given experimental conditions. We inferred that in this limit, $P_{1/2}$ is practically non-identifiable. This limit is also an extreme case of the function p being constant, implying that cell dispersal is independent of neighboring cell density in the given apparatus. If found to be statistically significant, the results also show an estimate of $Q_{1/2}$ much smaller than the values attained by cell density. In this limit, the movement of cells depends linearly on the gradient of growth rate, which is an extreme case of the function q being a constant for non-zero densities. Therefore,

our results imply that a simpler growth-dependent diffusion-like transport term in the rate of change of cell density is sufficient to capture the same dynamics:

$$\frac{\partial \rho}{\partial t} = \delta_0 \nabla^2 G + Gr$$

The above procedure of model reduction by eliminating q and p, in the biofilm model presented by Eberl and colleagues (See Equations 2.1 and 2.2), returns diffusion dynamics of cells.

$$\frac{\partial \rho}{\partial t} = \alpha_0 \nabla^2 \rho + \text{growth}$$

This is a special case of our model when growth is always exponential. By explicitly modelling nutrient concentration, our model successfully incorporates other growth regimes, such as nutrient-limited growth. It has been studied that the region of a bacterial colony with growing cells along the vertical axis is determined by proximity to the nutrient source due to limited diffusion across cells (Warren et al. 2019). The nutrients are available in the agarose pad in the apparatus used here, which is placed above the cells. As they grow into the pad, it may be hypothesized that the growing cells are localized at the topmost layers, relatively loosely packed. Hence, their movement may fail to show any dependence on cell density. This contrasts the case where the cells grow on top of a solid agar surface. Here, the growing cells are buried and packed beneath a column of non-growing cells and may pose non-linear effects in their density-dependent dispersal.

Now let us see possible alterations in the experimental protocol that help us estimate the parameters in the given model with greater certainty:

1. **Replicates:** The reason behind the lack of replicates in this study is the inability to replicate the initial cell density distribution. It is observed that strictly following the monoculture preparation protocol in Section 3.1.1 does not ensure similar distribution of cells. I suspect that varying evaporation rates of the droplet of bacterial culture and unevenness of the agar surface are two potential causes behind this observation. For calibrating against data from a monoculture with replicates, an alternate protocol that involves growing colonies from single cells is recommended (Bär et al. 2020). By overlaying images of different such colonies, we can get replicates of cell density observed at every point in space and time. The major limitation of this protocol is that it cannot be used for growing co-cultures. Another limitation is that it may take hours to days to see if and where the colonies are growing.

2. **Nutrient concentration:** Nutrient concentration is one of the state variables in the model which is not being measured. This introduces an inherent difficulty in estimating nutrient-related parameters such as the diffusion coefficient δ_s and half-maximal growth concentration K_s. We can increase the glucose concentration in the agarose pads to solve the redundancy problem in estimating K_s. This way, we can improve the chances that the non-linearity in the growth rate when nutrient

concentration is close to the magnitude of K_s is reflected in the obtained data. Additionally, we can calibrate the model over data obtained for different initial nutrient concentrations, thereby increasing the experimental information on nutrient concentration in the data and, consequently, greater confidence in the nutrient-related parameter estimates.

3. **Background scattering:** One of the significant sources of error in the data comes from the light scattered by non-cell disturbances that can be seen in the phase channel. To eliminate this, we can use a fluorescence channel to measure the densities of cells that express a suitable fluorescent protein. There has been some progress in developing modified fluorescent proteins whose emission intensities can be used as reliable proxies for cell densities, which is not possible for traditionally used GFP-derived proteins (Drepper et al. 2010).

Underfitting

Estimating the 95% confidence interval using the bootstrapping method shows that the minimum value of the objective function obtained by SA is much larger than the estimate. We typically infer this as a case of underfitting (Villaverde et al. 2021). However, it must be noted that the relative distance between the estimated confidence interval threshold and the estimated global minimum is sensitive to the characterization of the error model. Therefore, we must be sure about the quantification of error used to compute the objective function to infer underfitting. Due to the lack of replicates in this study, I have used stand-in values for standard deviation (See Section 3.2.1). Therefore, we require a modified experimental design that allows for accurate quantification of error, as described above, before estimating the confidence interval threshold and inferring underfitting.

