Modeling Collective Motion of Complex Systems using Agent-Based Models & Macroscopic Models

by

Sara Sami Jamous

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved July 2019 by the
Graduate Supervisory Committee:

Sebastien Motsch, Chair
Dieter Armbruster
Erika Camacho
Mohamed Moustaoui
Rodrigo Platte

ARIZONA STATE UNIVERSITY

August 2019

ABSTRACT

The main objective of mathematical modeling is to connect mathematics with other scientific fields. Developing predictable models help to understand the behavior of biological systems. By testing models, one can relate mathematics and real-world experiments. To validate predictions numerically, one has to compare them with experimental data sets. Mathematical modeling can be split into two groups: microscopic and macroscopic models. Microscopic models described the motion of so-called agents (e.g. cells, ants) that interact with their surrounding neighbors. The interactions among these agents form at a large scale some special structures such as flocking and swarming. One of the key questions is to relate the particular interactions among agents with the overall emerging structures. Macroscopic models are precisely designed to describe the evolution of such large structures. They are usually given as partial differential equations describing the time evolution of a density distribution (instead of tracking each individual agent). For instance, reaction-diffusion equations are used to model glioma cells and are being used to predict tumor growth. This dissertation aims at developing such a framework to better understand the complex behavior of foraging ants and glioma cells.

TABLE OF CONTENTS

CHAPTER Page

CHAPTER Page

LIST OF TABLES

LIST OF FIGURES

Figure Page

Figure Page

Figure Page

Figure Page

Figure Page

Figure Page

Figure Page

Figure Page

Chapter 1

INTRODUCTION

Mathematical modeling is essential to connect applied mathematics and other sciences such as biology, economy, technology, and engineering. Developing models help in learning and making decisions about real-life problems (Perkins and Simmons, 1988). By observing various experiences and behaviors, mathematical models allow predicting the actual world. From a statistical point of view, a model describes relationships between *dependent* and *independent* variables. However, in biology, the frontier between dependent and independent variables is often blurry. Testing predictive models is a way to clarify whether strong correlations imply strong interactions between variables (Hestenes, 2010).

To understand complex biological phenomena such as flocking and swarming, biologists often refer to *positive* and *negative* feed backs to model individual behaviors (Cipresso, 2015). Such assumptions lead mathematically to some *governing equations*. Solving these equations (analytically or numerically) lead to some testable predictions that can be compared to experimental data. We summarize the approach in figure 1.1.

The parameters used in the models should have some experimental meaning that will help to test some hypothesis and improve fitting the model with experimental data (Carrington *et al.*, 2005). Quantitative results can be compared by different observational assumptions that will identify the strength or weakness of the model (Judd *et al.*, 2011). If the predicted results of a model and the experimental data are close to each other, then the model can be considered a reasonable representation of the system.

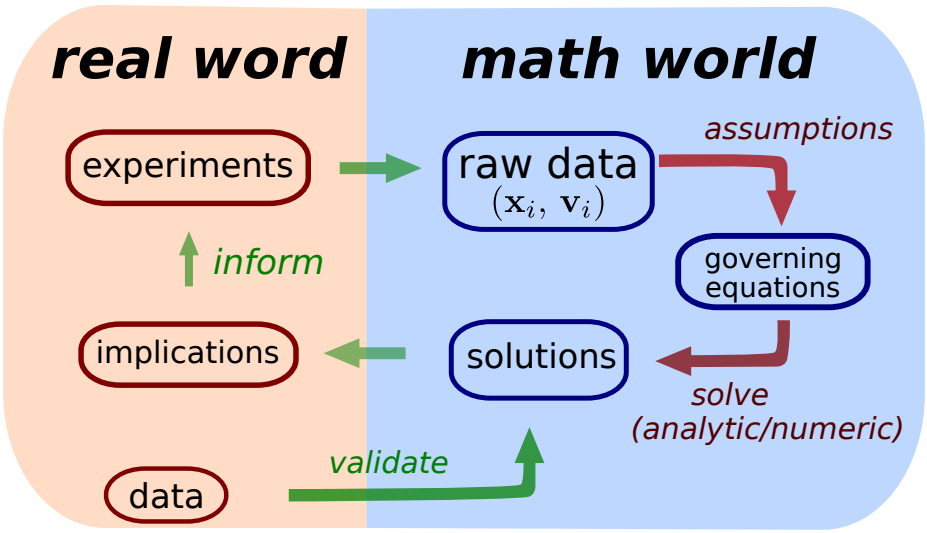


Figure 1.1: Schematic Representation of a Typical Mathematical Modeling Process

Mathematical models can be split into two groups: Agent-Based Models (ABMs) and Partial Differential Equations (PDEs). The split corresponds broadly to describe *microscopic* and *macroscopic* behavior. An agent-Based Model deals with the motion of each individual (referred here as *agent*). It is a (large) dynamical system which can be either deterministic or stochastic. Interactions between particles with their surrounding neighbors lead to change in their position and internal states (Barnes and Chu, 2015) (Politopoulos, 2007). One of the advantage of ABMs is to allow a larger spectrum of possibilities for modeling interaction among individuals. ABMs can reproduce pattern formation such as flocking or swarming. Often these biological abstractions permit researchers to analyze and explain these relationships. In this thesis, it is of particular interest that ABMs can be used to describe ants and cancer cells (chapter 2 and 3, resp.).

In chapter 2, we model ant foraging behavior process, i.e. ants moving back and forth between the nest and food source. We have been provided experimental data measuring the total flow of ants per second on a bridge where ants are crossing to

collect food. Investigating the flow, one can realize that the shape of the evolution curve increases to a maximum then start to decrease slowly over time. Therefore, our primary purpose is to create a predictive model that will fit this behavior. We develop a compartmental model to characterize the ant foraging dynamics. An ABM is introduced for ant foraging where each ant has an *internal state* describing whether it is in a *foraging state* or not. A loss function is used to estimate the best parameters needed for the model to fit the data. When we compare the experiments and the model, one can realize that, in the beginning the shape of the curves overlap, but over time they start to diverge. So, an extended model to the dynamics was introduced where the transition of ant foragers to non foragers is affected by the number of foragers. In this case, we can observe that experiments and model predictions are in better agreement, the predicted number of foraging ants almost overlaps the experimental data. Therefore, the extended model captures better the temporal evolution of the number of ant foragers.

Chapter 3 and 4 focus on glioblastoma which is the most aggressive and invasive malignant tumor. In chapter 3, glioma cells are responsive for brain tumor. A biological in vivo experimental observation shows different shapes of cancer cells (i.e circular and elliptical). Flock formation evolves only when cells are elongated. Such elongation can increase and influence the overall tumor growth. Therefore, to analyze how cancer cells manage to propagate throughout the organism system, we use ABMs to describe the essential features of the cell. In particular, we introduce eccentricity in the description of the cell. Adding this characteristic, we investigate how varying the density affect pattern formation. When the density is high, a new structure is noticeable, where cells move evenly in opposite direction to form a so-called stream.

In contrast to ABMs, Partial differential equations (PDEs) describe the time evolution of distribution. Instead of tracking individual particles (i.e cancer cells), we

are considering a *mass*. This macroscopic description is especially relevant when we consider tumor growth since the number of cancer cells can reach millions, it is often unpractical to keep track of the position of each cell. The macroscopic description is sometimes more helpful to observe and predict pattern formation. For instance, the investigation of reaction-diffusion equation is used to predict tumor growth.

In chapter 4, glioblastoma proliferates and spreads aggressively through the organism. The PDE mathematical model of reaction-diffusion was used to calculate the two coefficients of proliferation and invasion for patients. Some data of pre-treatment patients were shared with us, where they provided us the radius size of the tumor using different types of MRI taken on two separate days for a patient. By using the traveling wave solution and the steepness of the tumor (wave), one can estimate the proliferation and diffusion for each patient only if we are provided with two time points of imaging. Our main concern is to try to estimate these two parameters if we are only provided with one time point imaging, since in this case, we cannot evaluate the velocity. The hypothesis is to try to find a relationship (a strong correlation) between the velocity and the center size of the tumor (i.e necrosis) to be able to calculate these two parameters with one time imaging only.

Chapter 2

MODELING ANT FORAGING BEHAVIOR

2.1 Introduction

In nature, ants are able to optimize their foraging behavior in various environments from going back and forth between the nest and the food source. They regulate their foragers' activities based on some outgoing and returning neighbors that interact with them by positive and negative feedback (Robinson *et al.*, 2009; Prabhakar *et al.*, 2012; Mersch *et al.*, 2013). To study in more details this dynamics, Udiani et al. explain the collective of foraging ants which is regulated due to the environmental perturbation going from the nest to the food source and vice-versa (Udiani *et al.*, 2015). As shown in (Dussutour *et al.*, 2004), bi-directional traffic flow of ants going from the nest to the food source and vice-versa is influenced by several environmental factors. We would like to explore if our simple model can reproduce the key features observed experimentally.

As a source of inspiration, we use experimental data provided by Dussutour et al. (see figure 2.1). They have conducted several experiments recorded ants crossing a bridge to reach a food source. From these experiments, we observe that the flow of ants over time first increases rapidly and then slowly decays.

Our main motivation is to develop a mathematical model that will qualitatively predict the time evolution of the number of foragers as observed experimentally. We start by building a compartmental model inspired by epidemiology study (SIR model) It will contain three states inspired by the modeling of the diffusion of new products (Fibich *et al.*, 2012; Fibich, 2016).

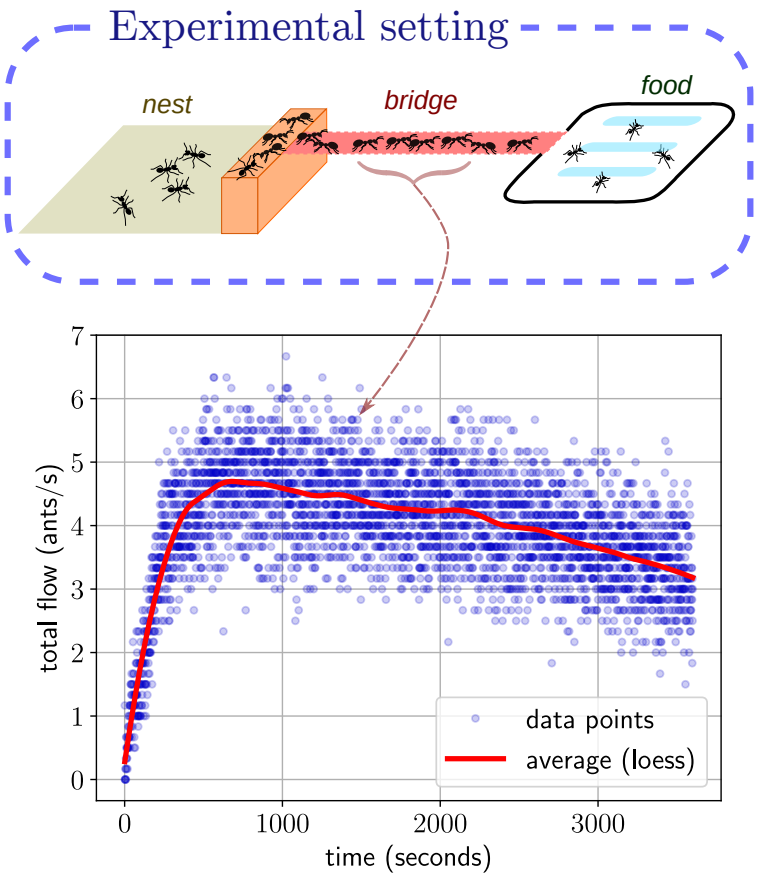


Figure 2.1: The Total Flow (in ants/second) of Ants Crossing the Bridge Back and Forth from the Nest to the Food Source as a Function of Time (in seconds). The Average of the Total Flow has been Plotted in Red by Using 'loess' Regression to Fit a Smooth Curve from the Data Points

First, we investigate the compartmental model analytically, denoted NFD model, and study its long-time behavior. We extend the model to incorporate a non-linearity necessarily to model the slow convergence toward equilibrium. The extended dynamics present some additional difficulties, in particular to characterize the stability of the equilibria. But data-model comparison using Metropolis-Hastings algorithm suggest that adding such non-linearity is necessarily to capture qualitatively the time evolution of the flow observed experimentally.

Then, we study a discrete version of the NFD model using Continuous-Time Markov Chains. We analyze the convergence of the discrete model to its continuous version and show the rate of convergence numerically. The discrete version of the model allows introducing spatial components in the dynamics. In other words, each ant could be described with a given location and velocity. We propose a simple model describing both the evolution of forager back and forth to a nest. Finally, we propose several methods to include congestion constraint in the dynamics.

2.2 Compartmental Model

We would like to model the evolution of the number of ant foragers over time. We do not take into consideration the position of the ants in this section. Thus, we mainly consider two quantities: the number of foragers and non-foragers. If one considers a first-order model, the evolution of the number of foragers could only be monotone. Indeed, we preserve the total number of ants, and therefore, there is only one degree of freedom for the dynamics (see remark below). For this reason, we introduce three states variables: an ant can be either Neutral, Foraging, or Deceptive. Thus, there are two states for a non-foraging ant (Neutral and Deceptive). Similar to a Susceptible SIR model (Hethcote, 2000), only Neutral ant can become forager whereas deceptive ant needs first to become neutral before being forager once again. Thus, the evolution of the behavior of an ant can be summarized in figure 2.2.

Remark 2.2.1 *If we model the evolution of total foragers using only foragers F and non-foragers N, a first-order model will give:*

$$F' = g(F, N)\,,\; N' = h(F, N)$$

where g and h are two functions to be determined. However, since the total number of ants denoted M must be preserved, we have: $N(t) = M - F(t)$ and therefore the

model reduces to:

$$F' = g(F, M - F).$$

Since the differential equation is autonomous, $F(t)$ will always be monotone for any choice of function g.

2.2.1 NFD Model

We denote by N, F, D the number of ants which are (resp.) Neutral, Foraging, and Deceptive. The total fixed number of ants is denoted M (i.e. $M = N + F + D$). We choose for the transition rates the following expression (see Fig. 2.2):

- Neutral to Foraging: $p + q\frac{F}{M}$ where $p > 0$ is a constant ('curiosity') and q a *positive feedback* ('active recruitment'), the more Foraging ants, the higher the rate,
- Foraging to Deceptive: constant rate $\lambda > 0$,
- Deceptive to Neutral: constant rate $r > 0$.

Notice that the transition rate from Neutral to Foraging depends on F, which implies a non-linearity. According to these rates, we define our first model.

Definition 1 [NFD model] *We consider the following dynamical system:*

$$N' = -(p + qF/M)N + rD \tag{2.2.1}$$

$$F' = (p + qF/M)N - \lambda F \tag{2.2.2}$$

$$D' = \lambda F - rD. \tag{2.2.3}$$

To study the dynamical system, we start by introducing the notations:

$$x = \frac{N}{M}\,,\ y = \frac{F}{M}\,,\ z = \frac{D}{M}.$$

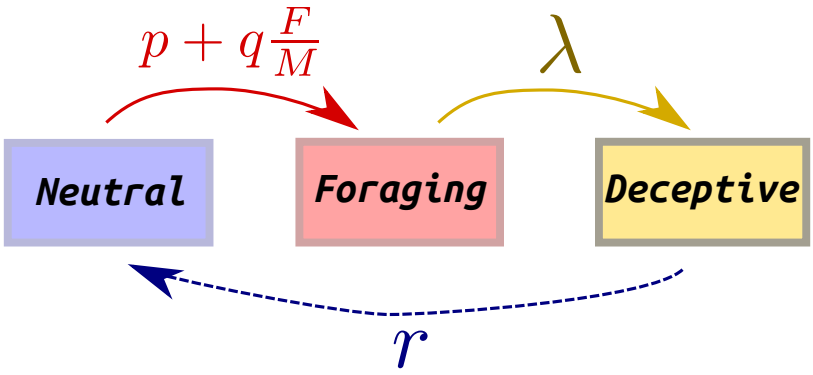


Figure 2.2: An Ant is in Three Possible States: Neutral, Foraging, or Deceptive. Over Time, each Ant Passes from Neutral to Foraging Behavior and then Deceptive, and the Circle Starts Again. Thus, the Model Requires Three Transition Rate

where the conservation of the mass gives $x + y + z = 1$. Thus, we can reduce the dynamical system (2.2.1)-(2.2.3) to only two equations (see Fig. 2.3):

$$x' = -(p + qy)x + r(1 - x - y) \tag{2.2.4}$$

$$y' = (p + qy)x - \lambda y \tag{2.2.5}$$

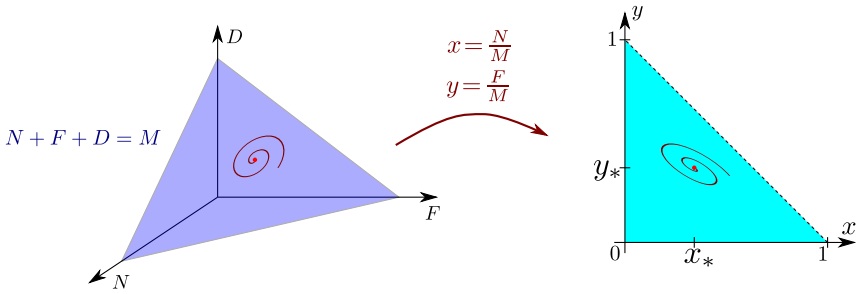


Figure 2.3: Using the Conservation of the Mass, we can Reduce the Dynamical System to a Two-Dimension Problem

Proposition 2.2.2 *Denote* $\Omega = \{(x, y) \in \mathbb{R}^2 | x \geq 0, y \geq 0, x + y \leq 1\}$. *The set* Ω *is invariant, in other words, for any* $(x_0, y_0) \in \Omega$, *the solution* $(x(t), y(t))$ *to* (2.2.4)-(2.2.5) *remains in* Ω *for all* $t \geq 0$.

Proof. We prove that the boundary of Ω forms a barrier for the dynamics (Hubbard and West, 2013), in other words any solution starting at the boundary of Ω will go towards the inside of Ω (see figure 2.3 - **right**). With this aim, we compute the scalar product of $\frac{d}{dt}(x(t), y(t))$ with the normal vector η of Ω. There are three cases to consider:

If $x = 0$ and $0 \leq y \leq 1$, then

$$\begin{pmatrix} x \\ y \end{pmatrix}' \cdot \begin{pmatrix} 1 \\ 0 \end{pmatrix} = x' = r(1-y) \geq 0.$$

If $y = 0$ and $0 \leq x \leq 1$, then

$$\begin{pmatrix} x \\ y \end{pmatrix}' \cdot \begin{pmatrix} 0 \\ 1 \end{pmatrix} = y' = px \geq 0.$$

If $x + y = 1$, $x \geq 0$ and $y \geq 0$, then

$$\begin{pmatrix} x \\ y \end{pmatrix}' \cdot \begin{pmatrix} -1/\sqrt{2} \\ -1/\sqrt{2} \end{pmatrix} = (-x' - y')/\sqrt{2} = \lambda y/\sqrt{2} \geq 0.$$

Hence, any point (x, y) starting inside Ω cannot escape out of it. Therefore, Ω is an invariant set. □

Proposition 2.2.3 *There exists a unique equilibrium (x^*, y^*) to the dynamical system* (2.2.1)-(2.2.3) *in Ω. Moreover, the equilibrium is asymptotically stable.*

Proof. Solving $y' = 0$ yields to $x^* = \frac{\lambda y^*}{p+qy^*}$. Then, $x' = 0$ leads to:

$$-(p + qy + r)\frac{\lambda y}{p + qy} \; + \; r(1-y) = 0 \quad \Rightarrow \quad P(y) = 0,$$

with $P(y) = q(\lambda + r)y^2 + (p\lambda + pr + \lambda r - qr)y - pr$. There exists two solutions y_1 and y_2 satisfying: $y_1 < 0$ and $0 < y_2 < 1$. Indeed, P is a parabola satisfying $P(0) < 0$ and $P(1) > 0$. Since $y > 0$, the equilibrium is given by: $y^* = y_2$.

To study the stability of the equilibrium, denote the dynamical system (2.2.4) (2.2.5) as:

$$\begin{pmatrix} x \\ y \end{pmatrix}' = F(x,y) \qquad \text{with} \quad F(x,y) = \begin{pmatrix} -(p+r+qy)x - ry + r \\ (p+qy)x - \lambda y \end{pmatrix}.$$

Taking the Jacobian of F gives:

$$DF(x,y) = \begin{bmatrix} -(p+r+qy) & -r-qx \\ p+qy & qx-\lambda \end{bmatrix}.$$

Notice that at equilibrium $qx^* = \frac{q\lambda y^*}{p+qy^*} < \lambda$, so $qx^* - \lambda < 0$.

Therefore Trace$(DF(x^*, y^*)) < 0$. Moreover, the determinant of $DF(x^*, y^*)$ is positive. We deduce that the equilibrium is necessarily stable (node or focus) in Ω. □

Following the proof, we have an explicit expression for the equilibrium. Introducing the coefficients:

$$\alpha = q(\lambda + r)\,, \qquad \beta = p\lambda + pr + \lambda r - qr\,, \qquad \gamma = pr,$$

we obtain the explicit expression:

$$F^* = M\,\frac{-\beta + \sqrt{\beta^2 + 4\alpha\gamma}}{2\alpha}, \qquad N^* = \frac{\lambda F^*}{p + qF^*/M}, \qquad D^* = \frac{\lambda F^*}{r}. \tag{2.2.6}$$

Figure 2.4 below is the solution of the model using Runge-Kutta method of order 4 using a initial condition $F_0 = D_0 = 0$ and $N_0 = M = 100$. We observe that the solution converges to a stationary state that corresponds to the formula obtained in (2.2.6).

2.2.2 Extended NFD Model

The convergence of the previous NFD model toward equilibrium is fast and does not match over time the shape of the data curve introduced in 2.1. Therefore, to slow

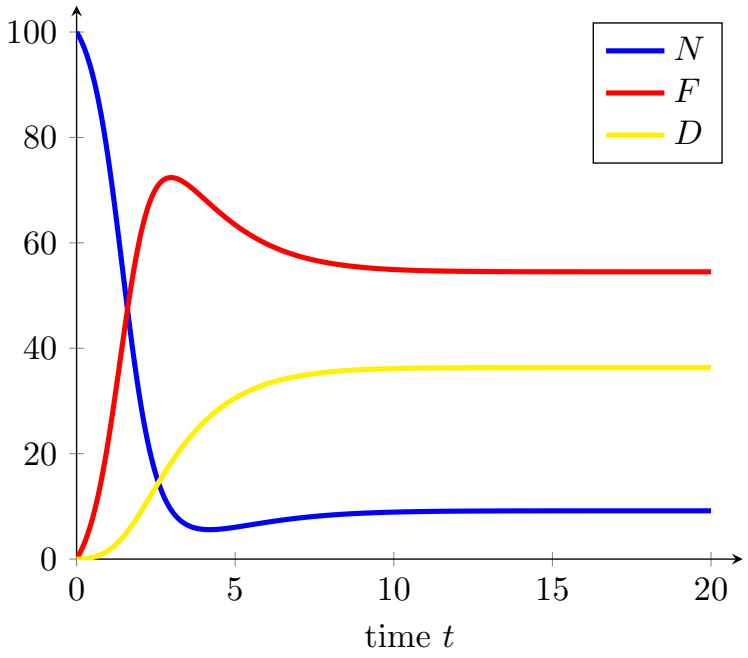


Figure 2.4: Numerical Solution (RK4) of the Dynamical System (2.2.1)-(2.2.3). The Solution Converges to an Equilibrium. Parameters of the Simulation: $\Delta t = 10^{-2}$, $p = 0.1$, $q = 2$, $\lambda = 0.2$, $r = 0.3$ and $M = 100$

down the convergence, we extend our NFD model. We start by exploring an extension of the previous dynamics where the transition rate from Foraging to Deceptive does also depend on the number of foragers: $\lambda(1 - \frac{F}{M})$ (i.e. the more Foraging ants, the lower the rate).

Definition 2 [Extended NFD model] *We consider the following dynamical system:*

$$N' = -(p + qF/M)N + rD \tag{2.2.7}$$

$$F' = (p + qF/M)N - \lambda F(1 - F/M) \tag{2.2.8}$$

$$D' = \lambda F(1 - F/M) - rD. \tag{2.2.9}$$

To study the equilibrium of the dynamics, we proceed as before and start by using

the conservation of mass (see Fig. 2.3) to reduce the dynamical system (2.2.7)-(2.2.9) to only two equations:

$$x' = -(p+qy)x + r(1-x-y) \tag{2.2.10}$$

$$y' = (p+qy)x - \lambda y(1-y) \tag{2.2.11}$$

Similarly to what we did in proposition 2.2.2, one can prove that Ω is also an invariant set for the extended dynamics. We start by considering the existence of equilibrium points based on some conditions.

Proposition 2.2.4 *Let $\lambda_c = \frac{r(p+q)}{(p+q+r)}$. If $\lambda < \lambda_c$, then there exists a unique equilibrium point to the dynamical system* (2.2.7)-(2.2.9) *denoted as (x_1^*, y_1^*) that is asymptotically stable and an attractor. Whereas, if $\lambda > \lambda_c$, (x_1^*, y_1^*) is unstable and there exists a second equilibrium point (x_2^*, y_2^*) in Ω.*

Proof. • To study the equilibrium points (x^*, y^*) of the two dimensional system (2.2.10) (2.2.11), solving $y' = 0$ in (2.2.11) gives

$$x^* = \frac{\lambda y^*(1-y^*)}{p+qy^*}. \tag{2.2.12}$$

Combining with $x' = 0$, we deduce from (2.2.10) the equation:

$$-(p+qy^*+r)\frac{\lambda y^*(1-y^*)}{p+qy^*} + r(1-y^*) = 0 \quad \Rightarrow \quad P(y^*) = 0.$$

We observe that $P(1) = 0$, so that $y = 1$ is a root of P. Indeed, a first equilibrium point is at $(x_1^*, y_1^*) = (0, 1)$. Therefore, we simplify P to get

$$P(y) = (y-1)Q(y) \qquad \text{where} \quad Q(y) = -q\lambda y^2 + (-p\lambda - r\lambda + rq)y + rp.$$

Note that $Q(0) = rp > 0$, thus there exists two root solutions y_1 and y_2 for Q satisfying: $y_1 < 0$ and $0 < y_2$. Since y is defined on $[0, 1]$ then we only consider

looking at the existence of y_2. Moreover, we have

$$Q(1) = r(p+q)\left(1 - \frac{\lambda}{\lambda_c}\right) \qquad \text{where} \quad \lambda_c = \frac{r(p+q)}{(p+q+r)}.$$

Thus, if $\lambda < \lambda_c$ we have that $Q(1) > 0$ and then $y_2 > 1$ is none reachable. Whereas, if $\lambda > \lambda_c$ we have that $Q(1) < 0$ so that $0 < y_2 < 1$. Thus, in this case, there exists a second equilibrium: (x_2^*, y_2^*) with x_2^* satisfying (2.2.12).

• To study the stability of the equilibrium points, denote the dynamical system (2.2.10) (2.2.11) as:

$$\begin{pmatrix} x \\ y \end{pmatrix}' = G(x,y) \qquad \text{with} \quad G(x,y) = \begin{pmatrix} -(p+qy)x + r(1-x-y) \\ (p+qy)x - \lambda y(1-y) \end{pmatrix}.$$

Taking the Jacobian of G gives:

$$DG(x,y) = \begin{bmatrix} -(p+qy+r) & -qx - r \\ p+qy & qx - \lambda + 2\lambda y \end{bmatrix}.$$

We start by looking at the stability of the equilibrium point $(x_1^*, y_1^*) = (0,1)$:

$$\begin{aligned} \text{Trace}(DG(x_1^*, y_1^*)) &= -(p+q+r) + \lambda, \\ \text{Det}(DG(x_1^*, y_1^*)) &= -\lambda(p+q+r) + r(p+q). \end{aligned}$$

For $\lambda < \lambda_c$, notice that $\lambda_c = \frac{r(p+q)}{(p+q+r)} < r$, so $\lambda < r$. Therefore the trace $\text{Trace}(DG(x_1^*, y_1^*)) < 0$. Moreover, the determinant $\text{Det}(DG(x_1^*, y_1^*))$ is positive. We deduce that the equilibrium (x_1^*, y_1^*) is necessarily stable. Now to specify if it is a node or focus, we look at

$$\text{Det} - \frac{\text{T}^2}{4} = -(p+q)\lambda - \frac{1}{4}[(p+q) - (\lambda + r)]^2 < 0$$

which proves that (x_1^*, y_1^*) is a local stable node when $\lambda < \lambda_c$.

When $\lambda > \lambda_c$, the determinant $\text{Det}(DG(x_1^*, y_1^*))$ is negative. Therefore, (x_1^*, y_1^*) is unstable saddle.

• It remains to show that in the case $\lambda < \lambda_c$, the equilibrium $(x_1^*, y_1*) = (0, 1)$ is an attractor. We use the Poincaré-Bendixson theorem (Arrowsmith and Place, 1992; Verhulst, 2006; Ciesielski, 2001) in dimension 2 knowing that Ω is an invariant compact set. Any solution has to converge to either an equilibrium or a periodic solution. Since there is only one equilibrium, we only have to rule out the convergence toward a periodic solution. But if we suppose that such periodic solution exists in Ω then it must contain an equilibrium point inside its orbit. However, there is only one equilibrium and it is located at the boundary of Ω, therefore no orbit inside Ω can contain it. □

Notice that we can not prove analytically the stability of the second equilibrium (x_2^*, y_2^*) when $\lambda > \lambda_c$. But numerically it is clear that the equilibrium is asymptotically stable as well as an attractor (see figure 2.5). In figure 2.6, we plot the bifurcation diagram of the equilibrium points x^* that satisfy equation (2.2.12) with respect to λ. We fix the values of the parameters, p, q and r same as in figure 2.4. One can observe that if $\lambda < \lambda_c$, only one solution exists and is stable at $x_1^* = 0$. Whereas, for $\lambda > \lambda_c$, we have existence of two equilibrium points, where x_1^* is unstable but x_2^* is stable. Although there is no simple expression for x_2^*, one can find the transition curve when $x_1^* = x_2^*$. Indeed, we can solve analytically:

$$\lambda = \lambda_c \quad \Rightarrow \quad (p+q)(\lambda - r) + r\lambda = 0$$

so that,

$$\frac{1}{4}(\lambda + r)^2 - \frac{1}{4}(\lambda - r)^2 + (p+q)(\lambda - r) = 0.$$

Therefore we get the equation of a hyperbola which explains the shape of the contour $\{x_* = 0\}$ in figure 2.6-right.

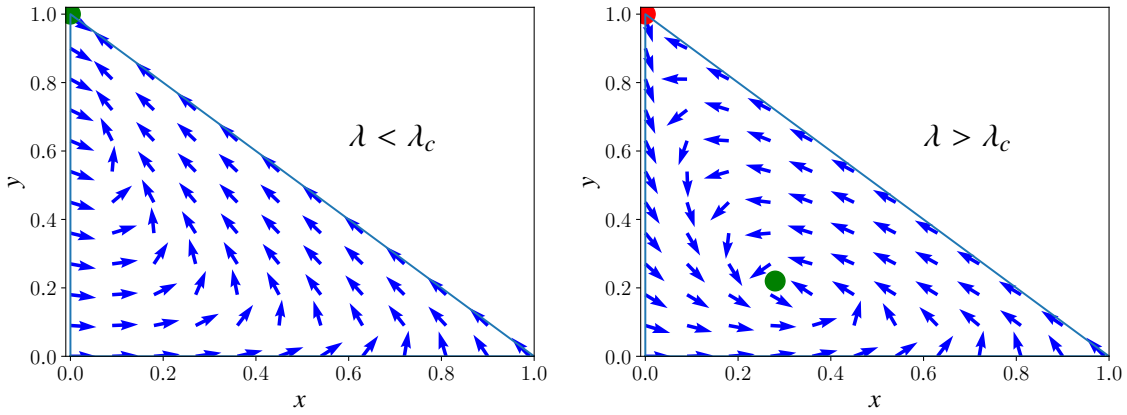


Figure 2.5: Phase Portrait for the Dynamics (2.2.10)- (2.2.11) Depending on λ and $\lambda_c = \frac{r(p+q)}{(p+q+r)}$. When $\lambda < \lambda_c$ (left), the Dynamics Converge Toward the Attractor $(x_1^*, y_1^*) = (0, 1)$. Whereas when $\lambda > \lambda_c$ (right), the Dynamics Converge Toward a Second Equilibrium (x_2^*, y_2^*) inside the Domain Ω

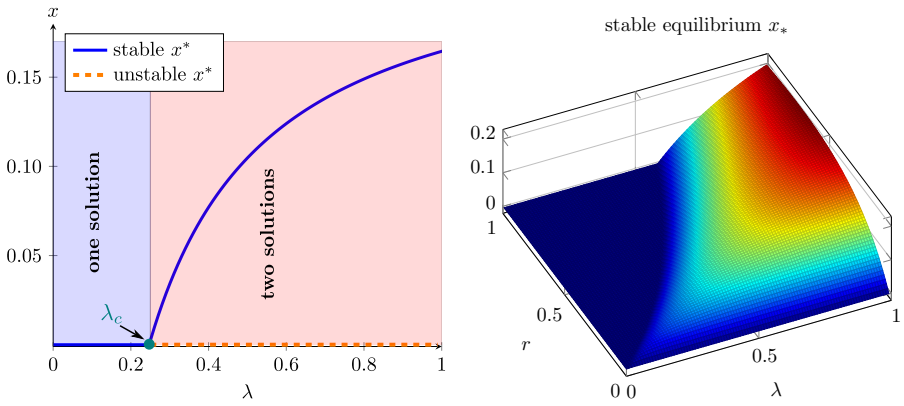


Figure 2.6: Left: The Bifurcation Diagram of the Equilibrium Points with Respect to λ, where the Parameters p, r and q are Fixed and Equal to the Values in Figure 2.4. When $\lambda \leq \lambda_c$, only One Solution Exists $x_1^* = 0$, and is Stable. Whereas, when $\lambda \geq \lambda_c$, we have Existence of Two Equilibrium Points, where x_1^* is Unstable but x_2^* is Stable. Right: The Numerical Observation of the Equilibrium Point x^*, Depending on λ and r

2.2.3 Data-Model Comparison

We would like to put to the test the two different models proposed by performing a qualitative comparison with the experimental data presented in figure 2.1. Unfortunately, we only have partial information about the number of foraging ants since we only have accessed to the flow measuring the amount of ants crossing the bridge per unit time. However, it is expected that the flow is, in a first approximation, proportional to the number of foraging ants (non-linear effects could make the relationship more complicated).

To perform a data-model comparison, we consider the target function to match the experimental average flow denoted $\mathcal{F}_{exp}$ (see the red curve in figure 2.1). We would like to compare $\mathcal{F}_{exp}(t)$ with the prediction of the amount of foragers $F(t)$ in our models. We have to normalize each curve such that their maximum is 1 (otherwise $\mathcal{F}_{exp}$ and $F(t)$ have different units). We use as a loss function ℓ the mean-square error:

$$\ell = \frac{1}{\mathrm{T}} \int_0^{\mathrm{T}} \left| \frac{\mathcal{F}(t)}{\|\mathcal{F}\|_\infty} - \frac{F(t)}{\|F\|_\infty} \right|^2 dt, \tag{2.2.13}$$

where $\|f\|_\infty = \max_t |f(t)|$ and T is the total time of the experiment.

To minimize the loss function ℓ with respect to the four parameters $\theta = (p, q, r, \lambda)$, we use the MetropolisâĂŞHastings (MH) Algorithm (Chib and Greenberg, 1995). This algorithm is slow compared to the approach of the traditional gradient-descent but it does not require to compare the gradient of the loss function ℓ. Starting from a given value for the parameters denoted $\theta_n = (p_n, q_n, r_n, \lambda_n)$, we perturb their values by adding a random variable: $\theta_* = \theta_n + \varepsilon_n$. We then keep the updated parameters (i.e. $\theta_{n+1} = \theta_*$) depending on its performance $\ell(\theta_*)$ by comparing it to $\ell(\theta_n)$ (see Algorithm 1). Note that the losses $\ell(\theta_n)$ could eventually increase with n which allows the algorithm to not get stuck into local minimum. There are two hyper-

parameters for the MH algorithm: the intensity of the noise applied to θ_n (denoted σ) and the 'temperature' (denoted D) governing the acceptance rate when the loss increases (taking $D = 0$ enforces a decay of the loss function).

Algorithm 1 Metropolis-Hasting

procedure METROPOLIS($\theta_n, \ell_n, \sigma, D$)
    $\theta_* = \theta_n + \sigma\varepsilon_n$
    $\ell_* = \ell(\theta_*)$
    if $\ell_* < \ell_n$ **then**
        Accepted: $\theta_{n+1} = \theta_*, \ell_{n+1} = \ell_*$
    else
        $U \sim \mathcal{U}[0, 1]$
        if $U < \exp(-(\ell_* - \ell_n)/D)$ **then**
            Accepted: $\theta_{n+1} = \theta_*, \ell_{n+1} = \ell_*$
        else
            Rejected: $\theta_{n+1} = \theta_n, \ell_{n+1} = \ell_n$
        end if
    end if
    Return θ_{n+1}, ℓ_{n+1}
end procedure

We apply the MH algorithm to our dynamical systems (2.2.1)-(2.2.3) and (2.2.7)-(2.2.9) starting with the same initial guess for the parameters. In figure 2.7-left, one can observe that the loss function of the original model is minimized at around $\ell \approx 4$ whereas for the extended model the minimum is close to $\ell \approx 1$. Plugging the parameters θ_* that minimize the loss ℓ (resp. $\theta_{NFD} = (.12, .22, .045, .021)$ and $\theta_{ext.NFD} = (.16, .30, 0.022, 0.059)$), we then plot the evolution of the curves $F(t)$

over time for the two models and compare them with the (normalized) total flux in figure 2.7-right. We notice that the NFD model converges quickly to an equilibrium, whereas the extended NFD is able to capture the slow decay of the number of foragers.

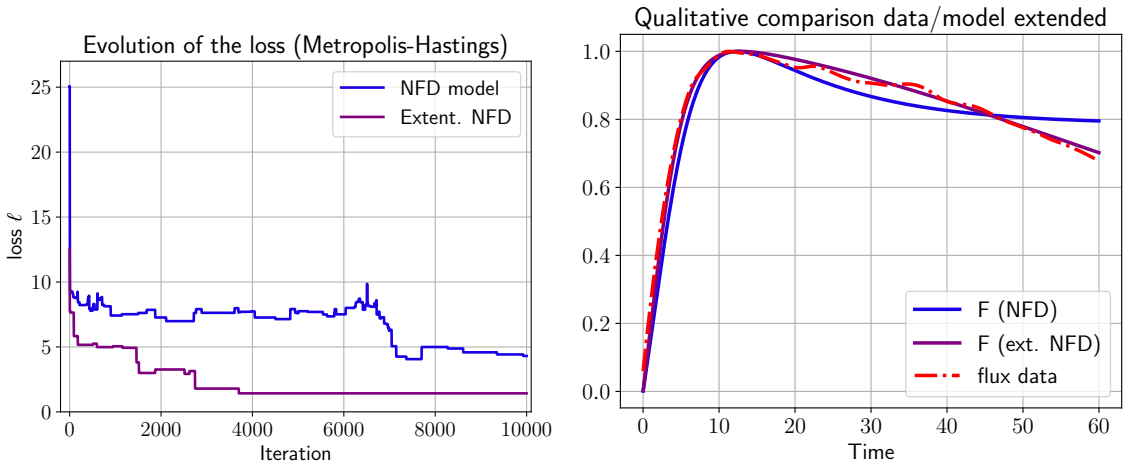


Figure 2.7: Left: Evolution of the loss function over the number of iterations by using Metropolis-Hastings for the original dynamical system (2.2.1)-(2.2.3) and the extended (2.2.7)-(2.2.9). The value that minimizes the loss function is lower in the extended model than the original model. Right: The corresponding curves $F(t)$ for both models using as parameters the minimizers of the loss function ℓ. We observe that the extended dynamics is able to capture the slow decay of the number of foragers

2.3 Agent-Based Models

2.3.1 Continuous-Time Markov Chains

We would like to describe our previous dynamics as a *discrete* number of ants by using a Continuous-Time Markov Chains (CTMC) model. The goal of this section is to show that we can recover similar dynamics as the NFD dynamics (2.2.1)-(2.2.3) in this setting. In particular, as the number of ants M growths, the discrete description will converge toward the continuous model.

We describe each ant with a state variable S_i that can take 3 values: $S_i \in \{1, 2, 3\}$ representing (resp.) Neutral, Foraging and Deceptive. Thus, the quantity N, F, D are given by:

$$N(t) = \#\{S_i(t) = 1\}, \quad F(t) = \#\{S_i(t) = 2\}, \quad D(t) = \#\{S_i(t) = 3\}, \tag{2.3.14}$$

where $\#$ represents the cardinal of a set. Notice that now N, F, and D can only take discrete values. As for the ODE system, the evolution of S_i over time depends on the rates at which it changes from 1 to 2 (i.e. Neutral to Foraging), from 2 to 3 etc...We use the same rates as for the ODE system (see Fig. 2.2). Mathematically, $S_i(t)$ is a (jump) Markov process with generator (Norris, 1998):

$$G(t) = \begin{bmatrix} -(p + q\frac{F(t)}{M}) & (p + q\frac{F(t)}{M}) & 0 \\ 0 & -\lambda & \lambda \\ r & 0 & -r \end{bmatrix} \tag{2.3.15}$$

where $F(t)$ is given by (2.3.14). For instance, if $S_i(0) = 2$, then S_i will jump to 3 after a time T given by an exponential law $T \sim \mathcal{E}(\lambda)$ (see Fig. 2.8).

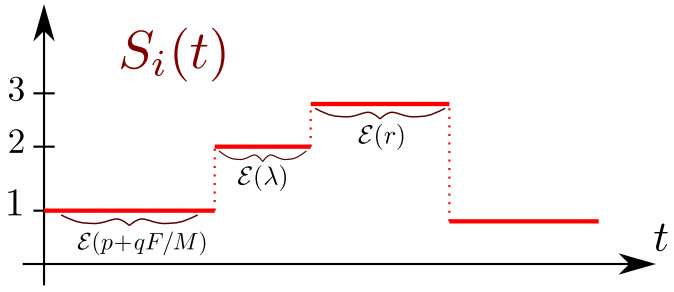


Figure 2.8: Illustration of the Poisson Process $S_i(t)$ with Transition Rates Given by Figure 2.2

Remark 2.3.1 *We have only described the evolution of a single ant i looking at the stochastic process $S_i(t)$. To fully describe the system, we would need to introduce*

the vector of all ants:

$$S(t) = \begin{bmatrix} S_1(t) \\ S_2(t) \\ \vdots \\ S_M(t) \end{bmatrix}. \tag{2.3.16}$$

One could write a generator for this vector.

Numerically, we approximate the jump process by a discrete Markov chain (Norris, 1998). Introducing a small time step Δt and supposing $S_i(0) = 1$. The probability P of the jump of S_i after a time Δt is approximated by:

$$P(S_i \text{ " jumps from 1 to 2 over "} [0, \Delta t]) \approx 1 - \exp(-(p + F/M)\Delta t),$$

or in other words

$$P(S_i(\Delta t) = 2 \,|\, S_i(0) = 1) \approx 1 - \exp(-(p + F/M)\Delta t).$$

Similarly, we have:

$$\begin{aligned} P(S_i(\Delta t) = 3 \,|\, S_i(0) = 2) &\approx 1 - \exp(-\lambda \Delta t), \\ P(S_i(\Delta t) = 1 \,|\, S_i(0) = 3) &\approx 1 - \exp(-r \Delta t). \end{aligned}$$

Note that the smaller Δt is, the smaller the probability will be to jump. Moreover, it is only an approximation as the stochastic process $S_i(t)$ could jump twice over a period $[0, \Delta t]$. However, the probability that two jumps occurs on such time interval is of the order $\mathcal{O}(\Delta t^2)$.

To illustrate the Markov process, we run simulations using the same parameters as in figure 2.4. In figure 2.9-left, we run one simulation with $M = 100$ ants and plot the evolution of the number of Neutral, Foraging and Defective ants. In contrast with the ODE system (2.2.1)-(2.2.3), the curves fluctuate due to the finite number of ants

M. To reduce these fluctuations, we perform 50 simulations in fig. 2.9-right and plot the average, denoted $\langle N\rangle$, $\langle F\rangle$ and $\langle D\rangle$. The width along the curves corresponds to the standard deviation. The results are now in good agreement with the solution of the ODE system (see Fig. 2.4).

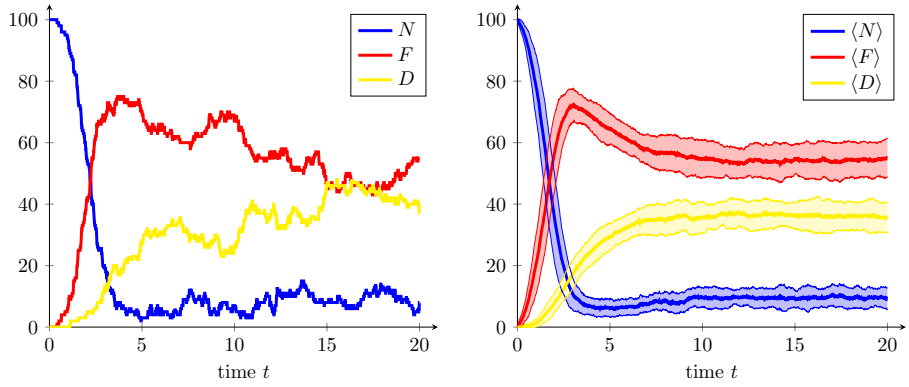


Figure 2.9: Left: The Evolution of N, F, D (2.3.14) from the Simulation of the Jump Process with Generator (2.3.15) for $M = 100$ Ants and Similar Parameters as in Fig. 2.4. Right: Averaging of 50 Simulations Along with the Standard Deviation

As we observe in figure 2.9, the evolution of N, F and D has qualitatively the same behavior as the ODE system (see fig. 2.4). We would like to analyze furthermore the agreement and discrepancy between the two dynamics, one being deterministic (ODE) and the other one stochastic (Markov process).

We notice that the ODE system has a scaling property. If $(N(t), F(t), D(t))$ is a solution to (2.2.1)-(2.2.3) with a total of M ants, then for any scalar $\alpha > 0$, then $(\alpha N(t), \alpha F(t), \alpha D(t))$ is the solution with αM ants, since:

$$(\alpha F)' = \alpha F' = \alpha\Big((p + qF/M)N - \lambda F\Big) = (p + \alpha qF/\alpha M)\alpha N - \lambda\alpha F.$$

We can proceed similarly for $N(t)$ and $D(t)$. As a result, varying M has no effects on the proportions $N(t)/M$, $F(t)/M$ and $D(t)/M$. However, increasing M will have an impact on the stochastic process.