6.2 Impact and going forward

In this thesis, I have successfully demonstrated the application of dynamic model calibration to a non-linear PDE, a class of mathematical model less studied in the context of calibration of biological models. Using this pipeline, I showed we could infer avenues for concise model formulation and better experimental design. Such a cross-talk between theory and practice is crucial in building mathematical models that allow us to predict real-life dynamics. PDE models are computationally less expensive than their commonly used alternatives, such as ABMs, that help us understand the underlying spatially-explicit events as mechanistic causes of the observed spatiotemporal dynamics.

Experiments involving the growth of non-motile cells on agar surfaces are widely used in research in synthetic ecology due to their ability to study community dynamics in controlled environments. A fully calibrated continuum model can be helpful in efficiently designing these experiments by predicting outcomes in a few seconds or minutes. In addition, the generalized multispecies framework allows us to use this framework to study any synthetic community where the species interact through diffusible entities. In Chapter 5, I demonstrated this by applying the model to an antibiotic cross-protection mutualism.

However, the comparison was purely qualitative, and there is still a need for systematic calibration of the multispecies model.

Although under simplistic assumptions, this is one of the few macroscopic-scale models derived from microscopic rules in the context of bacterial populations. There is a pressing need for modelling from first principles to better relate parameters at the cell-level events to macroscopic parameters. Here, q and p describe the non-linearity in scaling the pressure exerted by growth in response to the cell density. These may be derived separately to relate to physical laws such as Newtonian mechanics or Hertzian theory. A long-term goal is also to use models for predictions in "real" environments. These are very complex (soils, gut microbiota, and such communities) but could eventually be described. Likewise, models at the spatial scale studied here can be integrated into more standard biofilm models.

Just as we incorporated spatial structure to observe realistic dynamics, more such factors are at play in microbial systems that are often discarded as contingencies, such as stochasticity and anisotropy. Incorporating these effects into continuum models and dynamic model calibration is an important step forward. Concerning model calibration, relying on a simple sum of squared differences for model fitting becomes increasingly difficult when we deal with more complex mathematical models. For spatial models, for example, we need error models that can capture the spatial and temporal autocorrelation in error. Stochastic models, like stochastic partial differential equations (SPDEs), can capture these effects inherently.

Appendix A

Derivation of the growth-induced dispersal model

Let us start with a summary of the biofilm model by Khassehkhan, Hillen, and Eberl (2009). Firstly, a 1-dimensional discrete space continuous time model is presented. Assume that space is 1-dimensional and discretized into equally spaced compartments that can be denoted by integers in ascending order from left to right. Let ρ_i denote the cell density at site i. Note that i should not be confused with the strain index used in the multispecies model (Equation 2.5) and Appendix B. Then,

$$\frac{d\rho_i}{dt} = \mathcal{T}^+_{i-1}\rho_{i-1} + \mathcal{T}^-_{i+1}\rho_{i+1} - (\mathcal{T}^+_i + \mathcal{T}^-_i)\rho_i + g(\rho_i)\rho_i \tag{A.1}$$

where $\mathcal{T}^+_i$ and $\mathcal{T}^-_i$ are the rates of flow of cells to the right and left respectively, at the site i, and $g(\rho_i)$ is the local growth rate which is assumed to only depend on the local cell density. In order to agree with Assumption 2.1.1, Khassehkhan, Hillen, and Eberl (2009) used the following definition of flow rates:

$$\mathcal{T}^\pm_i = \alpha q(\rho_i)p(\rho_{i\pm1}) \tag{A.2}$$

where q and p are density dependent functions that satisfy the properties listed in Assumption 2.1.2, and α is a constant with the dimensions $[T^{-1}]$ that depends on the spatial resolution of the discretized model. If Δx is the distance between two neighbouring site, α satisfies the following property:

$$\lim_{\Delta x \to 0} \alpha(\Delta x)^2 = \alpha_0 > 0$$

Assume that ρ_i converges to a function ρ continuous in space and time in the limit that Δx goes to zero. Substituting Equation A.2 in A.1 and taking the limit of $\Delta x \to 0$, we get

$$\frac{\partial \rho}{\partial t} = \alpha_0 \nabla \cdot \left(D(\rho) \nabla \rho \right) + g(\rho)\rho$$
$$D(\rho) = pq + \rho \left(p \frac{dq}{d\rho} - q \frac{dp}{d\rho} \right)$$

Let us return to the 1-dimensional discrete space continuous time model given by Equation A.1. Consider the new state variable S_i which denotes the nutrient concentration at site i. Henceforth, the local growth rate will depend on the nutrient concentration at the site, i.e. $g(S_i)$.