To analyze how the total number of ants M impacts on the stochastic process, we run several simulations with various M. In figure 2.10-left, we plot the evolution of $F(t)$ for four values of M. We observe that the fluctuation reduces as M increases and the curves converge toward the solution of the ODE. To investigate further this convergence, we measure the average error between the stochastic and deterministic solution:

$$\text{Error}_F = \frac{1}{T}\left(\int_0^T |F(t) - F_{stoc}(t)|^2 \, dt\right)^{\frac{1}{2}}. \tag{2.3.17}$$

We introduce similar error function for N and D and denote the total error:

$$\text{Error} = \text{Error}_N + \text{Error}_F + \text{Error}_D. \tag{2.3.18}$$

In figure 2.10-right, we plot the total error depending on M. We observe that the decay is of the order $\mathcal{O}(1/\sqrt{M})$ as one could expect from the Central Limit theorem (Darling *et al.*, 2008).

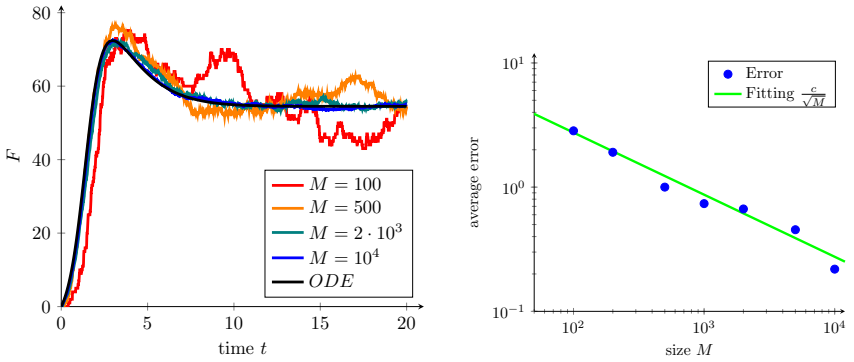


Figure 2.10: Left: Number of Foraging Ants Varying M. Fluctuation Reduce as the Total Number M Increases. Right: The Average Error Between the ODE and the Stochastic Process (2.3.18) Depending on M. The Decay is of the Order of $\mathcal{O}(1/\sqrt{M})$

2.3.2 *Spatial Model*

We would like to add spatial position into the model. Each ant i has now a position vector $\mathbf{x}_i(t) \in \mathbb{R}^2$. Denote by A the position of the nest, and B the position of the food source. Foraging ants are going back and forth between A and B at a constant speed c, thus:

$$\mathbf{x}_i' = c\,\omega(\mathbf{x}_i) + \sigma \dot{W}_i, \tag{2.3.19}$$

where σ is the noise intensity, W_i a Brownian motion, and ω_i is the direction of either nest or the source:

$$\omega_i = \frac{A - \mathbf{x}_i}{|A - \mathbf{x}_i|} \quad \text{or} \quad \omega_i = \frac{B - \mathbf{x}_i}{|B - \mathbf{x}_i|}$$

Each time the ant reaches its target, nest or food source (resp. A and B), the target direction is updated (A becomes B and vice-versa). Numerically, an ant reaches its target if it is at a distance less than a fixed radius R from the target. In the case where an ant is Neutral or Deceptive, the motion is simply a diffusive process:

$$\mathbf{x}_i' = \sigma \dot{W}_i, \tag{2.3.20}$$

there is no more incentive to go to a specific target. A schematic representation of the model is given below in figure 2.11.

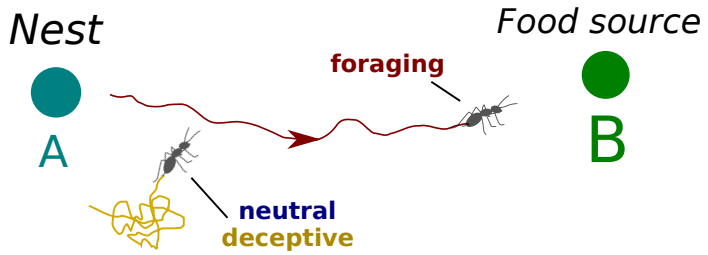


Figure 2.11: Schematic Representation of the Spatial Model of Ants. Foraging Ants (in Red) are Moving Back and Forth Between the Nest and the Food Source. Neutral and Deceptive are Moving Randomly

To investigate numerically the model, we look at the simulations at two different time (see figure 2.12). The parameters p, q, r and λ are as previously, whereas $c = 10$ and $\sigma = 0.2$. The nest is located at the origin $A = (0, 0)$ and the food source is at the right side at $B = (50, 0)$. The radius of both the nest and food source is estimated as $R = 1$. In figure 2.12, initially, the ants are all distributed in the nest. They start to spread slowly (see fig.2.12-left). As time evolves, ants will spread first toward the food source (x-direction) but also in the y-direction due the diffusion process (see fig.2.12-right).

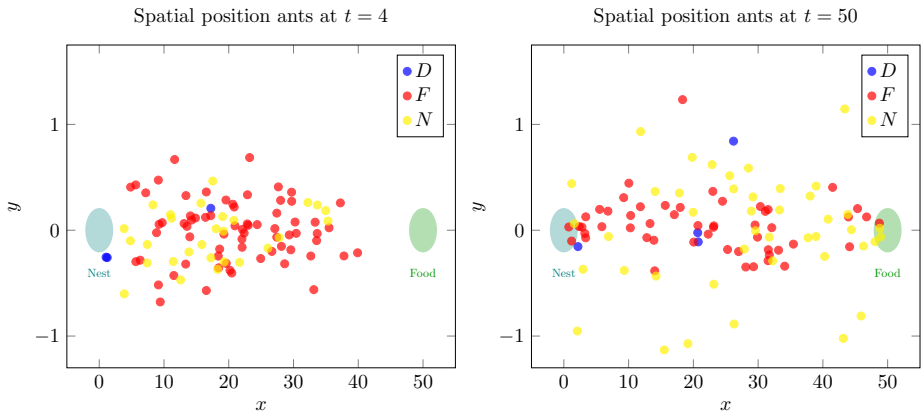


Figure 2.12: Simulation of the Model at two Different Time. Initially, the Ants are all Distributed in the Nest. As Time Evolves, Ants will Spread First Toward the Food Source (x-direction) but also in the y-Direction due the Diffusion Process

For local interaction between ants, we use the evolution strategy S_i of an ant i as defined previously in the Poisson process (see equation 2.3.14). We determine at each time step f_i the number of foraging , and $\aleph_i$ the total number of neighbors *nearby* ant $\mathbf{x}_i$ where $\aleph_i = n_i + f_i + d_i$. We count the neighbors f_i and $\aleph_i$ by looking at the strategy of the ants that are within the radius of interaction d of ant $\mathbf{x}_i$. Note that, by counting f_i and $\aleph_i$, one can observe that the number of its neighborhood includes

itself (see figure 2.15-**left**).

$$f_i(t) = \#\{j \mid S_j(t) = 2 \text{ and } |x_j - x_i| \leq d\}, \tag{2.3.21}$$

$$\aleph_i(t) = \#\{j \mid S_j(t) \in \{1,2,3\} \text{ and } |x_j - x_i| \leq d\}. \tag{2.3.22}$$

Remark 2.3.2 *We notice that there is a relationship between d and q. When d is very small (resp. very large), the average of evolution of N, F, D over 50 simulation of the local interaction will be similar to the standard dynamical ODE when $q = 0$ (resp. $q = 2$). The values of the other parameters used are similar to figure 2.4, the noise and speed are taken as $\sigma = 0.6$ and $c = 5cm/s$ (resp.) (see figures 2.13 - 2.14). When $d = 0$, no neighbors are around so that $f_i = \aleph_i = 0$ (i.e $qf_i = 0$ by looking at the local model), therefore the foragers will not interact with each others so that the curiosity coefficient $q = 0$. On the other hand, when d is large, local neighbors foragers are going to be equal to the entire followers $f_i = F$, therefore, $q = 2$ in this case like the original model.*

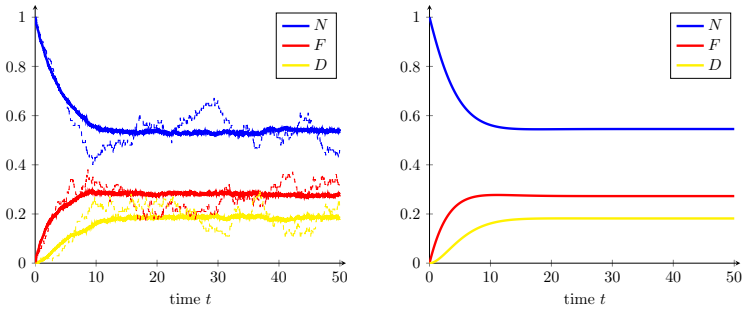


Figure 2.13: Left: The Average of Evolution of N, F, D of 50 Simulation of the Local Interaction with $d = 0.1$. Right: The Evolution of N, F, D of the Dynamical System with $q = 0$

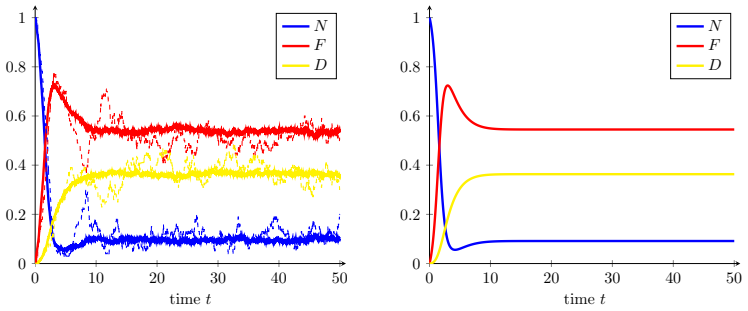


Figure 2.14: Left: The Average of Evolution of N, F, D of 50 Simulation of the Local Interaction with $d = 10$. Right: The Evolution of N, F, D of the Dynamical System with $q = 2$

2.3.3 Congestion Effect

Finally, we would like to discuss how to implement a congestion effect on the dynamics. When the density is high, ants have to slow down, i.e. the speed c is a (decaying) function of the density. Therefore, we propose two ways to incorporate congestion constraint.

A first possibility is to have the speed of each ant c_i depends on $\aleph_i$ the number of neighbors around it within the distance d, so that

$$c_i = c(\aleph_i), \tag{2.3.23}$$

where $c(\cdot)$ is a decaying function (see figure 2.15). The function is maximum when there is no neighborhood around it (i.e $\aleph_i = 1$).

Another possibility is to incorporate stopping time (Göttlich *et al.*, 2017): ants stop once they *hit* each other. Thus, the speed c has to take two values: 0 when it stops and $\bar{c}$ when it moves. This behavior can be modeled introducing a stochastic process Z_i with two possible values 0 and 1, meaning that the ant is (resp.) at rest or moving. We assume that $\bar{c}$ is 5 cm/s, so that the actual speed of an ant is then

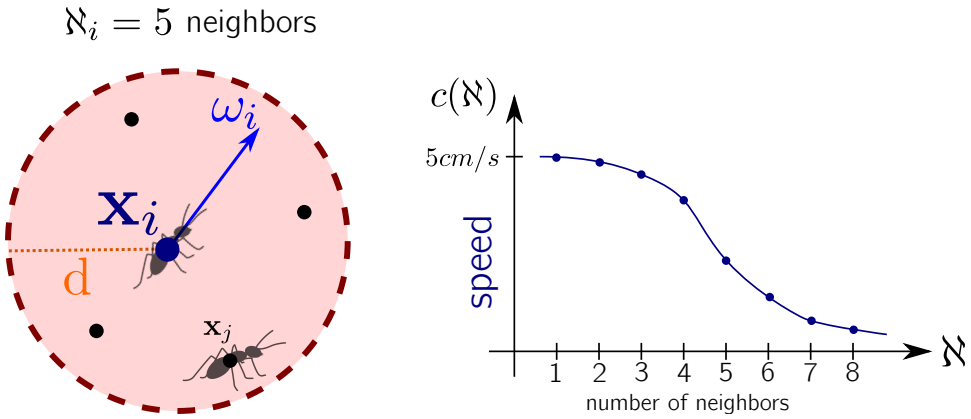


Figure 2.15: Left: The Neighbors $\aleph_i = 5$ of Ant $\mathbf{x}_i$ within the Radius of Interaction d. The Number of Neighborhood around $\mathbf{x}_i$ Includes Itself. Right: The Speed of an Ant with Respect to the Number of Ants in its Neighborhood

given by:

$$c_i = \bar{c} \cdot Z_i. \tag{2.3.24}$$

The evolution of the process Z_i depicted in figure 2.16-left will depend on two rates: $\lambda_{1\to 0}$ and $\lambda_{0\to 1}$ describing (resp.) the rate at which an ant stop and re-start. These rates must also depend on the number of neighbors where $\lambda_{1\to 0}$ and $\lambda_{0\to 1}$ should be (resp.) increasing and decreasing functions of $\aleph_i$ (figure 2.16 -right).

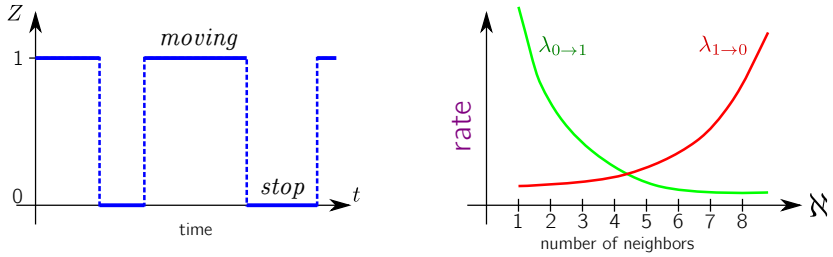


Figure 2.16: Left: The Stochastic Process Z Encodes when an Ant is Moving (i.e. $Z = 1$ Means Free Ride) or Resting (i.e. $Z = 0$). Right: The Rates $\lambda_{0\to 1}$ (Ant is at Rest then Moves) and $\lambda_{1\to 0}$ (Ant Moves then Stops) Depend on the Number of Ants $\aleph$ Nearby

2.4 Conclusion

We performed a data-model comparison using experimental data collected by Dussutour et al. At first we started by observing the experimental behavior of ants going back and forth from the nest to the food source through a bridge. Our main intention was to build a model that could fit the data of the average curves of the total flow over time. Then, we tried to estimate the parameters needed by minimizing a loss function between the model and the data. The results suggest that the extended model provides a much better fit to the data curve than the original model. We then used the mean field approximation to show that the dynamics could be approximated by a Continuous Time Markov Chain and investigate numerically how increasing the number of ants reduces the fluctuation.

To further improve the comparison between model and data, we would like to use more experimental information such as the number of ants moving from the nest to the food source and vice-versa. Such information will provide the number of foragers at the food source and therefore it would be possible to have a more accurate comparison between the model and the experiments.

Another challenge will be to incorporate the ant spatial positions into the data-model comparison. However, it is experimentally extremely time consuming to track each ant, although modern automatic tracking techniques could alleviate this work (as long as the density of ants is not too large).

2.5 Appendix

We explain how to simulate a stochastic process $S(t)$ with the following generator (Norris, 1998):

$$G(t) = \begin{bmatrix} -a(t) & a(t) & 0 \\ 0 & -b(t) & b(t) \\ c(t) & 0 & -c(t) \end{bmatrix} \tag{2.5.1}$$

where a, b, c are positive functions.

- For each time step Δt,
  - estimate the probability p that S jumps over the time interval Δt:

$$p = 1 - \exp(G(S,S)\Delta t),$$

  - with probability p

$$S \longrightarrow \begin{cases} S+1 & \text{if } S = 1 \text{ or } S = 2 \\ 1 & \text{if } S = 3. \end{cases}$$

Remark 2.5.1 *We will have to generate where we usually ended up with. But here we have no choice, we will only jump from 1 → 2, 2 → 3, and 3 → 1*

Chapter 3

MODELING GLIOMA CELLS USING ELLIPSES

Abstract

Modeling cancer cells is essential to better understand the invasive nature of glioma cells. Our goal is to study how the morphology of the cell influences the formation of so-called flock or stream formation. It has been observed experimentally that streams increase the speed of the overall tumor growth. We propose an original agent-based model modeling each cell as an ellipsoid and we show that by stretching the cells, one can increase stream formation.

A systematic numerical investigation of the model in $\mathbb{R}^2$ is performed and we deduce a phase diagram of the dynamics. Moreover, we study the effect of the density and show that, in contrast to classical model of flocking, increasing the density tends to reduce the formation of flocks. Preliminary work in $\mathbb{R}^3$ show that the dynamics tend to similarly generate streams when the density is large.

3.1 Introduction

Primary brain tumor is one of the most common causes of cancer related to death. The key to develop successful cancer therapy is to understand the essential mechanisms by which individual cancer cells manage to develop and spread throughout the organism. Of particular interest is the study of the morphology of the cells and how it could lead to different macroscopic behaviors (e.g. stream formation, diffusion behavior). Streaming promotes, facilitates, and influences the increased speed of the overall tumor growth. This raises the question on how does the cell morphology influence the pattern formation of the overall dynamics. This question is difficult to answer as

we rarely have access to the time evolution of the cell *in vivo*.

Modeling could provide a partial answer as it allows to explore various scenarios (Deisboeck and Couzin, 2009; Ribba *et al.*, 2004; Anderson and Quaranta, 2008; Harpold *et al.*, 2007). In particular, agent-based models (*microscopic level*) are convenient to include essential features of cell behavior (i.e. motility, cell-cell interaction) as it is used for various other so-called self-organized dynamics (e.g. pedestrians (Moussaid *et al.*, 2012), birds (Ballerini *et al.*, 2008), and fish (Couzin *et al.*, 2005)). To investigate the role of the cell morphology, we will model cells as ellipses or ellipsoids. Such model assumption is common in the study of bacteria, for instance viscoelastic ellipsoids have been used in (Palsson and Othmer, 2000) or self-propelled spheres in (Menzel and Ohta, 2012) (see also (Li *et al.*, 2017; Chen *et al.*, 2012; Pletjushkina *et al.*, 1994; Xiong *et al.*, 2010; Balagam *et al.*, 2014; Ariel *et al.*, 2013)). One of the specificity of our work is to study the effect on the density, while the dynamics are in a regime near a non-overlapping constraint. Our main finding is to show that the morphology of the cell influences the pattern formation of the tumor and how stretching the cells (ellipsoid shape) can increase stream formation.

The agent-based model proposed in this work contains essentially two mechanisms: i) self-propulsion, ii) cell-cell avoidance due to non-overlapping constraints. But since the cells have an ellipsoid shape, cell-cell avoidance lead to two effects: repulsion and steering. The larger the eccentricity of the cell, the larger the effect of the steering. In contrast to classical model of flocking (Cucker and Smale, 2007; Vicsek *et al.*, 1995), in our model cells do not take into account the velocity of their neighbors. First, we investigate numerically in $\mathbb{R}^2$ how the eccentricity enhances the formation of flocks (i.e. all the cells moving in the same direction) using as an indicator the so-called polarization of the configuration. Surprisingly when eccentricity becomes too large, flocks become less likely to occur. Then, we study how the density affects

the dynamics by increasing the number of cells. Since we do not suppose a mean-field type interaction (there is no averaging in the interaction), increasing slightly the density could lead to drastic change (Cho *et al.*, 2007). In our dynamics, we observe the emergence of streams when the density is larger, meaning that cells are aligned but not necessarily moving in the same direction. Streams are measured using the nematic average where we identify a vector ω and its opposite $-\omega$.

Beside the influence of the morphology of the cell, the other key component of the dynamics is the strength of both the repulsive effect and the steering effect which are determined by the two coefficients α and β (resp.). One could think that increasing steering effect (i.e. β larger) would enhance alignment and therefore lead to flocks or streams. Our numerical investigation reveals that this is not the case. In particular at large density, it is only in the regime where β is small that a flock or a stream could emerge. This result seems counter-intuitive but we have to emphasize that the alignment in our dynamics is only *indirect*, cells do not perceive each other velocity. Thus, it is an interplay between space constraint and steering that lead to the emergence of a stream or flock. Increasing one effect does not necessarily enhance alignment.

Although most of our simulations are done in dimension 2, we provide some preliminary results in $\mathbb{R}^3$. Stream formation is more challenging to observe in $\mathbb{R}^3$ since cells avoiding each other are no longer moving aligned or in opposite direction as in $\mathbb{R}^2$. However, our simulations show that stream formation do still occur in $\mathbb{R}^3$ providing that we maintain a large density of cells in the domain.

The complexity of the dynamics shows that it is difficult to predict a priory effect of each mechanism. Therefore, it would be of great interest to try developing a multi-scale approach to study the dynamics from a macroscopic viewpoint (Lowengrub *et al.*, 2009; Saut *et al.*, 2014; Deisboeck and Stamatakos, 2010). Moreover, if one would like to perform data-model comparison (Swanson *et al.*, 2003; Hawkins-Daarud

et al., 2013), most of the experimental observations are at a macroscopic scale. Thus, investigating the partial-differential equation associated with the dynamics (Morale *et al.*, 2005; Filbet *et al.*, 2005; Peruani *et al.*, 2008) could provide a way to bridge this gap.

The paper is organized as follows: we first present the agent-based model in section 3.2, then we study how the cell morphology influences the dynamics in section 3.3. A systematic numerical investigation of the model in $\mathbb{R}^2$ varying two key parameters is performed in section 3.4 which produces several phase diagrams of the dynamics at various density. We explore the model in $\mathbb{R}^3$ in section 3.5 and draw a conclusion and future work in section 3.6.

3.2 Agent-Based Model

We propose an agent-based model to describe the motion of individual cancer cells. The dynamics combine cell-motility (i.e. self-propulsion) and cell-cell interaction (e.g repulsion or adhesion). Specifically, we consider N cells described with a vector position $\mathbf{x}_i \in \mathbb{R}^d$ with d the spatial dimension ($d = 2$ or 3), moving with velocity $c\omega_i$ where $c > 0$ is the speed (supposed constant) and $\omega_i \in \mathbb{S}^{d-1}$ the velocity direction. The main novelty of the model is to consider an elliptic or ellipsoid shape for the cell. Thus, we consider two axis denoted a and b for (respectively) the major and minor axis (see figure 3.1-left). As two cells cannot occupy the same spatial position, cells will *push each other* if they are too close. Thus, we define an interaction potential V_i between cells that measures the *tension* exerted on cell i generated by the surrounding cells:

$$V_i = \sum_{j=1,\, j\neq i}^{N} \Phi(r_{ij}^2) \quad \text{with} \qquad r_{ij}^2 = \left| \frac{\langle \mathbf{x}_j - \mathbf{x}_i, \omega_i \rangle}{a} \right|^2 + \left| \frac{\langle \mathbf{x}_j - \mathbf{x}_i, \omega_i^{\perp} \rangle}{b} \right|^2 . \tag{3.2.1}$$

The quantity r_{ij} is referred to as the *normalized distance* between the membrane of cell i and the position of cell j. For instance, if $a = b$ we recover that r_{ij} is simply the norm $\|\mathbf{x}_j - \mathbf{x}_i\|/a$. The modification takes into account that the cell is more sensible to obstacle in front rather obstacle on the side. The model is defined in $\mathbb{R}^2$ (i.e. $d = 2$) but it can be generalized in $\mathbb{R}^3$ defining r_{ij} as followed:

$$r_{ij}^2 = \frac{1}{b^2}\Big(\|\mathbf{x}_j - \mathbf{x}_i\|^2 - e^2\Big[\,(\mathbf{x}_j - \mathbf{x}_i)\cdot\omega_i\Big]^2\Big) \tag{3.2.2}$$

where $e \in (0,1)$ is the eccentricity of an ellipse defined as $e = \sqrt{1 - \frac{b^2}{a^2}}$.