Rewrite the rate of flow as

$$\mathcal{T}_i^{\pm} = \delta g(S_i) q(\rho_i) p(\rho_{i\pm1}) \tag{A.3}$$

where δ is a dimensionless constant that depends on the spatial resolution of the discretized model such that it satisfies the following property:

$$\lim_{\Delta x \to 0} \delta (\Delta x)^2 = \delta_0 > 0$$

Using the definition of r in Section 2.2, we rewrite Equation A.1 as

$$\frac{d\rho_i}{dt} = \mathcal{T}_{i-1}^{+} \rho_{i-1} + \mathcal{T}_{i+1}^{-} \rho_{i+1} - (\mathcal{T}_i^{+} + \mathcal{T}_i^{-}) \rho_i + g(S_i) r(\rho_i) \rho_i \tag{A.4}$$

Let us use r_i to denote $r(\rho_i)$, and G_i to denote to total rate of cell growth at site i such that

$$G_i = G(\rho_i, S_i) = g(S_i)\rho_i$$

Substituting Equation A.3 in A.4 and rearranging the terms, we get

$$\begin{aligned}
\frac{d\rho_i}{dt} &= \delta(G_{i-1}q_{i-1} + G_{i+1}q_{1+1})p_i - \delta G_i q_i (p_{i-1} + p_{i+1}) + G_i r_i \\
&= \delta((G_{i-1}q_{i-1} + G_{i+1}q_{1+1})p_i - G_i q_i (p_{i-1} + p_{i+1}) + 2G_i q_i p_i - 2G_i q_i p_i) + G_i r_i \\
&= \delta(\Delta x)^2 \left(\frac{G_{i-1}q_{i-1} + G_{i+1}q_{1+1} - 2G_i q_i}{\Delta x^2} p_i - G_i q_i \frac{p_{i-1} + p_{i+1} - 2p_i}{\Delta x^2} \right) + G_i r_i
\end{aligned}$$

Assume that in the limit that Δx tends to zero, the discrete functions ρ_i and S_i converge to the functions ρ and S which are twice differentiable in space. Then, $g(S), r(\rho), q(\rho)$ and $p(\rho)$ are also twice differentiable in space if they are twice differentiable with respect to their dependent state variables. Furthermore, it is clear from the above equations that the transport of cells occur independently along each direction at the same speed. The net change in local cell density due to transport in two dimensions is then simply the sum of change in local density due to transport along each spatial axis.

Let $\nabla = \frac{\partial}{\partial x}\hat{i} + \frac{\partial}{\partial y}\hat{j}$ and $\nabla^2 = \nabla \cdot \nabla$, then

$$\lim_{\Delta x \to 0} \frac{d\rho_i}{dt} = \frac{\partial \rho}{\partial t} = \delta_0(p\nabla^2(Gq) - Gq\nabla^2 p) + Gr \tag{A.5}$$

Furthermore, let us say that the nutrient undergoes Fickian diffusion. It has been studied that packing of cells can reduce the diffusion constant of environmental metabolites (Dal Co et al. 2020). However, since we are dealing with agarose, it is safe to assume that there is sufficient cell-free region in the vertical direction to allow for fast diffusion along the two horizontal dimensions. As it is commonly done, I model nutrient depletion to be directly proportional to the growth rate of biomass, with a constant of proportionality which is the coefficient of converting unit mass of nutrient to unit biomass. Under these assumptions, I write the following PDE to describe the spatiotemporal dynamics of the nutrient concentration S:

$$\frac{\partial S}{\partial t} = \delta_s \nabla^2 S - \frac{Gr}{Y} \tag{A.6}$$

Appendix B

Derivation of the multispecies model

Let g_i be the growth rate per unit mass for strain k such that

$$g_k(S) = \mu_k \frac{S}{K_s + S}$$

where μ_i is its maximum growth rate. We assume different values for μ_k as the expression of certain genes may be costlier than some other and manifests itself in varying resource allotment for reproduction. Therefore, the total growth rate of each strain can be given by

$$G_k(\rho_k, S) = g_k(S)\rho_k$$

The total biomass growth rate is given by

$$G(\rho_1 ... \rho_N, S) = \sum_{k=1}^{N} G_k(\rho_k, S)$$

and the total growth pressure is given by the $Gq(\rho)$. Using our assumption that the cells of all strains are morphologically identical, the growth pressure, irrespective of the local composition, is uniformly distributed on all cells. The pressure experienced by cell biomass of a particular strain k is then simply the "partial growth pressure" given by

$$\frac{Gq\rho_k}{\rho}$$

Similarly, the penalty on growth rate r depends only on the total cell density and applies uniformly on all growth independent of the strain. Therefore, the net growth rate of strain k is given by $G_k r$.