To prevent overlapping, the function Φ has to be singular at the origin. We choose the following smooth function with compact support (see figure 3.1-right):

$$\Phi(s) = \begin{cases} \frac{1}{s}\exp\left(\frac{-1}{1-s}\right) & \text{if } 0 < s < 1 \\ 0 & \text{if } s \geq 1. \end{cases} \tag{3.2.3}$$

As r_{ij} decreases, Φ increases resulting into *repulsion*. We have now defined all the quantities required to define our agent-based model.

Definition 3 *Consider $(\mathbf{x}_i, \omega_i) \in \mathbb{R}^d \times \mathbb{S}^{d-1}$ for $i = 1..N$ and the dimension space $d = 2$ or $d = 3$. The self-propelled dynamics are defined as:*

$$\dot{\mathbf{x}}_i = \overbrace{c\,\omega_i}^{\text{self-propulsion}} - \overbrace{\alpha\nabla_{\mathbf{x}_i}V_i}^{\text{repulsion}} \tag{3.2.4}$$

$$\dot{\omega}_i = \underbrace{-\beta P_{\omega_i^\perp}\left(\nabla_{\omega_i}V_i\right)}_{\text{steering}} \tag{3.2.5}$$

where α, β and c are positive constant, V_i is the tension defined in (3.2.1) *and $P_{\omega_i^\perp} = \mathrm{Id} - \omega_i \otimes \omega_i$ is the projector operator onto the the normal plane to ω_i (it ensures that ω_i stays of norm 1).*

In order to reduce the tension generated by neighboring cells, a cell can either move away (i.e. *repulsion* effect) or change its direction (i.e. *steering* effect). Both

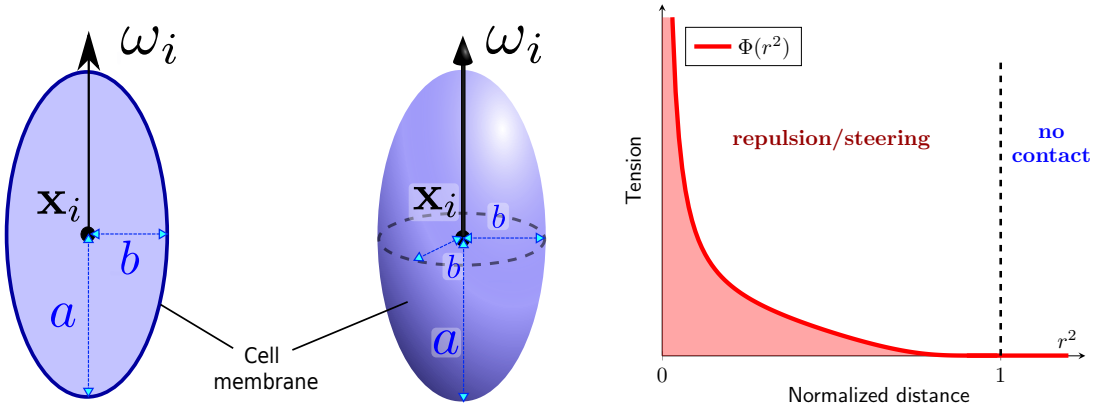


Figure 3.1: Left: A cell i is Described by its Position x_i, Orientation ω_i and its Elliptic Shape Determined by the Two Morphological Components a and b. Right: Function Φ Relies Spacing Between Cell r_{ij} (3.2.1) into *Tension* that Generates *Repulsion* when Two Cells Touch each Other

maneuvers are pondered by the coefficients α and β representing the strength of each effect. Using the expression of V_i (3.2.1), one can deduce an explicit expression of the dynamics (see appendix 3.7.1). Notice that if the cell has a circular shape (i.e. $a = b$ and the eccentricity $e = 0$), its orientation will remain constant i.e. $\dot{\omega}_i = 0$. Indeed, if that case, turning will have no effect on the reduction of the tension V_i. Thus, *steering effects* could only occur when $a \neq b$.

Remark 3.2.1 *Notice that r_{ij} cannot be considered a distance between cells i and j as it is not symmetric (i.e. $r_{ij} \neq r_{ji}$ in general). Thus, the influence of the cell i on j is in general different from the cell j on i (i.e $\Phi(r_{ij}^2) \neq \Phi(r_{ji}^2)$).*

The interaction would be symmetric if the repulsion was proportional to the overlapping of the two cells i and j. However, computing such overlapping is rather difficult for ellipse (details can be found in(Hughes and Chraibi, 2012)). Moreover, we would have to perform such computations for all the pairs (i, j) and at each time

step. Our model simplification allows to bypass this difficult.

3.3 Morphological Influence

3.3.1 Eccentricity Induces Alignment

Our first investigation of the agent-based model (3.2.4)-(3.2.5) is concerned with the impact of the morphology of the cell on the global behavior of the dynamics. As stated above, cells have perfect rounded shape when the two parameters a and b are equal (i.e. eccentricity e is zero) whereas they have elliptic or ellipsoid shape when $a > b$ (i.e. $e > 0$). We are going to vary the eccentricity e and measure how this change affects the cell distribution.

But before varying the morphological parameters a and b, there are several other parameters to be determined in our dynamics. When it is possible, we use experimental values that have been qualitatively estimated in vivo. For instance, it has been observed that glioma cell size varies in between $5\mu m$ to $20\mu m$ for their diameter and their speed varies around $10\mu m/h$ (Baker *et al.*, 2014a). However, some parameters cannot be infer from experimental observations such as the strength of the repulsion α and the steering β. A more detail investigation of these two parameters will be conducted in the next section. For now, we fix their values as indicated in table 3.1.

We perform the simulation in a square domain $\Omega = [0, L] \times [0, L]$ with periodic boundary conditions. For the initial condition, the position of the N particles $\{\mathbf{x}_i\}_{i=1\dots N}$ are distributed uniformly in Ω and their direction $\{\omega_i\}_{i=1\dots N}$ are taken randomly on the unit circle $\mathbb{S}^1$. In figure 3.2, we draw the final configuration of the dynamics after $T = 1000$ unit time for two types of cells: circular shape ($a = b = 4\mu m$, $e = 0$) and elliptic shape ($a = 5.5\mu m$, $b = 3\mu m$, $e = .84$). We observe that elliptic cells have formed clusters moving in the same directions while circular cells have no particular

Diameter cell (front/back)	**2a**	$\mathbf{8-14}\,\mu m$
Diameter cell (side)	**2b**	$\mathbf{6-8}\,\mu m$
Motility	c	$10\,\mu m/h$
Length domain Ω	L	$300\,\mu m$
Number of cells	N	1000
Cell-cell repulsion	Φ	eq. (3.2.3)
Strength repulsion	α	$40\,\mu m^2/h$
Strength steering	β	$1\,h^{-1}$

Table 3.1: Parameters Used for the Simulations of Figures 3.2 and 3.4

spatial organization.

3.3.2 Statistical Characterization

To further investigate the dynamics, we introduce several statistics to characterize the emergent behavior.

Definition 4 *Consider a velocity distribution* $\{\omega_i\}_{i=1..N} \subset \mathbb{S}^{d-1}$. *We denote by* ψ *the polarization:*

$$\psi = \frac{1}{N}\left|\sum_{i=1}^{N}\omega_i\right|. \tag{3.3.1}$$

Similarly, we define the nematic polarization (Li et al., *2017):*

$$\gamma = \sqrt{\langle\cos(2\theta_i)\rangle^2 + \langle\sin(2\theta_i)\rangle^2} \tag{3.3.2}$$

where θ_i *is the angle between the direction* ω_i *and the horizontal axis and* $\langle . \rangle$ *denotes the averaging over the indices* i *(i.e.* $\langle\cos(2\theta_i)\rangle = \frac{1}{N}\sum_{i=1}^{N}\cos(2\theta_i)$*).*

We say that the configuration is in a flocking configuration (i.e. cells moving in the same direction) when the polarization ψ is close to 1. Notice that the nematic

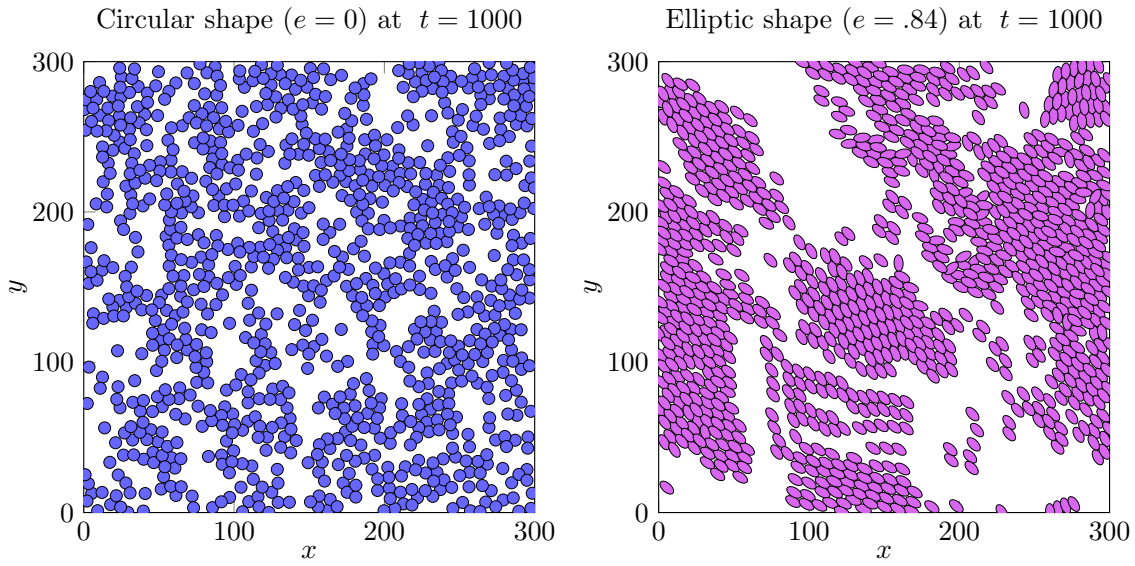


Figure 3.2: Snapshot of the Simulation of the Model Starting from a Uniform Distribution. After $t = 1000$ Unit of Time, Circular Cells ($a = b = 4\mu m$, $e = 0$) do not Form any Flocking Pattern (Left) whereas Elliptic Cells ($a = 5.5\mu m, b = 3\mu m$, $e = .84$) Move in a Common Direction (Right)

polarization γ will be also close to 1 in this case. We say that the configuration is in a stream formation (i.e. cells directions are parallel but not necessary moving in the same direction) when only the nematic polarization γ is close 1.

Remark 3.3.1 *The nematic polarization can be generalized in higher dimension (see appendix 3.7.2).*

The previous statistics only involve the velocity of the cells ω_i. We propose a third statistics to also characterize the spatial configuration.

Definition 5 *Consider a spatial configuration* $\{\mathbf{x}_i\}_{i=1\dots N} \subset \mathbb{R}^d$ *and a fixed radius* R. *We say that cell* i *is linked to cell* j *if the distance between the two particles is less*

than R. This defines a relationship (i.e. $i \sim j$) with $i \neq j$ defined:

$$i \sim j \quad \text{if and only if} \quad \|\mathbf{x}_j - \mathbf{x}_i\| \leq R. \tag{3.3.3}$$

Clusters $\mathcal{C}_k$ are defined as the connected components for this relationship: two cells i_0 and j_0 belong to the same cluster if there exists particles $i_1, \ldots, i_k$ (a path) such that

$$i_0 \sim i_1, \ i_1 \sim i_2, \ \cdots, \ i_k \sim j_0. \tag{3.3.4}$$

The cluster size $|\mathcal{C}_k|$ denotes the number of cells belonging in the cluster $\mathcal{C}_k$. The average cluster size $\overline{|\mathcal{C}|}$ is defined as the expected cluster size $|\mathcal{C}_k|$ over all the positions:

$$\overline{|\mathcal{C}|} = \frac{1}{N} \sum_{i=1}^{N} |\mathcal{C}(\mathbf{x}_i)|, \tag{3.3.5}$$

where $\mathcal{C}(\mathbf{x}_i)$ denotes the cluster containing the cell i.

We illustrate the three statistics used in figure 3.3.

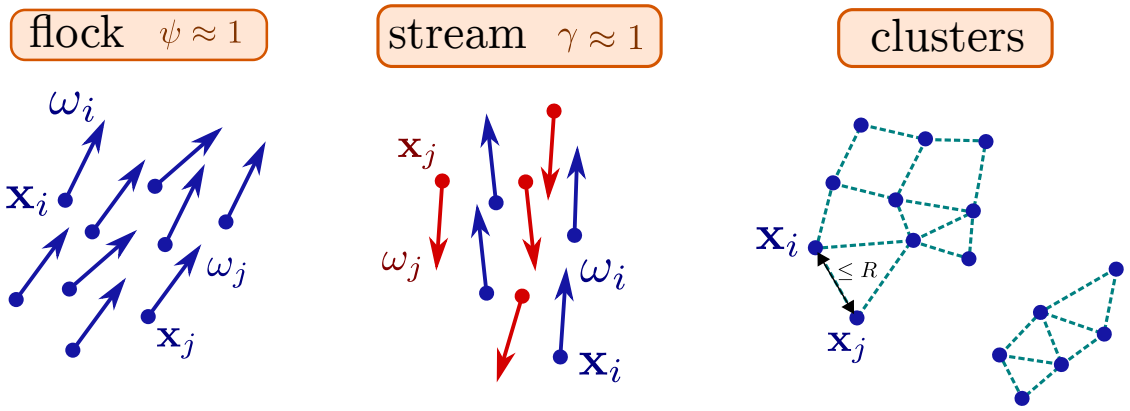


Figure 3.3: The Statistics Used to Characterize the Dynamics: The Polarization ψ (3.3.1), the Nematic Polarization γ (3.3.2) and the Clustering (3.3.3)-(3.3.4)

In figure 3.4 - left, we measure the value of the polarization ψ over time for different shapes of the cells by varying the coefficients a and b. When the cells have a circular shape ($a = b = 4\mu m$, $e = 0$), the polarization ψ remains close to zero

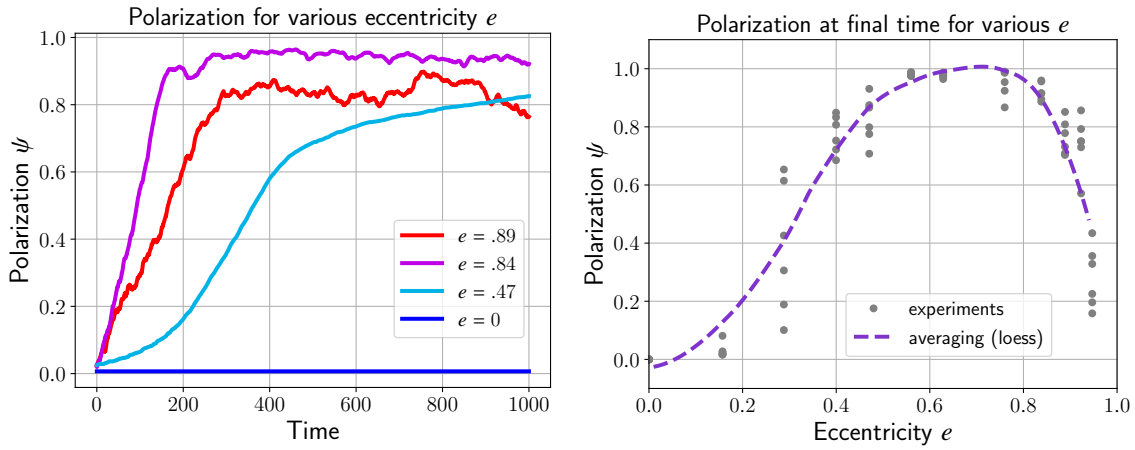


Figure 3.4: Polarization ψ (3.3.1) over Time while Varying the Eccentricity of the Cell e. Ellipsoid Cells that Align will Lead to an Increase of ψ Close to its Maximum 1. For the Circular Cells (Blue Curve), the Polarization Remains very Low as no Streams Emerge from the Dynamics (Left). The Polarization ψ over Eccentricity e During the Final Time $t = 1000$ of the Left Figure will Form a Parabola. By Increasing the Eccentricity, there is no Fundamental Impact on the Polarization Coefficient of the Cells (Right)

for all time whereas it increases up to a maximum close to 1 when the eccentricity is greater than zero. The relation between eccentricity and polarization is however non-trivial: increasing the eccentricity does not necessarily lead to large polarization. For instance, the polarization with eccentricity $e = .89$ is significantly smaller than with eccentricity $e = .84$.

To further investigate the relationship between polarization and eccentricity, we plot in figure 3.4 - right the polarization at the final time for several experiments (changing the seed for the initial condition) and various eccentricity e. We then perform a local regression ('loess' method) to estimate the expected polarization ψ as a function of e. We observe that increasing the eccentricity e leads to larger polarization

up to $e \approx .7$ but then the polarization quickly decays for larger eccentricity.

3.3.3 Indirect Alignment

In classical models flocking (Vicsek *et al.*, 1995; Cucker and Smale, 2007), each individual has access to the velocity of its neighbors. By relaxing its own velocity toward the average velocity of its neighbors, a flock emerges. However, in the agent-based model proposed (3.2.4)- (3.2.5), the cell i has no knowledge of the velocity of any of its neighbors, i.e. ω_j is not used to define the evolution of $(\mathbf{x_i}, \omega_i)$. Therefore, it is unclear at first why the dynamics proposed could generate similar flocking patterns.

To provide a partial answer, we provide a linear perturbation analysis of the model with respect to the eccentricity in the case of solely only two cells i and j. Let's denote θ_{ij} the angle between the direction of the cell ω_i and the relative position vector $\mathbf{x_j} - \mathbf{x_i}$ (see figure 3.5). Thus, one can write $\omega_i = (\cos\theta_{ij},\, \sin\theta_{ij})^T$ using the basis $\{\omega_i,\, \omega_i^\perp\}$. In particular,

$$\omega_i' = (-\sin\theta_{ij},\, \cos\theta_{ij})^T \theta_{ij}' = \omega_i^\perp \theta_{ij}'.$$

We deduce using the formulation (3.7.4) that:

$$\theta_{ij}' = \beta \frac{2e^2}{b^2} \Phi'(r_{ij}^2) x_{ij} y_{ij}.$$

where $x_{ij} = \langle \mathbf{x}_j - \mathbf{x}_i, \omega_i \rangle,\ y_{ij} = \langle \mathbf{x}_j - \mathbf{x}_i, \omega_i^\perp \rangle$. Moreover, elementary geometry shows that $x_{ij} = |\mathbf{x}_j - \mathbf{x}_i| \cos\theta_{ij}$ and $y_{ij} = |\mathbf{x}_j - \mathbf{x}_i| \sin\theta_{ij}$ and thus:

$$\theta_{ij}' = C \sin 2\theta_{ij}\,, \qquad \text{with } C = \beta \frac{e^2}{b^2} \Phi'(r_{ij}^2) |\mathbf{x}_j - \mathbf{x}_i|^2. \tag{3.3.6}$$

Notice that $C \leq 0$ since $\Phi'(r_{ij}^2) \leq 0$. As a consequence, if i and j stay close enough (e.g. $r_{ij}^2 < 1$), there are two stable equilibria for θ_{ij} at $\pm\pi/2$. Sketching the phase diagram in figure 3.5 indicates that ω_i will rotate toward the stable equilibrium to align with $(\mathbf{x_j} - \mathbf{x_i})^\perp$. By a similar argument, ω_j will be orthogonal to $\mathbf{x}_i - \mathbf{x}_j$ as well.

Therefore, instead of a direct alignment between ω_i and ω_j, we have an indirect alignment: both vectors will align with $(\mathbf{x_j} - \mathbf{x_i})^{\perp}$. Notice that this indirect form of alignment allow for the two vectors ω_i and ω_j to be *negatively* aligned, i.e. $\omega_i = -\omega_j$ which could lead to streaming formation. Moreover, in higher dimension, since $(\mathbf{x_j} - \mathbf{x_i})^{\perp}$ is an hyperplane, it is unclear whether such dynamics will also generate flocking or stream patterns.

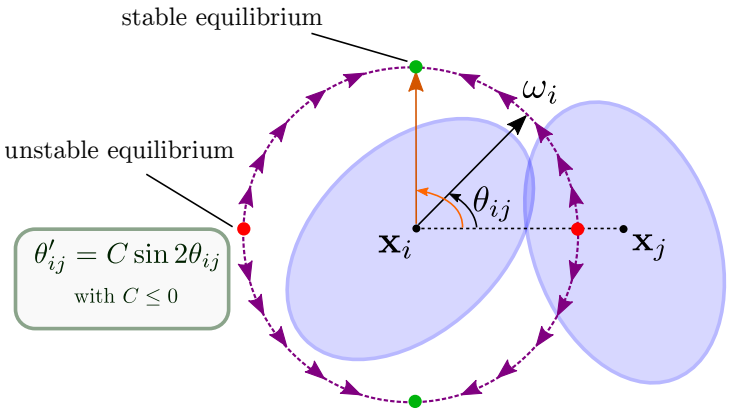


Figure 3.5: Indirect Alignment of Two Cells i and j. Both Cell i and j will Rotate to Align with the Orthogonal Vector to $(\mathbf{x_j} - \mathbf{x_i})$

3.4 Density Effect

In the previous section, we have investigated how the cell morphology (i.e. a, b) could favor the emergence of flocking patterns (i.e. cells moving in the same direction). Our formal analysis show that we could also observe stream formation (i.e. cell moving in opposite direction). In this section, we will show that such stream formations do emerge in a certain regime. We need for that to investigate the dynamics when we vary the parameters α (strength repulsion), β (strength steering) and N (density). We will fix the shape of the cells with $a = 5.5\mu m$ and $b = 3\mu m$ as they are the most common values experimentally. The range of the parameters are given in the table

3.2.

Diameter cell (front/back)	$2a$	$11\,\mu m$
Diameter cell (side)	$2b$	$6\,\mu m$
Motility	c	$10\,\mu m/h$
Length domain Ω	L	$300\,\mu m$
Number of cells	$\mathbf{N}$	$\mathbf{1000 - 2000}$
Cell-cell repulsion	Φ	eq. (3.2.3)
Strength repulsion	α	$\mathbf{10 - 200}\,\mu m^2/h$
Strength steering	β	$\mathbf{.1 - 10}\,h^{-1}$

Table 3.2: Parameters Used for the Simulations of Figures 3.6- 3.14

3.4.1 Emergence of Streams

To illustrate the formation of streams in the dynamics (3.2.4)-(3.2.5), we perform simulations in the regime: $\alpha = 100$, $\beta = .1$ (strong repulsion, low steering). In figure 3.6, we plot a snapshot of three simulations where we increase the density from $N = 1000$ to $N = 2000$ at $t = 1000$ unit of time. We color code the cells depending on their orientation: we determine the nematic average direction Ω_{nem} (3.7.2) (see appendix 3.7.2) and then color each cell i in blue if $\langle \omega_i, \Omega_{nem} \rangle > 0$ and in red if $\langle \omega_i, \Omega_{nem} \rangle < 0$.

We notice that when the number of particles is low with $N = 1000$ (figure 3.6 - **a**), almost all the cells are perfectly aligned in the same direction leading to a flocking configuration. The evolution of the polarization ψ given in figure 3.7 confirms this observation since ψ becomes close to 1 for $N = 1000$. As we increase the density with $N = 1500$ (figure 3.6 - **b**), we observe that number of cell moving in the opposite direction becomes larger making the polarization decay to only $\psi = .2$. Finally in the

case where $N = 2000$ (figure 3.6 - **c**), the number of cells moving in opposite direction become balanced and we observe the formation of a stream where the flow inside the domain is bidirectional. Indeed, the polarization ψ is close to zero for $N = 2000$ whereas the nematic polarization γ is around .9.

3.4.2 Local Minimum for the Energy

We conclude that increasing the density of cells is the underlying mechanism for stream formation. However, one has to notice that we always use as initial configuration random configurations for the velocities ω_i. If one would start from a perfect flock with no overlapping (i.e. $\omega_i = \Omega$ for all i and $r_{ij} > 1$ for all i, j), then the configuration will stay in the same configuration, it will be simply transported with a constant velocity. Thus, we will not observe the formation of a stream even at large density (e.g. $N = 2000$). In other words, flocking configuration can be as a globally stable configuration. In contrast, in a stream formation, the cells at the frontier between the regions moving in opposite direction are in an unstable equilibrium (see figure 3.5). However, if the steering coefficient β is small enough, streaming configuration can be maintained, the physical constraint of non-overlapping (through the repulsion α) prevent the cells to turn.

Another formal justification of the emergence of stream comes from the total potential energy V (3.2.1):

$$V = \sum_{i=1}^{N} V_i = \sum_{i,j=1,\, j\neq i}^{N} \Phi(r_{ij}^2), \tag{3.4.1}$$

with r_{ij} given by (3.2.1). The dynamics (3.2.4)-(3.2.5) is a gradient descent of the potential V (i.e. $\mathbf{x}_i' = -\alpha \nabla_{\mathbf{x}_i} V_i$, $\omega_i' = -\beta P_{\omega_i^\perp}(\nabla_{\omega_i} V_i)$) combined with a free transport part (i.e. $\mathbf{x}_i' = c\omega_i$). The gradient descent decays the potential V whereas the free transport could either increase or decrease V. But as we increase α, the dynamics

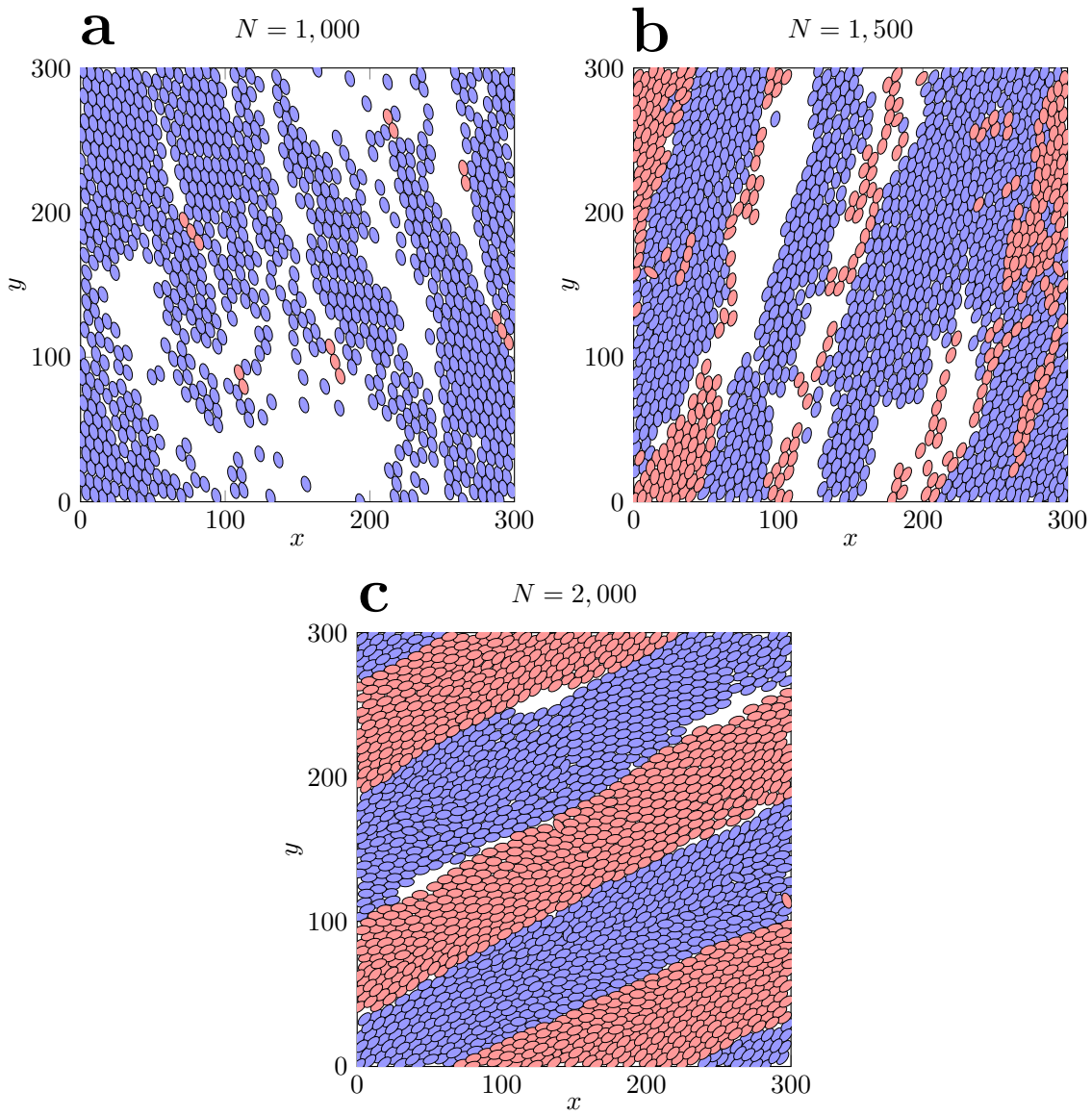


Figure 3.6: Snapshots of the Simulation of the Dynamics at $t = 1000$ Unit Time for Various Cell Densities ($N = 1000, 1500$ and 2000). Red Cells are Moving in Opposite Direction to the Blue Cells. Flocking Appears when the Density is Low (a) but then the Dynamics Start to Converge to Stream Formation as we Increase the Density (c)

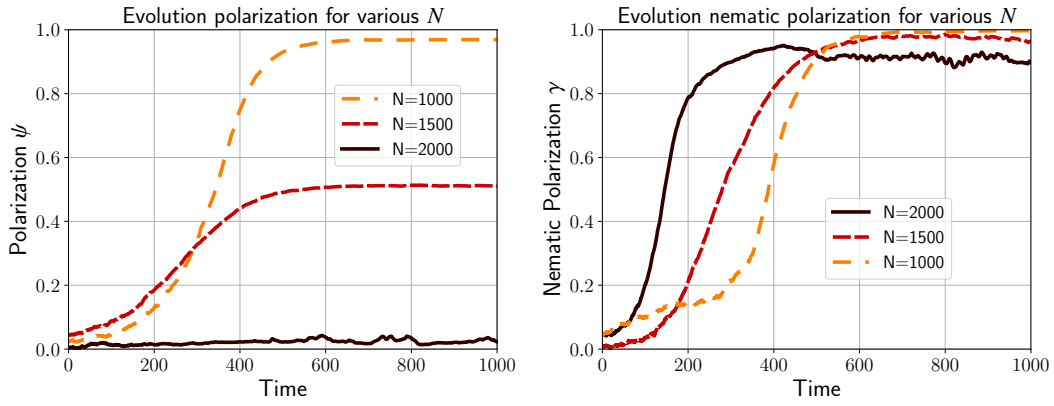


Figure 3.7: The Polarization ψ, and the Nematic Polarization γ are Represented for the Previous Cases in Figures 3.6. When Flocking Formation Appears, ψ and γ Converge Approximately to 1 over Time. When Stream Emerges, $\psi \approx 0$, and $\gamma \approx 1$

become more likely to stuck into local equilibrium (i.e. stream). The perturbations from the free transport part of the dynamics are insufficient to move the configuration away from a local equilibrium (see fig. 3.8).

3.4.3 Phase Diagram

We have identified so far two types of configurations: flocking when the cells are aligned (i.e. $\psi \approx 1$ (3.3.1)), and streaming when the cells are moving in opposite direction (i.e. $\gamma \approx 1$ (3.3.2) and $\psi \ll 1$). The convergence of the dynamics toward one of these configurations depend on the density N (figure 3.6) and on the parameters α and β. We would like to study the effects of repulsion and alignment (i.e the coefficients α and β resp.) for the three distinct cases of N.

With this aim, we fix the shape of the cells ($a = 5.5\mu m$ and $b = 3\mu m$) as mentioned previously and we make a systematic analysis by varying continuously $\alpha \in [0, 200]$ and $\beta \in [0, 10]$. For each value of α and β, we perform 5 simulations and compute several statistics after $t = 1000$ unit of time. For instance, we plot in figure 3.9

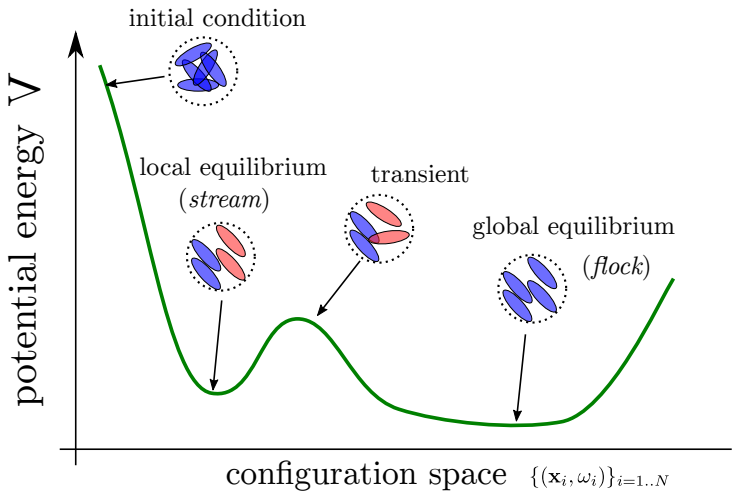


Figure 3.8: Sketch Representation of the Potential Energy V (3.4.1) over the Configuration Space $\{(x_i, \omega_i)\}_{i=1,\dots,N}$. Stream can be Seen as Local Equilibrium whereas Flock are Global Equilibrium. When the Parameter α (Repulsion) is Increased, the Stream Configuration Become more Stable

the polarization ψ depending on α and β in the case $N = 1000$. The scatter plot represents all the data points $(\alpha_j, \beta_j, \psi_j)$. Notice that for a given set (α, β) we find different polarization ψ due to the randomness of the initial condition. We represent as a surface the average polarization $\langle\psi\rangle$ computed from the 5 final configurations with similar parameters (α, β). To reduce the fluctuation, we estimate a local averaging ('loess') in figure 3.9 which makes possible the estimation of smooth region where the polarization ψ is higher than a given threshold. Regarding of using averaging or smoothing, we observe that the polarization is surprising small when β is large and α small. There is also another region where the polarization decays when α is large and β is small.

A better visualization is to draw the polarization ψ as a heat-map depending on α and β (figure 3.10-left). Thanks to the smoothing, we can also estimate contours at different thresholds ($\psi = .5$ and $\psi = .8$). We proceed similarly with the nematic

polarization γ (3.3.2) in figure 3.10-right. We notice that in contrast to the polarization ψ, the nematic polarization remains large even when β is small and α is large. Indeed, it is the regime where we observe the formation of streams.

Finally, we also use a third statistics to characterize the configuration using the average clusters size $\overline{|\mathcal{C}|}$ (3.3.5). We use the radius $R = 10\mu m$ to define the clusters (i.e. two cells are connected if their distance $\|\mathbf{x}_j - \mathbf{x}_i\|$ is less than $10\mu m$). The average size cluster $\overline{|\mathcal{C}|}$ is then estimated in figure 3.11. We observe two regions: cluster sizes are (relatively) smaller when α is small and independent of β. Thus, the repulsion α governs the formation of clusters.

We combine the three statistics (polarization ψ, nematic polarization γ, cluster size) to create a phase diagram in the parameter space (α, β). Three regions are delimited:

i) **flocking**: $\{\alpha, \beta \text{ such that } \psi > .8\}$,

ii) **streaming**: $\{\alpha, \beta \text{ such that } \gamma > .7 \text{ and } \psi < .8\}$,

iii) **scattering**: $\{\alpha, \beta \text{ such that } \overline{|\mathcal{C}|} < 600\}$.

The results are given in figure 3.12. For most of the parameters α and β, the dynamics converge to a flock.

Performing a similar investigation for $N = 1500$ and $N = 2000$ lead to drastically different results. The regions where flocking occurs are more narrow (figure 3.13-a). But surprisingly stream formation is still occurring for all values of α as long as the steering coefficient β is small enough (figure 3.13-b). Only the cluster formation through the statistics $\overline{|\mathcal{C}|}$ remain similar (see figure 3.13-c) as in the case $N = 1000$. As a result, the phase diagrams for $N = 1500$ and $N = 2000$ contain large region unidentified as either flock or stream (figure 3.14). Notice that increasing density

does not penalize the formation of stream in the region where β is small and α is large.

3.5 Dynamics in 3D

We would like to finally study the dynamics (3.2.4)-(3.2.5) in $\mathbb{R}^3$. In contrast to the 2D case, the indirect alignment (see section 3.3.3) lead to various possible equilibria. Indeed, two nearby cells i and j must have their velocity ω_i and ω_j orthogonal to $(\mathbf{x_j} - \mathbf{x_i})$. In $\mathbb{R}^2$, we conclude that $\omega_i = \omega_j$ or $\omega_i = -\omega_j$ at equilibrium. But this is no longer the case in $\mathbb{R}^3$. Therefore, it is unclear if the dynamics will produce flock or stream.

Diameter cell (front/back)	a	$11\,\mu m$
Diameter cell (side)	b	$6\,\mu m$
Motility	c	$10\,\mu m/h$
Length domain Ω	L	$70\,\mu m$
Number of cells	$\mathbf{N}$	$\mathbf{1000 - 2000}$
Cell-cell repulsion	Φ	eq. (3.2.3)
Strength repulsion	α	$100\,\mu m^2/h$
Strength steering	β	$.1\,h^{-1}$

Table 3.3: Parameters Used for the Simulations in $\mathbb{R}^3$ (Figures 3.15-3.16)

Another difference in $\mathbb{R}^3$ is that the nematic polarization γ is no longer defined. However, we have already provided an alternative in appendix 3.7.2 using the quantity J (3.7.1). We also use a smaller domain for our simulation in order to keep the same density as in $\mathbb{R}^2$ (maintaining a similar ratio "volume occupied/volume domain") thus we reduce the size of the box to $L = 70\mu m$. Otherwise, the other parameters are similar to the previous simulations (see table 3.3), in particular we use $\alpha = 100$

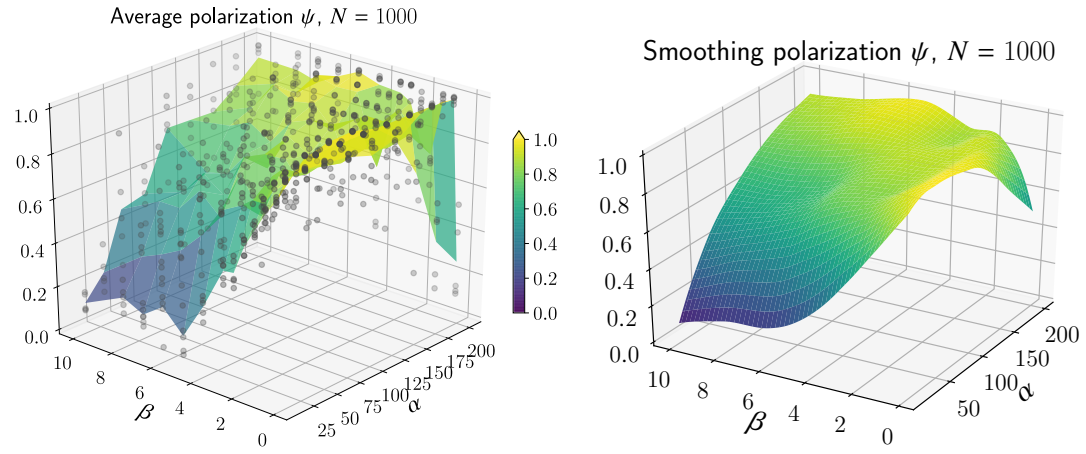


Figure 3.9: Left: For Each Pair α, β, we Estimate the Polarization ψ (Scatter Point) at the End of 5 Simulations. The Average is then Computed to Construct a Surface Plot. Right: We Use a Local Regression ('loess') to Estimate ψ as a Function of α and β to Reduce the Fluctuation

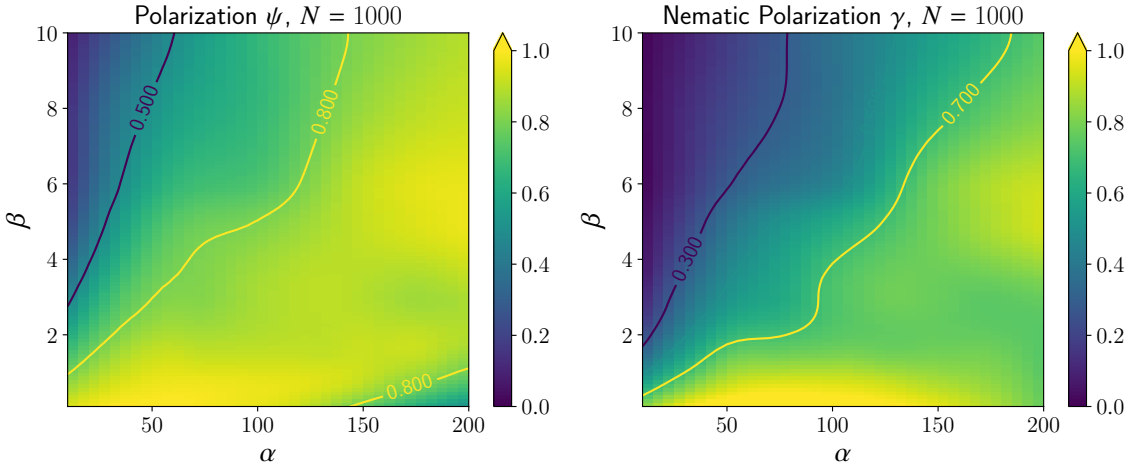


Figure 3.10: Left: Heat-Map Representation of the Polarization ψ as a Function of α, β Using the Smooth Estimation from Figure 3.9-right. The Contour $\psi = .8$ will be Used to Determine the Region when the Dynamics Generate Flock. Right: We Perform a Similar Analysis as the Left Figure Using the Nematic Polarization γ

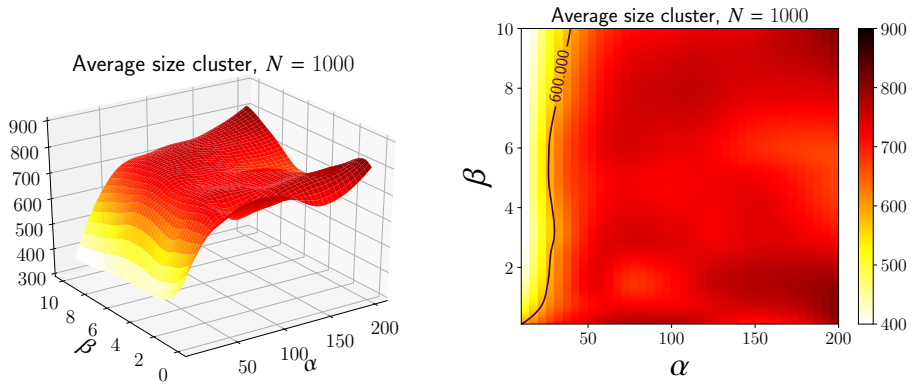


Figure 3.11: The Average Size Cluster $\overline{|\mathcal{C}|}$ (3.3.5) for Various Values of α and β. The Estimation has been Smoothed Using Local Regression ('loess'). We Deduce the Region when the Cluster Size is Below a Certain Threshold

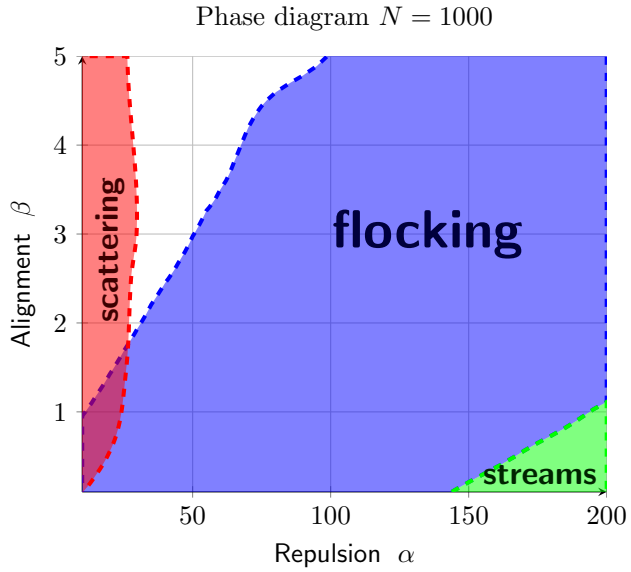


Figure 3.12: Combining the Results of the Figures 3.10 - 3.11, we Create a Phase Diagram Consisting of Three Regions for the Configuration: Flocking ($\psi > .8$), Streams ($\gamma > .7$ and $\psi < .8$) and Scattering ($\overline{|\mathcal{C}|} < 600$)