Having found the growth rate and growth pressure on each strain, we can write the following PDE to describe the spatiotemporal dynamics of the constituent species

$$\frac{\partial \rho_k}{\partial t} = \delta_0 \left(p\nabla^2 \left(\frac{Gq\rho_k}{\rho} \right) - \frac{Gq\rho_k}{\rho} \nabla^2 p \right) + G_k r$$

It is easy to check that Equation 2.5 for all species $k = 1...N$ add up to Equation 2.3.

Appendix C

Derivation of likelihood function

C.1 Independent normally distributed error

Assume the error model where error in measuring every data point is drawn from independent normal distributions. Under the conditions of independent error distributions, the likelihood function is simply the product of likelihood of obtaining each data point when the parameter vector is given. For an ODE model, let $\mathcal{D}_i(t_k)$ be the value of the i^{th} observable measured at time t_k.

$$\mathbb{P}(\mathcal{D}|\theta) = \prod_i \prod_k \mathbb{P}(\mathcal{D}_i(t_k)|\theta)$$

For a normal distribution, the likelihood is given by

$$\mathbb{P}(\mathcal{D}_i(t_k)|\theta) = \frac{1}{\sqrt{2\pi\sigma_i^2(t_k)}} \exp\left(\frac{-(\mathcal{D}_i(t_k) - \mathcal{X}_i(t_k,\theta))^2}{2\sigma_i^2(t_k)}\right)$$

Substituting and taking the log-transformation, we obtain the following:

$$-2\ln\mathbb{P}(\mathcal{D}|\theta) = \sum_i \sum_k \left(\frac{(\mathcal{D}_i(t_k) - \mathcal{X}_i(t_k,\theta))^2}{\sigma_i^2(t_k)} + 2\pi\sigma_i^2(t_k)\right)$$

Maximization of the likelihood function is equivalent to minimization of the above function. Additionally, since optimization happens over different values of θ, all the terms in the sum of this function that is independent of θ remains a constant. Therefore, Maximum Likelihood Estimation is achieved by the minimization of the following objective function:

$$\mathcal{O}(\mathcal{D}|\theta) = -2\ln\mathbb{P}(\mathcal{D}|\theta) - \sum_i \sum_k 2\pi\sigma_i^2(t_k) = \sum_i \sum_k \frac{(\mathcal{D}_i(t_k) - \mathcal{X}_i(t_k,\theta))^2}{\sigma_i^2(t_k)}$$

C.2 Ad hoc likelihood function

Let $\mathcal{I}(\omega)$ be the value of cell density measured at the pixel ω (which indicates the spatial and temporal coordinates, i.e., $\omega = (x, y, t)$) in the Phase contrast time-lapse data. Let $\rho(\omega, \theta)$ be the predicted cell density for the parameter vector θ. For the given parameter vector θ, let us first fix the spatial coordinate (x, y). Let us look at the likelihood of obtaining the set of data points at every time point at this spacial point, denoted by $\{\mathcal{I}(\omega) \quad \forall\omega \in \Omega_{\mathsf{C}}(x, y)\}$ where $\Omega_{\mathsf{C}}(x, y)$ is the set of all pixels at the spatial coordinate (x, y) with presumably non-zero cell density (See Section 3.2.1). Under Assumption 3.2.1, I propose the following error model:

$$\mathcal{I}(\omega) = \rho(\omega, \hat{\theta}) + \epsilon(\omega) + \epsilon_{\mathsf{BG}}(x, y)$$

where

$$\epsilon(\omega) \sim \mathcal{N}(0, \sigma(\omega), \quad \epsilon_{\mathsf{BG}}(x, y) \sim \mathcal{N}(0, \sigma_{\mathsf{BG}})$$

Therefore,

$$\begin{aligned}\mathbb{P}(\mathcal{I}(\omega)|\theta) &= \mathbb{P}\Big(\epsilon(\omega) + \epsilon_{\mathsf{BG}}(x, y) = \mathcal{I}(\omega) - \rho(\omega, \theta)\Big) \\ &= \int_{-\infty}^{\infty} \mathbb{P}\Big(\epsilon_{\mathsf{BG}}(x, y) = z\Big)\mathbb{P}\Big(\epsilon(\omega) = \mathcal{I}(\omega) - \rho(\omega, \theta) - z\Big)dz\end{aligned}$$

For all the time points, we can then write

$$\begin{aligned}\mathbb{P}\{\mathcal{I}(\omega)\forall\omega \in \Omega_{\mathsf{C}}(x, y)|\theta\} &= \int_{-\infty}^{\infty} \mathbb{P}\Big(\epsilon_{\mathsf{BG}}(x, y) = z\Big)\mathbb{P}\Big(\bigcap_k \{\epsilon(\omega) = \mathcal{I}(\omega) - \rho(\omega, \theta) - z\}\Big)dz \\ &= \int_{-\infty}^{\infty} \mathbb{P}\Big(\epsilon_{\mathsf{BG}}(x, y) = z\Big) \prod_k \mathbb{P}\Big(\epsilon(\omega) = \mathcal{I}(\omega) - \rho(\omega, \theta) - z\Big)dz \\ &= C \int_{-\infty}^{\infty} \exp\left(\frac{-z^2}{2\sigma_{\mathsf{BG}}^2}\right) \prod_k \exp\left(\frac{-(\mathcal{I}(\omega) - \rho(\omega, \theta) - z)^2}{2\sigma^2(\omega)}\right)dz\end{aligned}$$

where C is the product of normalization constants of the normal distributions within the integral. Next, we use the standard result that the product of normal distributions is a normal distribution (Bromiley 2013). Let μ_{err} and σ_{err} denote the mean and standard deviation of this new normal distribution. Using this, we can rewrite the likelihood as

$$\mathbb{P}\{\mathcal{I}(\omega)\forall\omega \in \Omega_{\mathsf{C}}(x, y)|\theta\} = \frac{C(x, y)}{\sqrt{2\pi\sigma_{\mathsf{err}}^2}} \int_{-\infty}^{\infty} \exp\left(\frac{-(z - \mu_{\mathsf{err}})^2}{2\sigma_{\mathsf{err}}^2}\right)dz = C(x, y)$$

where

$$C(x,y) = \frac{1}{(2\pi)^{n/2}} \sqrt{\frac{\sigma_{\text{err}}^2}{\sigma_{\text{BG}}^2 \prod_{\omega \in \Omega_{\text{C}}(x,y)} \sigma^2(\omega)}} \exp\left[-\frac{1}{2} \left(\frac{\mu_{\text{BG}}^2}{\sigma_{\text{BG}}^2} + \sum_{\omega \in \Omega_{\text{C}}(x,y)} \frac{(\mathcal{I}(\omega) - \rho(\omega,\theta))^2}{\sigma^2(\omega)} - \frac{\mu_{\text{err}}^2}{\sigma_{\text{err}}^2} \right) \right] \tag{C.1}$$

$$\sigma_{\text{err}}(x,y) = \left(\frac{1}{\sigma_{\text{BG}}^2} + \sum_{\omega \in \Omega_{\text{C}}(x,y)} \frac{1}{\sigma^2(\omega)} \right)^{-1/2} \tag{C.2}$$

$$\mu_{\text{err}}(x,y,\theta) = \sigma_{\text{err}}^2(x,y) \sum_{\omega \in \Omega_{\text{C}}(x,y)} \left(\frac{\mathcal{I}(\omega) - \rho(\omega,\theta)}{\sigma^2(\omega)} \right) \tag{C.3}$$

According to our assumption that error at all spatial points follow independent distributions,

$$\begin{aligned} \mathbb{P}(\mathcal{I}|\theta) &= \prod_{(x,y)} \mathbb{P}\{\mathcal{I}(\omega) \forall \omega \in \Omega_{\text{C}}(x,y) | \theta\} = \prod_{(x,y)} C(x,y) \\ \implies -2 \ln \mathbb{P}(\mathcal{I}|\theta) &= \sum_{(x,y)} -2 \ln C(x,y) \end{aligned} \tag{C.4}$$

Taking the log-transform of Equation C.1, we get

$$-2 \ln C(x,y) = \sum_{\omega \in \Omega_{\text{C}}(x,y)} \left(\frac{(\mathcal{I}(\omega) - \rho(\omega,\theta))^2}{\sigma^2(\omega)} \right) - \frac{\mu_{\text{err}}^2(x,y,\theta)}{\sigma_{\text{err}}^2(x,y)} + \text{terms independent of } \theta$$

Substituting in Equation C.4, and ignoring all the terms independent of θ, we get the objective function

$$\mathcal{O}(\mathcal{I}|\theta) = \sum_{(x,y)} \left[\sum_{\omega \in \Omega_{\text{C}}(x,y)} \left(\frac{(\mathcal{I}(\omega) - \rho(\omega,\theta))^2}{\sigma^2(\omega)} \right) - \frac{\mu_{\text{err}}^2(x,y,\theta)}{\sigma_{\text{err}}^2(x,y)} \right]$$

such that the MLE ($\hat{\theta}$) satisfies

$$\mathcal{O}(\mathcal{I}|\hat{\theta}) = \min_{\theta} \mathcal{O}(\mathcal{I}|\theta)$$

Bibliography

Alamino, R. C. (2020). “An agent-based lattice model for the emergence of anti-microbial resistance”. *Journal of Theoretical Biology* 486.