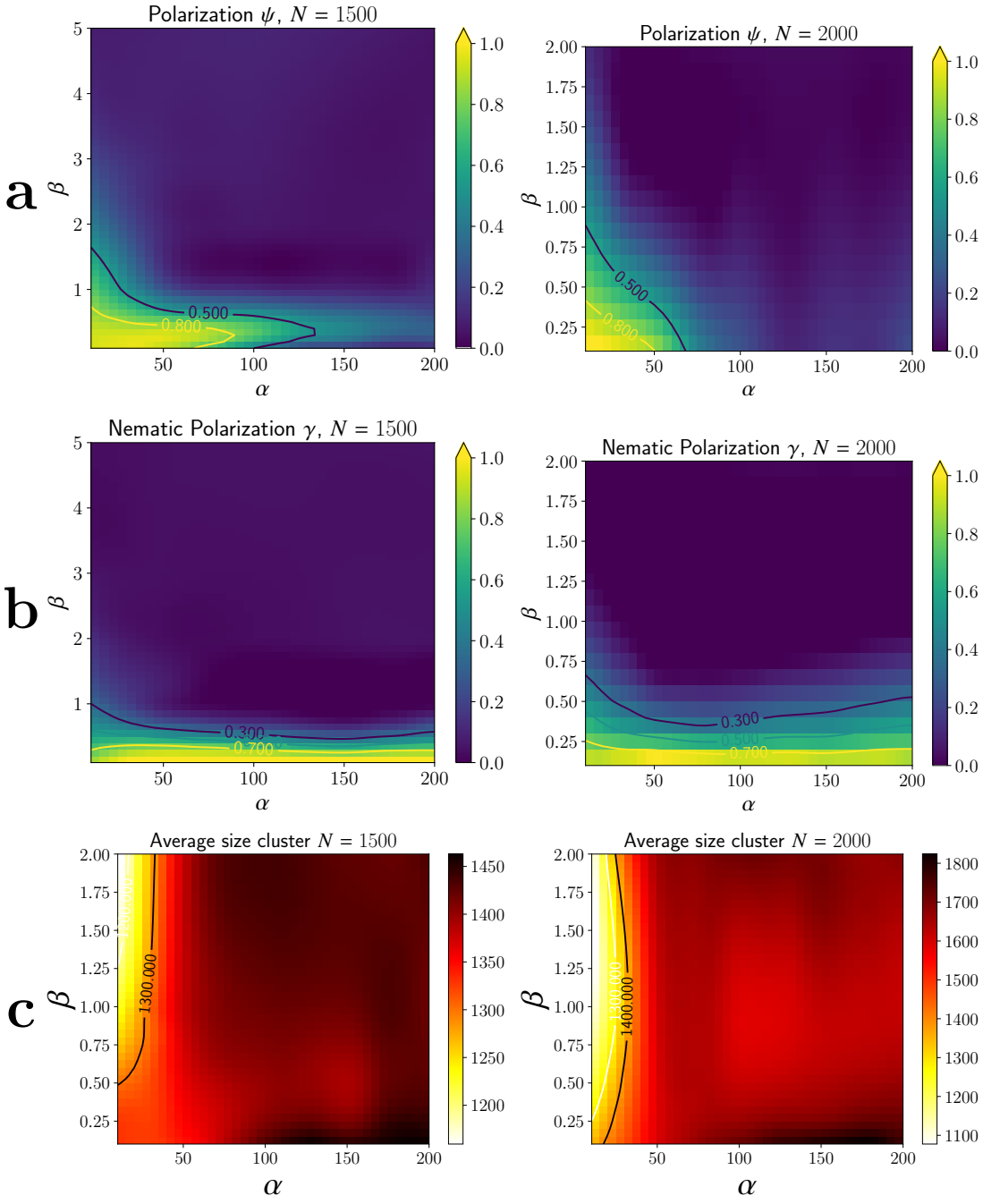


Figure 3.13: a) Average Polarization ψ for Various Parameters α and β with $N = 1500$ and $N = 2000$. Notice that the Polarization is Significantly Smaller Compared to the Case $N = 1000$ (figure 3.10). b) Nematic Polarization γ for $N = 1500$ and $N = 2000$. γ Remains Close to 1 for any Values of α when β is Small. c) The Average Cluster size $\overline{|\mathcal{C}|}$ Behave Similarly as in the Case $N = 1000$ with Smaller Clustering for Small Value of α

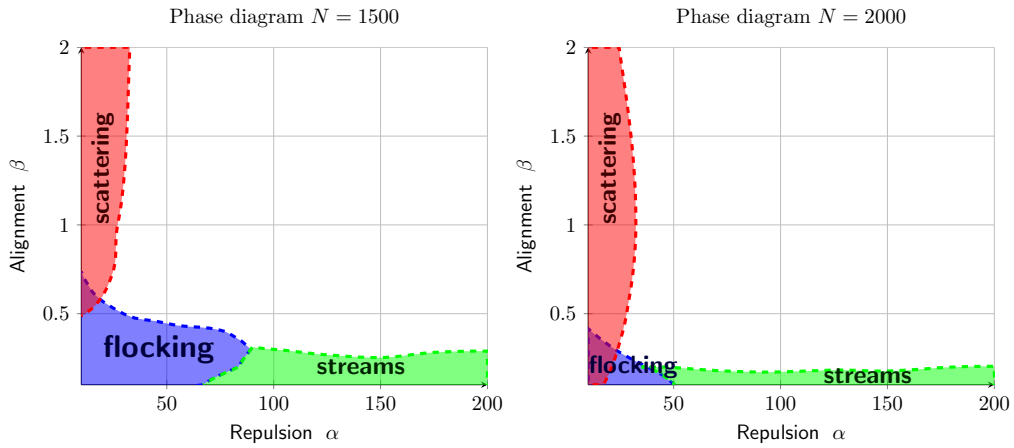


Figure 3.14: Phase Diagram when the Total Number of Cells N is 1500 (Left) and 2000 (Right). As we Increase the Density, the Regions for Flocking Configurations Drastically Reduce. However, Streams still Form when β is Small

and $\beta = .1$ to be in the regime more susceptible of stream formation (at least in $\mathbb{R}^2$).

First, we investigate with $N = 1000$ cells (low density). We plot the final configuration at $t = 1000$ unit times starting from two different initial conditions in figure 3.15. As in the previous figure, we color code the cells depending on their orientation to help visualize cells moving in opposite direction. We observe the formation of a flock (figure 3.15-top left) and of a stream (figures 3.15-top right). Note that even when the flock develops (top left), few isolated cells are still moving in opposing direction of the flow. Thus, both flock and stream configuration could emerge when the density is low.

The situation is different when we increase the density with $N = 1500$ and $N = 2000$, we only observe the formation of streams (see figures 3.15-bottom). Similar to the situation in $\mathbb{R}^2$, increasing the density reduces the possibility for the cells to rotate and therefore streams are more likely to occur. Plotting the time evolution of both the polarization ψ and nematic polarization J in figure 3.16 confirms our observation. The quantity J always converge to 1 whereas the polarization ψ stays

low except when $N = 1000$.

3.6 Conclusion

We have proposed an agent-based model describing the motion of cancer cells and study the emergence of pattern formations. In our model, the morphology of the cells played a key role since eccentricity allows the cells to align (indirectly) to each other and eventually form a flock or a stream. The emergence of such configuration is also governed by other parameters (α, β) as well as the density. Several phase diagrams summarizing the effects of these parameters are estimated for various density. In contrast to mean-field type models, the density drastically changed the dynamics and flocking configurations became more spare as the density increases.

There are several extensions that will be interesting to pursue. For instance, it will be interesting to mix cells with different morphological shapes (i.e. with different values for a and b) since not all the cells are identical (see for instance (Kansal *et al.*, 2000)). Or we could also have the shape of the cells evolving over time that could even lead to a birth/death process. But increasing the density is also challenging numerically as the dynamics become singular when two cells overlap which is more likely with a birth process. To avoid this complication, one could look for a continuous description of the dynamics through a Partial Differential Equation (PDE) (Morale *et al.*, 2005; Filbet *et al.*, 2005; Peruani *et al.*, 2008). Such PDE description might provide some hindsight about the emergence of flock or stream in certain regimes (e.g. $\alpha \gg 1$, $\beta \ll 1$).

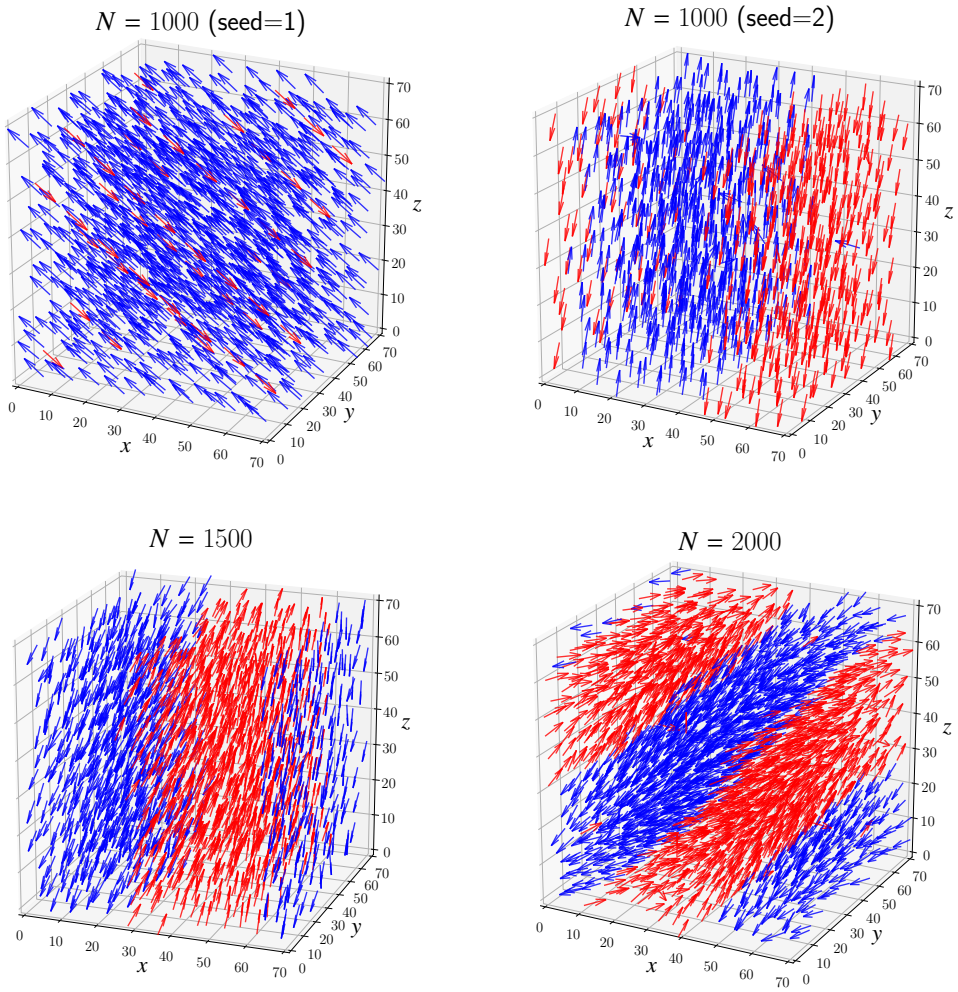


Figure 3.15: Snapshots of Simulation in $\mathbb{R}^3$ at $t = 1000$ Unit Time with Number of Particles N Equal to 1000 (Top), 1500 and 2000 (bottom). Flocking and Streaming Appear when the Number of Particles is Low Depending on the Initial Condition (top Left and Top Right (resp.)). Whereas only Stream Emerges when the Number of Particles is Higher, N Equal 1500 and 2000 (Bottom - Left and Right (resp.)). We Color Code the Cells in Blue or Red Depending on the Direction in Comparison to the Nematic Average (see Appendix 3.7.2)

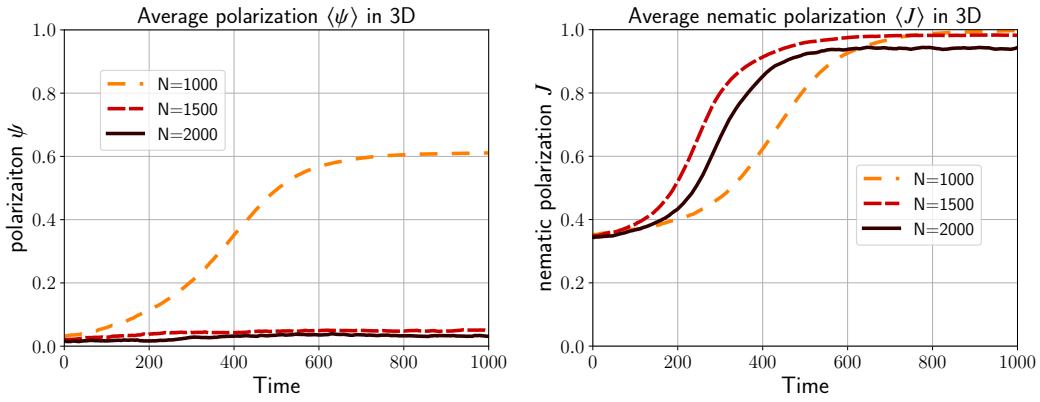


Figure 3.16: The Evolution of the Polarization ψ and the Nematic Polarization J for the Simulations Presented in Figure 3.15. When the Number of Particle is Low ($N = 1000$), Flock and Stream Emerge Depending on the Initial Condition Leading to an Average Value for ψ is around 0.6 and J Close to 1. However, for Larger Density ($N = 1500$ and $N = 2000$), only Streams Emerge since the Polarization ψ is Low and the Nematic Polarization J is Close to 1

3.7 Appendix

3.7.1 Explicit Expression of the Agent-Based Model

Model in $\mathbb{R}^2$

We would like to provide an explicit expression of the dynamics (3.2.4)-(3.2.5) by computing explicitly the gradient of the potential $\nabla_{\mathbf{x}_i} V_i$ and $\nabla_{\omega_i} V_i$. We first introduce the local coordinates:

$$x_{ij} = \langle \mathbf{x}_j - \mathbf{x}_i, \omega_i \rangle, \quad y_{ij} = \langle \mathbf{x}_j - \mathbf{x}_i, \omega_i^{\perp} \rangle \tag{3.7.1}$$

where $\omega_i^{\perp}$ is the orthogonal vector of ω_i. Simple computations show that:

$$\nabla_{\mathbf{x}_i} x_{ij} = -\omega_i \ , \ \nabla_{\mathbf{x}_i} y_{ij} = -\omega_i^{\perp} \ , \ \nabla_{\omega_i} x_{ij} = \mathbf{x}_j - \mathbf{x}_i \ , \ \nabla_{\omega_i} y_{ij} = -(\mathbf{x}_j - \mathbf{x}_i)^{\perp}.$$

We deduce that:

$$\begin{aligned}\nabla_{\mathbf{x}_i} V_i &= \sum_{j\neq i} \Phi'(r_{ij}^2)\nabla_{\mathbf{x}_i} r_{ij}^2 = 2\sum_{j\neq i} \Phi'(r_{ij}^2)\left(\frac{x_{ij}\nabla_{\mathbf{x}_i} x_{ij}}{|a|^2} + \frac{y_{ij}\nabla_{\mathbf{x}_i} y_{ij}}{|b|^2}\right)\\ &= -2\sum_{j\neq i} \Phi'(r_{ij}^2)\left(\frac{x_{ij}}{|a|^2}\,\omega_i + \frac{y_{ij}}{|b|^2}\,\omega_i^{\perp}\right), \end{aligned} \tag{3.7.2}$$

which provides an explicit expression to estimate $\mathbf{x}_i'$ in (3.2.4). To estimate ω_i' in (3.2.5), we compute similarly:

$$\begin{aligned}\nabla_{\omega_i} V_i &= 2\sum_{j\neq i} \Phi'(r_{ij}^2)\left(\frac{x_{ij}\nabla_{\omega_i} x_{ij}}{|a|^2} + \frac{y_{ij}\nabla_{\omega_i} y_{ij}}{|b|^2}\right)\\ &= 2\sum_{j\neq i} \Phi'(r_{ij}^2)\left(\frac{x_{ij}(\mathbf{x}_j-\mathbf{x}_i)}{|a|^2} - \frac{y_{ij}(\mathbf{x}_j-\mathbf{x}_i)^{\perp}}{|b|^2}\right).\end{aligned}$$

Using the explicit expression of the projection operator $P_{\omega_i^{\perp}} = \mathrm{Id} - \omega_i \otimes \omega_i$, we deduce:

$$P_{\omega_i^{\perp}}(\mathbf{x}_j - \mathbf{x}_i) = y_{ij}\cdot\omega_i^{\perp} \quad \text{and} \quad P_{\omega_i^{\perp}}\Big((\mathbf{x}_j-\mathbf{x}_i)^{\perp}\Big) = x_{ij}\cdot\omega_i^{\perp}. \tag{3.7.3}$$

Thus,

$$\begin{aligned}P_{\omega_i^{\perp}}\left(\nabla_{\omega_i} V_i\right) &= 2\sum_{j\neq i} \Phi'(r_{ij}^2)\left(\frac{x_{ij}y_{ij}\cdot\omega_i^{\perp}}{|a|^2} - \frac{y_{ij}x_{ij}\cdot\omega_i^{\perp}}{|b|^2}\right)\\ &= 2\sum_{j\neq i} \Phi'(r_{ij}^2)x_{ij}y_{ij}\left(\frac{1}{|a|^2} - \frac{1}{|b|^2}\right)\omega_i^{\perp}\\ &= -\frac{2e^2}{b^2}\sum_{j\neq i} \Phi'(r_{ij}^2)x_{ij}y_{ij}\,\omega_i^{\perp}. \end{aligned} \tag{3.7.4}$$

Model in higher dimension

In higher dimensions, i.e. $\mathbb{R}^3$, we cannot define anymore the local coordinate y_{ij} since $\omega_i^\perp$ is now an hyperplane. However, we can still define $x_{ij} = \langle \mathbf{x}_j - \mathbf{x}_i, \omega_i \rangle$ and we will use the second formulation for r_{ij} (3.2.2). Some elementary computations show that:

$$\begin{aligned}
\nabla_{\mathbf{x}_i} \Big[(\mathbf{x}_j - \mathbf{x}_i) \cdot \omega_i \Big]^2 &= 2 \Big[(\mathbf{x}_j - \mathbf{x}_i) \cdot \omega_i \Big] (-\omega_i) = -2 x_{ij} \omega_i, \\
\nabla_{\omega_i} \Big[(\mathbf{x}_j - \mathbf{x}_i) \cdot \omega_i \Big]^2 &= 2 \Big[(\mathbf{x}_j - \mathbf{x}_i) \cdot \omega_i \Big] (\mathbf{x}_j - \mathbf{x}_i) = 2 x_{ij} (\mathbf{x}_j - \mathbf{x}_i).
\end{aligned}$$

We deduce:

$$\begin{aligned}
\nabla_{\mathbf{x}_i} V_i &= \sum_{j \neq i} \Phi'(r_{ij}^2) \nabla_{\mathbf{x}_i} r_{ij}^2 \\
&= \frac{1}{b^2} \sum_{j \neq i} \Phi'(r_{ij}^2) \left(\nabla_{\mathbf{x}_i} \|\mathbf{x}_j - \mathbf{x}_i\|^2 - e^2 \nabla_{\mathbf{x}_i} \Big[(\mathbf{x}_j - \mathbf{x}_i) \cdot \omega_i \Big]^2 \right) \\
&= \frac{1}{b^2} \sum_{j \neq i} \Phi'(r_{ij}^2) \Big(2(\mathbf{x}_i - \mathbf{x}_j) + 2 e^2 x_{ij} \omega_i \Big).
\end{aligned}$$

Similarly, we have:

$$\begin{aligned}
\nabla_{\omega_i} V_i &= \sum_{j \neq i} \Phi'(r_{ij}^2) \nabla_{\omega_i} r_{ij}^2 \\
&= \frac{1}{b^2} \sum_{j \neq i} \Phi'(r_{ij}^2) \left(\nabla_{\omega_i} \|\mathbf{x}_j - \mathbf{x}_i\|^2 - e^2 \nabla_{\omega_i} \Big[(\mathbf{x}_j - \mathbf{x}_i) \cdot \omega_i \Big]^2 \right) \\
&= \frac{2e^2}{b^2} \sum_{j \neq i} \Phi'(r_{ij}^2) x_{ij} (\mathbf{x}_i - \mathbf{x}_j).
\end{aligned}$$

Using that the projector operator $P_{\omega_i^\perp} = \mathrm{Id} - \omega_i \otimes \omega_i$, we deduce:

$$P_{\omega_i^\perp}(\mathbf{x}_j - \mathbf{x}_i) = (\mathbf{x}_j - \mathbf{x}_i) - [(\mathbf{x}_j - \mathbf{x}_i) \cdot \omega_i] \cdot \omega_i = (\mathbf{x}_j - \mathbf{x}_i) - x_{ij} \cdot \omega_i$$

Thus,

$$P_{\omega_i^\perp} (\nabla_{\omega_i} V_i) = -\frac{2e^2}{b^2} \sum_{j \neq i} \Phi'(r_{ij}^2) x_{ij} \left((\mathbf{x}_j - \mathbf{x}_i) - x_{ij} \cdot \omega_i \right).$$

Notice that in any dimension, if the cell i has a circular shape (i.e. $a = b$ and the eccentricity becomes $e = 0$), then $\dot{\omega}_i = -\beta P_{\omega_i^\perp} (\nabla_{\omega_i} V_i) = 0$. Thus, the dynamics will have no effect on the orientation of the cell.

3.7.2 Nematic Average Velocity

Flocking pattern is primary described using average velocity from which we deduce the polarization of the flock. Denoting $(\omega_i)_{i=1..N}$ the velocity of all the cells, the average velocity $\langle\omega_i\rangle$ and polarization ψ are given by:

$$\langle\omega_i\rangle_i = \frac{1}{N}\sum_{i=1}^{N}\omega_i \quad \text{and} \quad \psi = |\langle\omega_i\rangle_i|.$$

However, for streaming formation, it is more difficult to define the average *nematic velocity* or the average *nematic direction*. With this aim, we need to introduce an optimization problem. For instance, the direction of the average velocity can be seen as the maximization of:

$$\max_{\Omega\in\mathbb{S}^{d-1}} \frac{1}{N}\sum_{i=1}^{N}\langle\omega_i,\Omega\rangle.$$

By Cauchy-Schwarz inequality, the maximizer Ω_* must be in the direction of the average $\langle\omega_i\rangle_i$. We would like to define similarly the average nematic direction as the direction Ω_{nem} that maximizes the function:

$$J(\Omega) = \frac{1}{N}\sum_{i=1}^{N}\langle\omega_i,\Omega\rangle^2. \tag{3.7.1}$$

Notice in particular that $J(\Omega) = J(-\Omega)$.

Nematic average in $\mathbb{R}^2$

Denote θ_i the angle between the velocity ω_i of the cell i and the horizontal axis. Also, define the angle $\bar{\theta}$ of the maximizer Ω_{nem} with respect to the x-axis. Notice that:

$$\max_{\Omega\,\in\,\mathbb{S}^1} J(\Omega) \;=\; \max_{\Omega\,\in\,\mathbb{S}^1}\frac{1}{N}\sum_{i=1}^{N}\langle\omega_i,\,\Omega\rangle^2 = \max_{\theta}\frac{1}{N}\sum_{i=1}^{N}\cos^2\left(\theta_i-\theta\right).$$

To maximize J, we must solve $\frac{\partial J(\bar{\theta})}{\partial \theta} = 0$. We find:

$$\begin{aligned}\frac{\partial J(\bar{\theta})}{\partial \theta} &= \frac{1}{N}\sum_{i=1}^{N} 2\cos\left(\theta_i - \bar{\theta}\right)\sin\left(\theta_i - \bar{\theta}\right) = \frac{1}{N}\sum_{i=1}^{N}\sin\left(2\left(\theta_i - \bar{\theta}\right)\right) \\ &= \frac{1}{N}\sum_{i=1}^{N}\left[\sin\left(2\theta_i\right)\cos\left(2\bar{\theta}\right) - \cos\left(2\theta_i\right)\sin\left(2\bar{\theta}\right)\right] \\ &= \langle\sin\left(2\theta_i\right)\rangle_i \cos\left(2\bar{\theta}\right) - \langle\cos\left(2\theta_i\right)\rangle_i \sin\left(2\bar{\theta}\right).\end{aligned}$$

Therefore, $\frac{\partial J(\bar{\theta})}{\partial \theta} = 0$ leads to:

$$\frac{\langle\sin\left(2\theta_i\right)\rangle_i}{\langle\cos\left(2\theta_i\right)\rangle_i} = \frac{\sin\left(2\bar{\theta}\right)}{\cos\left(2\bar{\theta}\right)}.$$

This motivates the following definition.

Definition 6 *The nematic average direction Ω_{nem} is defined as:*

$$\Omega_{nem} = (\cos\bar{\theta},\, \sin\bar{\theta}) \qquad \textit{with} \quad \bar{\theta} = \frac{1}{2}\arctan\left(\frac{\langle\sin\left(2\theta_i\right)\rangle_i}{\langle\cos\left(2\theta_i\right)\rangle_i}\right). \tag{3.7.2}$$

Remark 3.7.1 *If we denote the nematic average:*

$$\mathbf{u}_{nem} = \left(\langle\cos(2\theta_i)\rangle_i,\, \langle\sin(2\theta_i)\rangle_i\right)^T \tag{3.7.3}$$

the nematic polarization defined in (3.3.2) *is given by $\gamma = |\mathbf{u}_{nem}|$. This formula is similar to the expression of the usual polarization ψ giving as the norm of the average velocity $|\langle\omega_i\rangle_i|$.*

Nematic polarization in higher dimension

Note that the previous section is only applicable in $\mathbb{R}^2$ since we use polar coordinates. To find the maximizer Ω_* of J (3.7.1) in higher dimension, we introduce the Lagrangian:

$$\mathcal{L}(\Omega, \lambda) = J(\Omega) - \lambda g(\Omega)$$

with $g(\Omega) = |\Omega|^2 - 1$. The maximizer Ω_{nem} must be a critical point to $\mathcal{L}$ which leads to the existence of a Lagrange multiplier λ_* satisfying:

$$\begin{aligned} \nabla_\Omega J(\Omega_*) = \lambda_* \nabla_\Omega g(\Omega_*) \quad &\Rightarrow \quad \frac{2}{N}\sum_{i=1}^{N} \langle \omega_i, \Omega_* \rangle \omega_i = 2\lambda_* \Omega_* \\ &\Rightarrow \quad \frac{1}{N}\sum_{i=1}^{N} [\omega_i \otimes \omega_i]\Omega_* = \lambda_* \Omega_*. \end{aligned}$$

Denoting the matrix $A = \frac{1}{N}\sum_{i=1}^{N}[\omega_i \otimes \omega_i]$, we deduce that λ_* and Ω_* are eigenvalue/eigenvector of A. Therefore, J is maximized at the eigenvector $\Omega_\star$ that is associated to the largest eigenvalue of the matrix A.

Remark 3.7.2 *A is a positive semi-definite matrix, so the maximized quantity J is recognized as a Rayleigh quotient (Horn and Johnson, 2012).*

Remark 3.7.3 *In dimension larger than 2, the nematic average vector $\mathbf{u}_{nem}$ (3.7.3) is no longer defined. Thus, we cannot define nematic polarization as $\gamma = |\mathbf{u}_{nem}|$. Instead, we can use as statistics the value of J at the maximizer Ω_{nem}. Similarly, we find that $J(\Omega_{nem})$ is close to 1 when the vectors form a stream.*

Moreover, in $\mathbb{R}^2$, we observe that the nematic polarization γ and $J(\Omega_{nem})$ are related to each other since:

$$\begin{aligned} J(\Omega_*) &= \frac{1}{N}\sum_{i=1}^{N} \langle \omega_i, \Omega \rangle^2 = \frac{1}{N}\sum_{i=1}^{N} \cos^2(\theta_i - \bar{\theta}) = \frac{1}{2N}\sum_{i=1}^{N} \Big(1 + \cos\Big(2(\theta_i - \bar{\theta})\Big)\Big) \\ &= \frac{1}{2} + \frac{1}{2}\langle \cos(2(\theta_i - \bar{\theta}))\rangle_i = \frac{1}{2} + \frac{1}{2}\Big[\langle \cos(2\theta_i)\rangle_i \cos(2\bar{\theta}) + \langle \sin(2\theta_i)\rangle_i \sin(2\bar{\theta})\Big] \\ &= \frac{1}{2} + \frac{1}{2}|\mathbf{u}_{nem}| \end{aligned}$$

since $\mathbf{u}_{nem}$ and the vector $(\cos 2\bar{\theta}, \sin 2\bar{\theta})^T$ are parallel. Thus:

$$J(\Omega_*) = \frac{1}{2} + \frac{1}{2}\gamma \quad \Rightarrow \quad \gamma = 2J(\Omega_*) - 1.$$

Therefore, using γ or $J(\Omega_*)$ provide similar information about the nematic alignment of the cells. However, the advantage of $J(\Omega_*)$ is to generalize in any dimensions.

Chapter 4

NECROTIC HYPOTHESIS INVESTIGATION

4.1 Introduction

Glioblastoma Multiforme (GBM) is the most aggressive and highly invasive primary brain tumor in adults (Boyle *et al.*, 2008). It is a high grade stage IV cancer, and accounts for 60% to 70% of all malignant glioma (Baker *et al.*, 2014b; Alvord, 1995; Harpold *et al.*, 2007). Despite aggressive treatment, nearly all cases recur after six to seven months. The median of survival is approximately twelve to fifteen months.

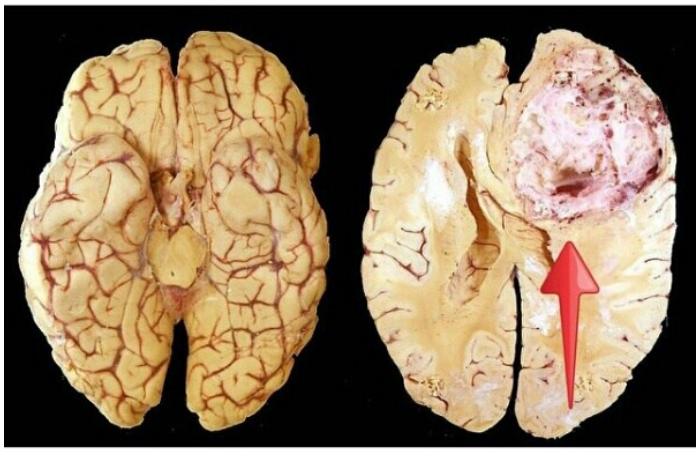


Figure 4.1: The Frontal Lobe of a Brain Showing the Most Aggressive Brain Tumor, Glioblastoma Multiforme (GBM)

GBM is known for two terms, the proliferation and the invasion (Harpold *et al.*, 2007). The proliferation is defined in terms of aggressiveness of the rapid evolution and growth of cancer cells, and the invasion, in terms of the high velocity of the

disease spread via movement.

Since the aggressiveness and spread of GBM usually differ in an extremely heterogeneous manner across patient, we will study the net rates of diffusion D and proliferation ρ of different patients. We will start by introducing two models, PI and PIHNA, that we used during our study to calculate the D and ρ terms. Then, we will explain the six types of MRI that we will use to calculate the invisibility index (D/ρ). Under the PI model, the velocity will asymptotically approach the radial growth velocity $v = 2\sqrt{D\rho}$ by using two pretreatments imaging points. By combining, these two equations with two unknowns, we will be able to calculate D and ρ.

The usual methods require two time points imaging to estimate the velocity. By having only a single time point of imaging, we can only estimate D/ρ, so our hopes that we could calibrate both D and ρ from one single day. Hence, our necrotic hypothesis, was to see if the size of necrosis relative to the rest of tumor can be used, as a surrogate of the velocity ,to estimate the patient-specific rates of net proliferation ρ and invasion D.

4.2 Models

4.2.1 Proliferation Invasion (PI) Model

The mathematical model for the PI model is the regular reaction diffusion equation, (Kot, 2001):

$$\overbrace{\frac{\partial n}{\partial t}}^{\text{Rate of change of glioma cell density}} = \overbrace{\nabla \cdot (D(x)\nabla n)}^{\text{Net diffusion}} + \overbrace{\rho n(1 - \frac{n}{k})}^{\text{Net proliferation}} \tag{4.2.1}$$

The PI model describes the rate of change of the concentration of the glioma cells $n = n(x,t)$ in position x at time t ($cells/mm^3$), based on the two net rates: the invasion or diffusion (D) ($mm^2/year$) and the proliferation or growth (ρ) ($1/year$).

k is the tumor cell carrying capacity ($cells/mm^3$). The ratio D/ρ is related to the degree of invasion which is also called the invisibility index. Since PI model is portrays of the spread of the glioma, so we will be looking for a traveling wave solution which in this case, the model will be similar to the Fisher's equation. This yields to the same predictable pattern for Fisher's equation of linear radial growth that we usually observe in low and high grade of gliomas. Hence, the velocity of growth predicated by this equation related to the velocity of radial growth Kot (2001) will be $v = 2\sqrt{D\rho}$ (Appendix 4.6.1).

In the PI model, the tumor cells proliferate at a constant rate which is independently of the availability of nutrients and the cell population is the same type. In the next section, the PIHNA model is built upon the logical extension of the PI model.

4.2.2 Proliferation Invasion Hypoxia Necrosis Angiogenesis (PIHNA) Model

The PIHNA model incorporates nutrient availability and oxygenation by using the vasculature (v) and angiogenic (a) (which is the development of new blood vessels) factors. In this model, the tumor cell populations are partitioned into three different classes which are the well-oxygenated, the deficiency of oxygen, and the dead cells, called respectively, the normoxic (c), hypoxic (h) and necrotic (n) tumors cells.

The mathematical model of PIHNA is represented with these equations:

$$\begin{aligned}\frac{\partial c}{\partial t} = & \overbrace{\nabla \cdot (D(1-T)\nabla c)}^{\text{Net dispersal of normoxic glioma cells}} + \overbrace{\rho c(1-T)}^{\text{Net proliferation of normoxic glioma cells}} \\ & + \overbrace{\gamma h V}^{\text{Conversion of hypoxic to normoxic}} - \overbrace{\beta c(1-V)}^{\text{Conversion of normoxic to hypoxic}} \\ & - \overbrace{\alpha_n n c}^{\text{Conversion of normoxic to necrotic}}\end{aligned} \tag{4.2.2a}$$

$$\begin{aligned}\frac{\partial h}{\partial t} = & \overbrace{\nabla \cdot (D(1-T)\nabla h)}^{\text{Dispersal of hypoxic glioma cells}} - \overbrace{\gamma h V}^{\text{Conversion of hypoxic to normoxic}} \\ & + \overbrace{\beta c(1-V)}^{\text{Conversion of normoxic to hypoxic}} - \overbrace{\alpha_h h(1-V) + \alpha_n n h}^{\text{Conversion of hypoxic to necrotic}}\end{aligned} \tag{4.2.2b}$$

$$\frac{\partial n}{\partial t} = \overbrace{\alpha_h h(1-V) + \alpha_n n(c+h+v)}^{\text{Conversion of hypoxic, normoxic and vasculature to necrotic}} \tag{4.2.2c}$$

$$\begin{aligned}\frac{\partial v}{\partial t} = & \overbrace{\nabla \cdot (D_v(1-T)\nabla v)}^{\text{Dispersal of vasculature}} + \overbrace{\mu \frac{a}{K_m + a} v(1-T)}^{\text{Net proliferation of vasvulature}} \\ & - \overbrace{\alpha_n n v}^{\text{Conversion of vasculature to necrotic}}\end{aligned} \tag{4.2.2d}$$

$$\begin{aligned}\frac{\partial a}{\partial t} = & \overbrace{\nabla \cdot (D_a \nabla a)}^{\text{Diffusion of angiogenic factors}} + \overbrace{\delta_c c + \delta_h h}^{\text{Net production of angiogenic factors}} \\ & - \overbrace{q\mu \frac{a}{K_m + a} v(1-T) - \omega a v}^{\text{Net consumption of angiongenic factors}} - \overbrace{\lambda a}^{\text{Decay of angiogenic factors}}\end{aligned} \tag{4.2.2e}$$

where the vessel efficiency is $V = \frac{v}{(c+h+v)}$ and $T = \frac{(c+h+v+n)}{k}$, such that k is the carrying capacity (cells/mm^3). In addition, the normoxic cells proliferate at a rate ρ (1/year) and invade at a rate D (mm^2/year), while the hypoxic cells only invade and necrotic cells which are considered dead just take up place. Also, α_n (1/year) is the rate at which the cells undergo necrosis when in contact with necrotic cells, γ (1/year) and β(1/year) are the maximum conversion rates between the hypoxic to normoxic cells and vise versa respectively. When the oxygen levels fall too low, the hypoxic cells will become a necrotic cells at a rate α_h (1/year). Both hypoxic

and normoxic cells produce some tumor angiogenic factors at a rate of δ_h (1/year) and δ_c (1/year), respectively, with $\delta_h > \delta_c$. The vasculature cells dispersal rate is D_v (mm^2/year) and proliferation rate is μ (1/year). The angiogenic factors are consumed by the vasculature for both vasculature proliferation with rate q ((1/year) and for regular vasculature maintenance with rate ω (1/year). Once the angiogenic factors are produced by cells, they disperse in the tissue at a rate D_a (mm^2/year) and decay at a rate λ (1/year). For more details about this model, a good referral will be (Swanson *et al.*, 2011).

This mathematical PIHNA model has seventeen unknown parameters, which are difficult to estimate and may be estimated separately at best. The values of the parameters in our case will be derived from the work done by Levine *et al.* (2001, 2011); Mac Gabhann and Popel (2004); Serini *et al.* (2003); Sherratt and Murray (1990). We will be using some previous work that showed that a marked quantitative heterogeneity in the net rates of proliferation ρ and invasion D, within the grade IV of GBM (Harpold *et al.*, 2007).

4.3 Estimating Diffusion Coefficient D and Proliferation Rate ρ

The Proliferation Invasion (PI) model is a Patient-Specific virtual control model that we are aware of which provide some clinically valuable metrics for diffusion and treatment response. The two net parameters of invasion D and proliferation ρ can be calculated for an individual patient by providing only two time point imaging pre-treatment of Magnetic Resonance Imaging (MRIs).

4.3.1 Magnetic Resonance Imaging (MRI) of the Bain

There are six types of MRI that we used during our process which are T1-weighted, T1-Gd, T2-weighted, fluid-attenuated inversion-recovery (FLAIR), T0 and T0-Gd.

T1-weighted MRI imaging is usually infused with Gadolinium (Gd) which is a non-toxic paramagnetic contrast enhancement agent. Gd enhanced a bright white signal in the blood vessels which makes it easier to visualize the MRIs. In the case where we add Gd, the MRI of T1 will be called T1-Gd. When injecting Gd during a scan on T1-weighted images, the Gd will pass through the blood-barrier barium (BBB) and will breakdown in the vasculature structure. Usually, in a normal brain, there is no infiltration of Gd through out BBB. But in a malignant tumor, due to the immature and malformed of the blood vessels in the tumor, the BBB will break and the gadolinium will leak into the surrounding tissue.

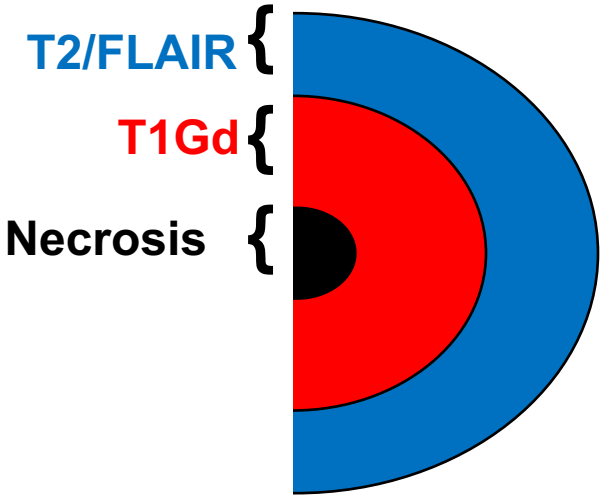


Figure 4.2: Different Proportions of MRI Abnormalities Depending on the Tumor's Metabolic Demands

According to Mayo-Clinic, T0 and T0-Gd are some measurements that are not usually used outside the Lab. They assume that the dark center surrounding within the T1-Gd regions will show the necrotic tissue which is the T0 and in case of brightness it will be T0-Gd.

Once BBB is broken and the blood vessels are malformed, the water starts in-

creasing around the cells and bypassing into the tissue, as edema. T2-weighted (T2) and FLAIR MRI will reveal the increase of the concentration of water or edema which is contained around the tumor cells. Note that T2 and FLAIR images are both indicating edema, so we treat them as interchangeable with preference given to FLAIR. Therefore, the different proportions of MRI will be resumed as see in figure 4.2.

All these different types of MRI will be used to give us the radii of the sizes of the tumors, which will be needed in the next section to calculate D and ρ from the PI model.

4.3.2 Patient-Specific Virtual Controls

The velocity used by the data was the average velocity:

$$v_{ave} = \bar{v} = \frac{R(t_2) - R(t_1)}{t_2 - t_1} = \frac{1}{t_2 - t_1} \int_{t_1}^{t_2} v(s)ds \tag{4.3.3}$$

where R is the mean size of the tumor in millimeter (mm) at time t_i in years. The two velocities needed to be calculated were T1-Gd and T2/FLAIR. If the size of the tumor was estimated using FLAIR for first and last day for the patient, we will calculate the velocity using only FLAIR measurement, in other words we did not use T2 measurement. On the other hand, if we have only access to one measurement with FLAIR, we used T2 to obtain the measurement on the other day, and we will ignore the velocity in this case. If we do not have access to any measurements with FLAIR on the first and last day, but we do with T2, then we will calculate the velocity using only T2 measurement. Notice that we always chose the first and last day for the MRI sizes and volumes, ignoring measurement in between these two dates.

Once we have an estimation of the average velocity (using both T1-Gd and T2/FLAIR) for each patient, we equated our result to the previous velocity of the PI

model writing:

$$v_{ave} = 2\sqrt{D\rho}. \tag{4.3.4}$$

Since we have two estimations of v_{ave} (using either T1-Gd or T2/FLAIR), we have two possible choices for equation (4.3.4). Note also that this equation is not sufficient to estimate D and ρ. We need another equation.

Based on previous results from PNT Lab, a specific function called DpCal was created to calculate the invisibility index D/ρ by using two different MRI sizes which are T1-Gd and T2/FLAIR images:

$$\frac{D}{\rho} = \text{DpCal}(R_1, R_2). \tag{4.3.5}$$

Combining (4.3.4) and (4.3.5), we obtain an estimation of D and ρ for each patient.

Since we had two different dates for the MRI (first and last day), we had two estimation for D/ρ using either first or last day. We ended up with two equations which are T1-Gd or T2/FLAIR velocities and D/ρ from the first day and from last day with two unknowns for each patient. For each patient, it turn out to be up-to four possible combinations for D and ρ from the PI model.

4.4 Necrotic Investigation

4.4.1 Necrotic Hypothesis

Based on some previous results Swanson (2000), Dalrymple *et al.* (1994), Kelly (1993), Kelly *et al.* (1987), the density of cells at the edge of the tumor as seen on T1-Gd and T2/FLAIR images correspond to an 80 % and 16% of the carrying capacity k respectively (see figure 4.3). Therefore, the relationship between these two radii describes the steepness of the tumor cell profile or the shape of the traveling wave which is related to D/ρ (fig 4.3).

While we can get the net invasion rate D and the net proliferation rate ρ for patients with **two** time points imaging by combining the two equations (4.3.4) and (4.3.5), we can not get both D and ρ if there is only **one** time point of imaging since we can not estimate the radial velocity. In this case, we will only be able to find the shape of the wave profile by using equation (4.3.5). By looking at figure 4.4, we can see that it is possible to have tumors with different velocities (slow or fast) but with the same shape. Since the size of the necrosis found by T0 image is less than the size of the tumor calculated in T1-Gd (see figure 4.2), then we can presume that the necrosis is located on the hill of the wavefront. Therefore, the necrotic hypothesis is to try to find if the size of the tumor T0 in combination with the other tumor sizes (see figure 4.4) will at a single time point serve as a surrogate for the radial velocity so that we can calculate D and ρ. So, we could hopefully say of tumors with the same shape, but different sizes of necrosis, which one was growing faster.

To investigate the necrotic hypothesis, we were given some data, that will help us explore more some alternative options.

4.4.2 Data Given

The data given were all pre-treatments and approved prior to May 4th 2016. It included 219 unique patients, with an identity patient number to each one and with 563 unique times imaging for all of them. Each patient had different types of MRI taken on different dates. Note that we considered that all the volumetric measurements (mm^3) of the MRIs given are converted to spherically so that the radii sizes of the tumors needed was calculated by using the spherical volume:

$$\text{Volume} = V = \frac{4}{3}\pi r^3 \tag{4.4.6}$$

Based, on the data given, there were always three patients off limit that were

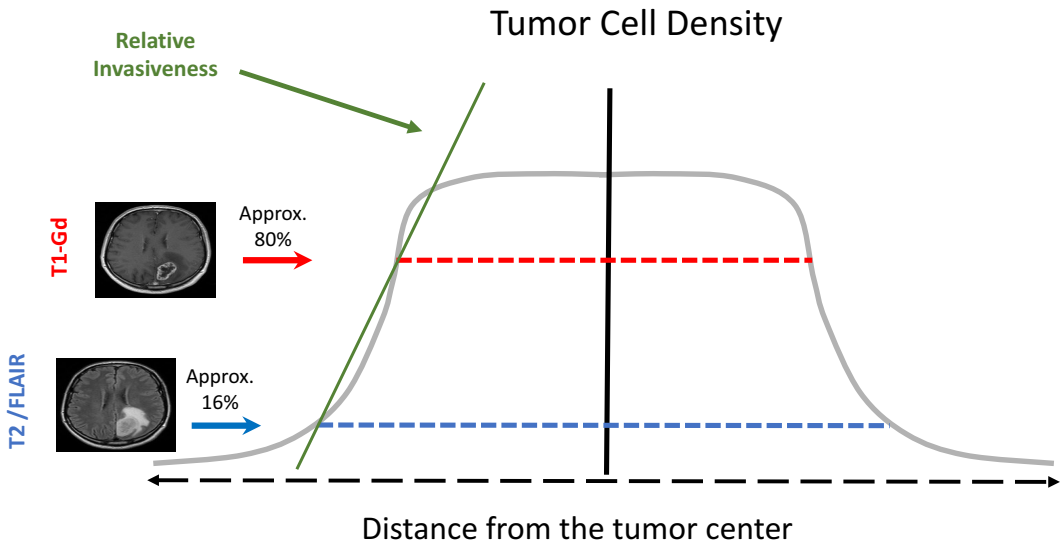


Figure 4.3: Pre-Treatments Time Points Images of T1Gd and T2/FLAIR, Used to Approximate the Correspond to an 80 % and 60% Tumor Cell Density Threshold Respectively. The Steepness of the Wave can be Estimated by Calculating the Relative Invasiveness D/ρ by using T1Gd and T2/FLAIR Images

ruining the linear regression lines. One of them, the tumor was growing around 2500 $mm/year$. We did study more the necrotic hypothesis by examining and studying different plots that will be discussed in the next section.