Anderson, J. C. et al. (2006). “Environmentally controlled invasion of cancer cells by engineered bacteria”. *Journal of Molecular Biology* 355, 619–627.

Ashyraliyev, M. et al. (2009). “Systems biology: parameter estimation for biochemical models”. *The FEBS Journal* 276, 886–902.

Atkinson, A. C. and Donev, A. N. (1992). *Optimum Experimental Designs*. Oxford Statistical Science Series. Oxford, New York: Oxford University Press.

Balleza, E., Kim, J. M., and Cluzel, P. (2018). “Systematic characterization of maturation time of fluorescent proteins in living cells”. *Nature Methods* 15, 47–51.

Balsa-Canto, E., Alonso, A. A., and Banga, J. R. (2010). “An iterative identification procedure for dynamic modeling of biochemical networks”. *BMC Systems Biology* 4, 11.

Bär, J. et al. (2020). “Efficient microbial colony growth dynamics quantification with ColTapp, an automated image analysis application”. *Scientific Reports* 10.

Barker, G. C. and Grimson, M. J. (1994). “Scaling behaviour of an extended Eden model”. *Journal of Physics A: Mathematical and General* 27, 653.

Bengtsson-Palme, J. (2020). “Microbial model communities: To understand complexity, harness the power of simplicity”. *Computational and Structural Biotechnology Journal* 18, 3987–4001.

Bromiley, P. (2013). “Products and Convolutions of Gaussian Probability Density Functions”. In.

Byrne, H. and Drasdo, D. (2009). “Individual-based and continuum models of growing cell populations: a comparison”. *Journal of Mathematical Biology* 58, 657–687.

Cao, Y. et al. (2021). “Repulsive expansion dynamics in colony growth and gene expression”. *PLOS Computational Biology* 17.

Cardinale, B. J. (2011). “Biodiversity improves water quality through niche partitioning”. *Nature* 472, 86–89.

Cases, I. and Lorenzo, V. de (2005). “Genetically modified organisms for the environment: stories of success and failure and what we have learned from them”. *International Microbiology: The Official Journal of the Spanish Society for Microbiology* 8, 213–222.

Dal Co, A. et al. (2020). “Short-range interactions govern the dynamics and functions of microbial communities”. *Nature Ecology & Evolution* 4, 366–375.

Davis, B. D. (1987). “Mechanism of bactericidal action of aminoglycosides”. *Microbiological Reviews* 51, 341–350.

DeAngelis, D. L. and Yurek, S. (2017). “Spatially Explicit Modeling in Ecology: A Review”. *Ecosystems* 20, 284–300.

Deegan, R. D. et al. (1997). “Capillary flow as the cause of ring stains from dried liquid drops”. *Nature* 389, 827–829.

Dolinšek, J., Goldschmidt, F., and Johnson, D. R. (2016). “Synthetic microbial ecology and the dynamic interplay between microbial genotypes”. *FEMS Microbiology Reviews* 40. Ed. by J. R. van der Meer, 961–979.

Drepper, T. et al. (2010). “Flavin Mononucleotide-Based Fluorescent Reporter Proteins Outperform Green Fluorescent Protein-Like Proteins as Quantitative In Vivo Real-Time Reporters”. *Applied and Environmental Microbiology* 76, 5990–5994.

Eberl, H. J., Parker, D. F., and Van Loosdrecht, M. C. M. (2001). “A New Deterministic Spatio-Temporal Continuum Model for Biofilm Development”. *Computational and Mathematical Methods in Medicine* 3, 161–175.

Escalante, A. E. et al. (2015). “Ecological perspectives on synthetic biology: insights from microbial population biology”. *Frontiers in Microbiology* 6.

Farrell, F. D. C. et al. (2013). “Mechanically Driven Growth of Quasi-Two-Dimensional Microbial Colonies”. *Physical Review Letters* 111.

Feder, P. I. (1968). “On the Distribution of the Log Likelihood Ratio Test Statistic When the True Parameter is ”Near” the Boundaries of the Hypothesis Regions”. *The Annals of Mathematical Statistics* 39, 2044–2055.

Fujikawa, H. and Matsushita, M. (1991). “Bacterial Fractal Growth in the Concentration Field of Nutrient”. *Journal of the Physical Society of Japan* 60, 88–94.

Gorochowski, T. E. (2016). “Agent-based modelling in synthetic biology”. *Essays in Biochemistry* 60, 325–336.

Graf, F. E. et al. (2019). “Inhibiting conjugation as a tool in the fight against antibiotic resistance”. *Drug Development Research* 80, 19–23.

Greulich, P. et al. (2015). “Growth-dependent bacterial susceptibility to ribosome-targeting antibiotics”. *Molecular Systems Biology* 11, 796.

Grimson, M. J. and Barker, G. C. (1994). “Continuum model for the spatiotemporal growth of bacterial colonies”. *Physical Review E* 49, 1680–1684.

Hallatschek, O. et al. (2007). “Genetic drift at expanding frontiers promotes gene segregation”. *Proceedings of the National Academy of Sciences of the United States of America* 104, 19926–19930.

Harcombe, W. (2010). “Novel Cooperation Experimentally Evolved Between Species”. *Evolution* 64, 2166–2172.

Harvey, R. J. and Koch, A. L. (1980). “How partially inhibitory concentrations of chloramphenicol affect the growth of Escherichia coli.” *Antimicrobial Agents and Chemotherapy* 18, 323–337.

Hass, H. et al. (2019). “Benchmark problems for dynamic modeling of intracellular processes”. *Bioinformatics* 35, 3073–3082.

Hol, F. J. H. et al. (2015). “The idiosyncrasy of spatial structure in bacterial competition”. *BMC Research Notes* 8, 245.

Ingber, L. (1993). “Simulated annealing: Practice versus theory”. *Mathematical and Computer Modelling* 18, 29–57.

Jeong, S.-T. (2018). “Stability of Finite Difference Schemes on the Diffusion Equation with Discontinuous Coefficients”. In.

Kareiva, P. (1994). “Special Feature: Space: The Final Frontier for Ecological Theory”. *Ecology* 75, 1.

Khassehkhan, H., Hillen, T., and Eberl, H. J. (2009). “A Nonlinear Master Equation for a Degenerate Diffusion Model of Biofilm Growth”. In: *Computational Science – ICCS*

2009. Ed. by G. Allen et al. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, pp. 735–744.

Kirkpatrick, S. (1984). “Optimization by simulated annealing: Quantitative studies”. *Journal of Statistical Physics* 34, 975–986.

Klein, T. A., Ahmad, S., and Whitney, J. C. (2020). “Contact-Dependent Interbacterial Antagonism Mediated by Protein Secretion Machines”. *Trends in Microbiology* 28, 387–400.

Kolmogorov, A., Petrovskii, I., and Piscounov, N. (1937). “Study of a Diffusion Equation That Is Related to the Growth of a Quality of Matter and Its Application to a Biological Problem”. *Moscow University Mathematics Bulletin* 1, 1–26.

Kreutz, C. et al. (2013). “Profile likelihood in systems biology”. *The FEBS Journal* 280, 2564–2571.

Kuhn, D. et al. (2010). “Systems biotechnology – Rational whole-cell biocatalyst and bioprocess design”. *Engineering in Life Sciences* 10, 384–397.

Kwok, R. (2010). “Five hard truths for synthetic biology”. *Nature* 463, 288–290.

Leenheer, P. D., Schuster, M., and Smith, H. (2019). “Strong cooperation or tragedy of the commons in the chemostat”. *Mathematical Biosciences and Engineering* 16, 139–149.

Lötstedt, P. (2021). “Derivation of continuum models from discrete models of mechanical forces in cell populations”. *Journal of Mathematical Biology* 83, 75.

Martin, L. (1947). “Phase-contrast methods in microscopy”. *Nature* 159, 827–830.

Mattei, M. R. et al. (2018). “Continuum and discrete approach in modeling biofilm development and structure: a review”. *Journal of Mathematical Biology* 76, 945–1003.

Mimura, M., Sakaguchi, H., and Matsushita, M. (2000). “Reaction–diffusion modelling of bacterial colony patterns”. *Physica A: Statistical Mechanics and its Applications* 282, 283–303.

Momeni, B., Waite, A. J., and Shou, W. (2013). “Spatial self-organization favors heterotypic cooperation over cheating”. *eLife* 2. Ed. by D. Tautz.

Morehead, M. S. and Scarbrough, C. (2018). “Emergence of Global Antibiotic Resistance”. *Primary Care* 45, 467–484.

Müller, M. J. I. et al. (2014). “Genetic drift opposes mutualism during spatial population expansion”. *Proceedings of the National Academy of Sciences of the United States of America* 111, 1037–1042.

Murphy, S. A. and Van Der Vaart, A. W. (2000). “On Profile Likelihood”. *Journal of the American Statistical Association* 95, 449–465.

Murray, J. D. (2003). “Bacterial Patterns and Chemotaxis”. In: *Mathematical Biology: II: Spatial Models and Biomedical Applications*. Interdisciplinary Applied Mathematics. New York, NY: Springer, pp. 253–310.

Myung, I. J. (2003). “Tutorial on maximum likelihood estimation”. *Journal of Mathematical Psychology* 47, 90–100.

Naylor, J. et al. (2017). “Simbiotics: A Multiscale Integrative Platform for 3D Modeling of Bacterial Populations”. *ACS Synthetic Biology* 6, 1194–1210.

Peters, D. P. C. et al. (2004). “Strategies for Ecological Extrapolation”. *Oikos* 106, 627–636.

Pipe, L. Z. and Grimson, M. J. (2008). “Spatial-temporal modelling of bacterial colony growth on solid media”. *Molecular bioSystems* 4, 192–198.

Purnick, P. E. M. and Weiss, R. (2009). “The second wave of synthetic biology: from modules to systems”. *Nature Reviews Molecular Cell Biology* 10, 410–422.

Rahman, K. A., Sudarsan, R., and Eberl, H. J. (2015). “A Mixed-Culture Biofilm Model with Cross-Diffusion”. *Bulletin of Mathematical Biology* 77, 2086–2124.

Raue, A. et al. (2009). “Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood”. *Bioinformatics* 25, 1923–1929.

Regoes, R. R. et al. (2004). “Pharmacodynamic Functions: a Multiparameter Approach to the Design of Antibiotic Treatment Regimens”. *Antimicrobial Agents and Chemotherapy* 48, 3670–3676.

Sandle, T. (2010). “History and development of microbiological culture media”. *Institute of Science and Technology Journal*.

Savage, D. F., Way, J., and Silver, P. A. (2008). “Defossiling Fuel: How Synthetic Biology Can Transform Biofuel Production”. *ACS Chemical Biology* 3, 13–16.

Tan, J., Zuniga, C., and Zengler, K. (2015). “Unraveling interactions in microbial communities - from co-cultures to microbiomes”. *Journal of Microbiology (Seoul, Korea)* 53, 295–305.

Tronnolone, H. et al. (2018). “Diffusion-Limited Growth of Microbial Colonies”. *Scientific Reports* 8.

Van Dyken, J. D. et al. (2013). “Spatial population expansion promotes the evolution of cooperation in an experimental Prisoner’s Dilemma”. *Current biology* 23, 919–923.

Villaverde, A. F. et al. (2021). “A protocol for dynamic model calibration”. *Briefings in Bioinformatics* 23.

Wade, M. J. et al. (2016). “Perspectives in mathematical modelling for microbial ecology”. *Ecological Modelling* 321, 64–74.

Warren, M. R. et al. (2019). “Spatiotemporal establishment of dense bacterial colonies growing on hard agar”. *eLife* 8.

Wieland, F.-G. et al. (2021). “On structural and practical identifiability”. *Current Opinion in Systems Biology* 25, 60–69.

Xun, X. et al. (2013). “Parameter Estimation of Partial Differential Equation Models”. *Journal of the American Statistical Association* 108.

Yip, A. et al. (2022). “Calibrating spatiotemporal models of microbial communities to microscopy data: A review”. *PLOS Computational Biology* 18.

Young, J. W. et al. (2012). “Measuring single-cell gene expression dynamics in bacteria using fluorescence time-lapse microscopy”. *Nature Protocols* 7, 80–88.

Yurtsev, E. A., Conwill, A., and Gore, J. (2016). “Oscillatory dynamics in a bacterial cross-protection mutualism”. *Proceedings of the National Academy of Sciences* 113, 6236–6241.

Zomorrodi, A. R. and Segrè, D. (2016). “Synthetic Ecology of Microbes: Mathematical Models and Applications”. *Journal of Molecular Biology*. Engineering Tools and Prospects in Synthetic Biology 428, 837–861.