4.4.3 Data Analysis

At first, we tried to find a relationship between the velocity using either T1-Gd or T2/FLAIR and the size of T0-Gd in combination with the rest of the tumors which are T0-Gd size, T0-Gd size over T1-Gd size, and T0-Gd size over T2/FLAIR size which were taken on the first and last day. Plotting these first graphs were essentially seeing if it would be as simple as the velocity would be related to something of the

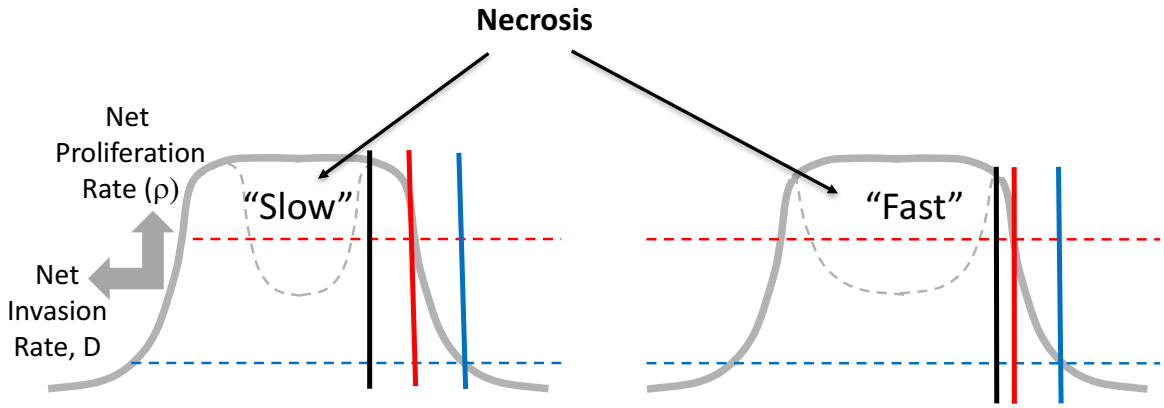
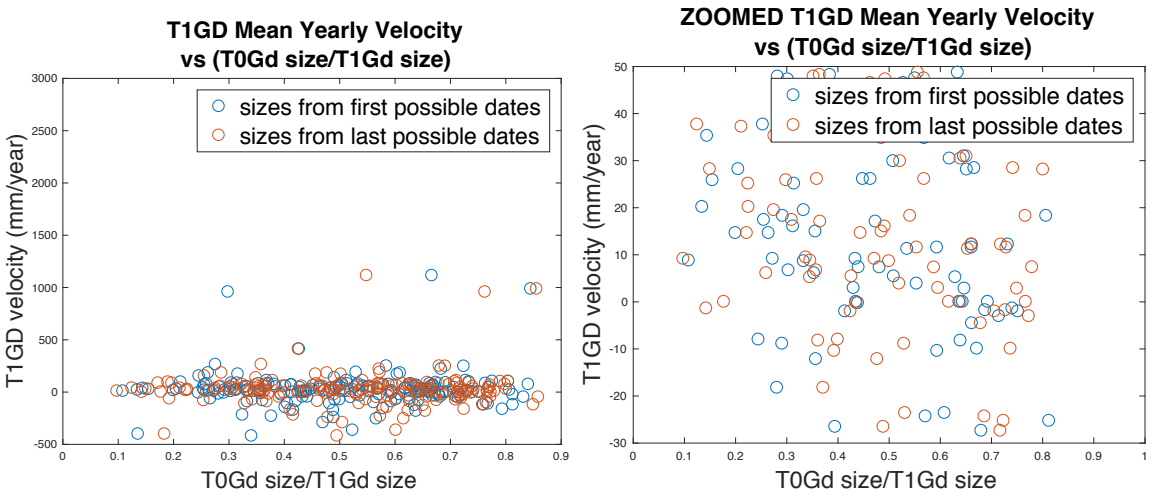


Figure 4.4: The Edge of T0 Image Gives the Size of the Tumor (Necrosis) in Relative to the Rest of the Tumor can be Used to Substitute the Velocity of the Wavefront (as Slow or Fast), so we can Estimate the Patient Specific Rates of Net Proliferation ρ and Invasion D. The Colors are Related to the MRI Images in Figure 4.2

necrosis directly. By looking at all these graphs, there was not an obvious predictive pattern. We display one example of these graphs, in figure 4.5, the plot of the velocity using T1-Gd versus (T0-Gd size/T1-Gd size) on first and last day. As we can observe that few patients were off limit in the figure (a), so we had to zoom the graph to be able to see in figure (b) if there is any strong correlation.

Then, we wondered if maybe the predictive relationship would be different depending on the shape of the tumor which is given by D/ρ. After examining, the 24 scatter plots between D/ρ and the velocity where the colors of the dots were based on the associated sizes, we were not able to find any strong pattern. In figure 4.6, we demonstrated one of the scatters, which shows D/ρ from last day versus (the velocity using T1Gd) and the color dots are (T0Gd size/T1Gd size from last day). We didn't see any strong correlation, in the scatters as well. Note that, some white dots are shown in figure 4.6, which means that T0-Gd size is not measured during that day, or T0-Gd size and T1-Gd size are taken on different dates so we can't find the ratio



(a) Regular plot

(b) Zoomed graph

Figure 4.5: The Velocity Using T1Gd vs (T0Gd size/T1Gd size) Taken on First and Last day

of T0-Gd/T1-Gd size.

In other words, the first set of plots were seeing if a single variable regression which is the associated size in our case would be enough for the velocity. The second set was seeing if we used two variables (D/ρ and the associated sizes) if we could find a good regression.

We also tried one step further, we keep the same scatters but we batched the T2/FLAIR sizes of the tumor in groups of 5 mm. Our prediction were that once we arrange the T2/FLAIR sizes in groups we tried to see if the predictive relationship actually needed a third variable which is the T2/FLAIR size. Again, we didn't see any pattern or strong correlation (see figure 4.7 which is a particular example where T2/FLAIR size were between 20 - 25 mm).

Therefore, they provided me a look up table by using the parameters values by previous work for the PIHNA model to calculate D and ρ for each patient. Our presumptions were that the sets of four combinations between D and ρ of the PI and

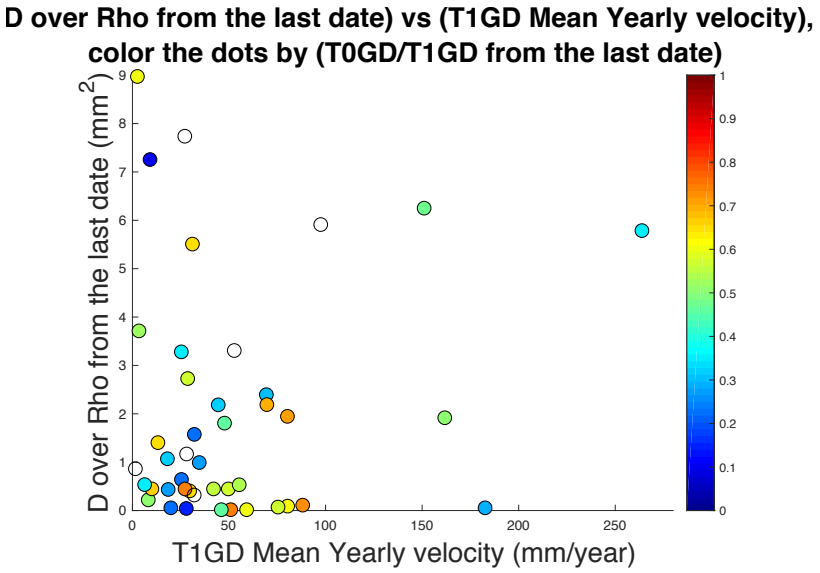


Figure 4.6: The Scatter Showing (D / ρ from Last Day) vs (the Velocity Using T1Gd) and the Color Dots are (T0Gd Size/T1Gd Size from Last Day)

PIHNA model were supposed to give us a positive correlation. Regrettably, we were not able to find a strong positive correlation between D or ρ of PI and PIHNA model. An example of one of these plots where graphed in figure 4.8, where it is noticeable that the linear regression does not show a strong correlation between D using the look up table (PIHNA model) and D from the previous explanation in PI model.

4.5 Conclusion

Our analysis reveal that the necrotic investigation using only a single day image was insufficient to estimate the diffusion rate D and the proliferation rate ρ. Our initial assumption was to surrogate the combination of the necrosis size with the rest of the tumors to estimate the radial velocity and deduce the rates D and ρ. However, our analysis did not show strong correlation between necrotic size and tumor velocity. This can be explained by the strong heterogeneity of cancer type among patients. Moreover, it would be helpful to perform an error analysis of the

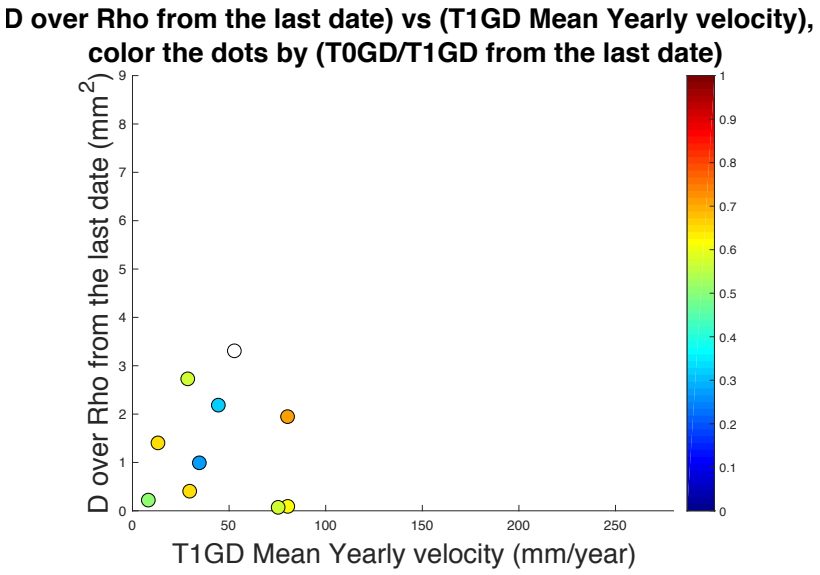


Figure 4.7: The Scatter Showing (D / ρ from Last Day) vs (the Velocity Using T1Gd) and the Color Dots are (T0Gd Size/T1Gd Size from Last Day) where T2/FLAIR Size is Between 20 and 25 mm

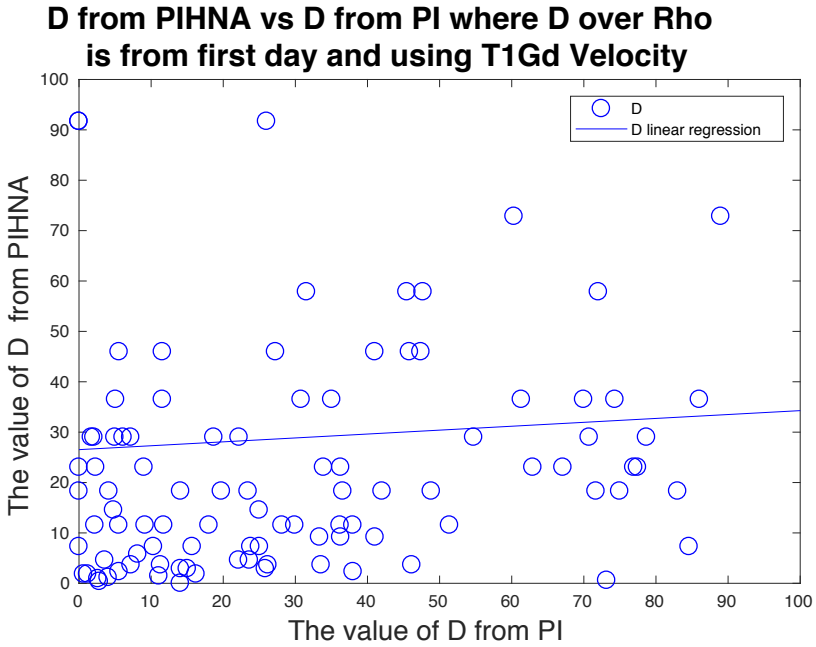


Figure 4.8: Linear Regression of D from PIHNA Model versus D in PI Model by Using D/ ρ from the First Day and the Velocity Using T1-Gd

same type of MRI measurements by comparing successive MRI taken on the same day for the same patient. We would have an estimation of the variation that might exceed the accuracy needed to estimate the rates D and ρ.

4.6 Appendix

4.6.1 Explicit Expression of the Velocity of the PI Model

Under the PI model, equation (4.2.1), the velocity should asymptotically approach to $v = 2\sqrt{D\rho}$. Assume D is a constant and x is in $\mathbb{R}$. To simplify equation (4.2.1), we will do some change of variables:

$$u = \frac{n}{k}, \qquad \tau = \rho t, \qquad y = \sqrt{\frac{\rho}{D}}x. \tag{4.6.7}$$

So we can reduce our PI equation to

$$\frac{\partial u}{\partial \tau} = \frac{\partial^2 u}{\partial y^2} + u(1-u). \tag{4.6.8}$$

We are trying to model the spread of cells, we will look at a traveling wave equation, where we will assume that the wave is moving to the right. Let $z = y - c\tau$ where c is the positive speed of the wave. Thus, $u(y,\tau) = \phi(y - c\tau) = \phi(z)$. By transforming our PDE to ODE, we will get:

$$\phi'' + c\phi' + \phi(1-\phi) = 0. \tag{4.6.9}$$

By writing the previous equation as a system, we will obtain the following:

$$\begin{cases} \phi' = w \\ w' = \phi'' = -cw - \phi(1-\phi). \end{cases} \tag{4.6.10}$$

Solving $\phi' = 0$, and $w' = 0$ leads to two equilibrium points (0,0) and (1,0). To study the stability of the equilibrium, denote the system (4.6.10) as:

$$\begin{pmatrix} \phi \\ w \end{pmatrix}' = F(\phi, w) \qquad \text{with} \quad F(\phi, w) = \begin{pmatrix} w \\ -cw - \phi(1-\phi) \end{pmatrix}.$$

Taking the Jacobian of F gives:

$$DF(\phi, w) = \begin{bmatrix} 0 & 1 \\ -1 + 2\phi & -c \end{bmatrix}.$$

At (1,0), the Jacobian will give a negative trace and negative determinant. By Routh-Hurwitz criterion, the equilibrium (1,0) is a saddle point. On the other hand, at (0,0), by solving the characteristic equation of the Jacobian which is $\lambda^2 + c\lambda + 1 = 0$, we will get

$$\lambda = \frac{-c \pm \sqrt{c^2 - 4}}{2}. \tag{4.6.11}$$

Therefore, if $0 < c < 2$, we have a stable focus, and if $c \geq 2$, so we have a stable node. To ensure that $\phi > 0$ (so $u > 0$), we need (0,0) to be a stable node. Hence, $c \geq 2$ is necessary condition for the travel wave. So the minimum wave speed for the traveling wave is at $c_{min} = 2$

$$2 = c_{min} = \frac{dy}{d\tau} = \frac{\sqrt{\frac{\rho}{D}}dx}{\rho dt} = \frac{1}{\sqrt{\rho D}}\frac{dx}{dt} \tag{4.6.12}$$

If the initial propagating wave with speed $c > 2$ is specified, the speed will change into the minimum value $c = 2$ (Gazdag and Canosa, 1974). Hence, $v = \frac{dx}{dt} = 2\sqrt{D\rho}$.

Chapter 5

CONCLUSION

The overall focus of this dissertation has been based on developing microscopic and macroscopic models to explain the observation of some biological in vivo experiment and to fit some natural science data. The application of different models was given in each chapter to illustrate the overall understanding of complex biological phenomena. Each chapter provides some unique contribution to the growing field of applied mathematical modeling.

The novelty of these chapters is the connection that links the global umbrella. Chapters 2 and 3 describe the development of ABMs of individual particles (ants and glioma cells (resp.)) that clarify the biological clarification of experiments. Whereas, chapter 4, illustrate how data can be used to calculate parameters; through PDE models, to try to find a strong correlation between two quantities. In addition, chapters 3 and 4, study brain tumor while using microscopic and macroscopic models (resp.).

In chapter 2, ant foragers are able to navigate a path by going back and forth from the nest to food source through a bridge. According to the data provided to us, one can observe that the evolution of the total flow on the bridge increases to reach a peak then starts to decrease slowly. Our main intention was to predict a model that will fit the data curve. We developed compartmental models with three stages. The latter model is an extended dynamic from the first one. We compare the shape of the curve of the number of foragers of the two models with the data. One can realize that the extended model will fit much better over time with the data than the previous one.

Chapter 3 focuses on modeling glioma cells which are responsive for brain tumor. Experimental observation show that these cells have different shapes, and this might lead to different macroscopic behaviors such as stream and flock formation. To explore how cancer cells manage to spread throughout the organism, an agent-based model was introduced to describe the essential features of the cell (e.g. motility, cell-cell interaction). Depending on the shape of the cells, the stream formation can increase and influence the overall tumor growth. We investigate how the density affects these pattern formation.

In chapter 4, glioblastoma is the most aggressive tumor that proliferates and spreads through the organism. The data that was provided to us helped us calculate the two coefficients of proliferation and invasion, with two time imaging, by using the reaction diffusion model, which help us calculate the velocity by using the traveling wave solution. Our main goal was to try to find if it is possible to estimate these two coefficients with one time imaging only. Therefore, based on some previous assumptions, we investigated that by trying to see if there is any correlation between the central size of the tumor and the velocity.

In addition, in the appendix, we explained the comparison of Verlet and Cell-linked methods. In our numerical code, in chapters 2 and 3, we used Cell-linked to calculate the number of neighbors around each ant, and glioma cell (resp.) within a specific radius. Verlet method evaluates the number of neighbors, by constructing a list of neighboring for each particles in the system. Instead of computing the distance between all the ants on every time steps, Cell-linked method will split the domain to cell grids and locate the particles in every cell. Then, it will calculate the number of neighbors only by counting the number of particles in the cell neighbors that are within the radius provided. Therefore, one can conclude, that with a large number of particles, and small radius, Cell-linked method will be more efficient to calculate

the number of neighbors.

Alvord, E., "Patterns of growth of gliomas.", American Journal of Neuroradiology **16**, 5, 1013–1017 (1995).

Anderson, A. R. and V. Quaranta, "Integrative mathematical oncology", Nature Reviews Cancer **8**, 3, 227 (2008).

Ariel, G., A. Shklarsh, O. Kalisman, C. Ingham and E. Ben-Jacob, "From organized internal traffic to collective navigation of bacterial swarms", New Journal of Physics **15**, 12, 125019 (2013).

Arrowsmith, D. and C. M. Place, *Dynamical systems: differential equations, maps, and chaotic behaviour*, vol. 5 (CRC Press, 1992).

Baker, G. J., V. N. Yadav, S. Motsch, C. Koschmann, A.-A. Calinescu, Y. Mineharu, S. I. Camelo-Piragua, D. Orringer, S. Bannykh, W. S. Nichols *et al.*, "Mechanisms of glioma formation: iterative perivascular glioma growth and invasion leads to tumor progression, vegf-independent vascularization, and resistance to antiangiogenic therapy", Neoplasia **16**, 7, 543–561 (2014a).

Baker, G. J., V. N. Yadav, S. Motsch, C. Koschmann, A.-A. Calinescu, Y. Mineharu, S. I. Camelo-Piragua, D. Orringer, S. Bannykh, W. S. Nichols *et al.*, "Mechanisms of glioma formation: iterative perivascular glioma growth and invasion leads to tumor progression, vegf-independent vascularization, and resistance to antiangiogenic therapy", Neoplasia **16**, 7, 543–561 (2014b).

Balagam, R., D. B. Litwin, F. Czerwinski, M. Sun, H. B. Kaplan, J. W. Shaevitz and O. A. Igoshin, "Myxococcus xanthus gliding motors are elastically coupled to the substrate as predicted by the focal adhesion model of gliding motility", PLoS computational biology **10**, 5, e1003619 (2014).

Ballerini, M., N. Cabibbo, R. Candelier, A. Cavagna, E. Cisbani, I. Giardina, V. Lecomte, A. Orlandi, G. Parisi, A. Procaccini *et al.*, "Interaction ruling animal collective behavior depends on topological rather than metric distance: Evidence from a field study", Proceedings of the national academy of sciences **105**, 4, 1232–1237 (2008).

Barnes, D. J. and D. Chu, *Guide to simulation and modeling for biosciences* (Springer, 2015).

Boyle, P., B. Levin *et al.*, *World cancer report 2008.* (IARC Press, International Agency for Research on Cancer, 2008).

Carrington, P. J., J. Scott and S. Wasserman, *Models and methods in social network analysis*, vol. 28 (Cambridge university press, 2005).

Chen, X., X. Dong, A. BeâĂŹer, H. L. Swinney and H. Zhang, "Scale-invariant correlations in dynamic bacterial clusters", Physical review letters **108**, 14, 148101 (2012).

Chib, S. and E. Greenberg, "Understanding the metropolis-hastings algorithm", The american statistician **49**, 4, 327–335 (1995).

Cho, H., H. Jönsson, K. Campbell, P. Melke, J. W. Williams, B. Jedynak, A. M. Stevens, A. Groisman and A. Levchenko, "Self-organization in high-density bacterial colonies: efficient crowd control", PLoS biology **5**, 11, e302 (2007).

Ciesielski, K., "On the poincaré–bendixson theorem", IMUJ **19** (2001).

Cipresso, P., "Modeling behavior dynamics using computational psychometrics within virtual worlds", Frontiers in psychology **6**, 1725 (2015).

Couzin, I. D., J. Krause, N. R. Franks and S. A. Levin, "Effective leadership and decision-making in animal groups on the move", Nature **433**, 7025, 513 (2005).

Cucker, F. and S. Smale, "Emergent behavior in flocks", IEEE Transactions on automatic control **52**, 5, 852–862 (2007).

Dalrymple, S. J., J. E. Parisi, P. C. Roche, S. C. Ziesmer, B. W. Scheithauer and P. J. Kelly, "Changes in proliferating cell nuclear antigen expression in glioblastoma multiforme cells along a stereotactic biopsy trajectory", Neurosurgery **35**, 6, 1036–1045 (1994).

Darling, R., J. R. Norris *et al.*, "Differential equation approximations for markov chains", Probability surveys **5**, 37–79 (2008).

Deisboeck, T. S. and I. D. Couzin, "Collective behavior in cancer cell populations", Bioessays **31**, 2, 190–197 (2009).

Deisboeck, T. S. and G. S. Stamatakos, *Multiscale cancer modeling* (CRC press, 2010).

Dussutour, A., V. Fourcassie, D. Helbing and J.-L. Deneubourg, "Optimal traffic organization in ants under crowded conditions", Nature **428**, 6978, 70 (2004).

Fibich, G., "Bass-sir model for diffusion of new products in social networks", Physical Review E **94**, 3, 032305 (2016).

Fibich, G., A. Gavious and E. Solan, "Averaging principle for second-order approximation of heterogeneous models with homogeneous models", Proceedings of the National Academy of Sciences **109**, 48, 19545–19550 (2012).

Filbet, F., P. Laurençot and B. Perthame, "Derivation of hyperbolic models for chemosensitive movement", Journal of Mathematical Biology **50**, 2, 189–207 (2005).

Gazdag, J. and J. Canosa, "Numerical solution of fisher's equation", Journal of Applied Probability **11**, 3, 445–457 (1974).

Göttlich, S., S. Knapp and P. Schillen, "A pedestrian flow model with stochastic velocities: Microscopic and macroscopic approaches", arXiv preprint arXiv:1703.09134 (2017).

Harpold, H. L., E. C. Alvord Jr and K. R. Swanson, "The evolution of mathematical modeling of glioma proliferation and invasion", Journal of Neuropathology & Experimental Neurology **66**, 1, 1–9 (2007).

Hawkins-Daarud, A., S. Prudhomme, K. G. van der Zee and J. T. Oden, "Bayesian calibration, validation, and uncertainty quantification of diffuse interface models of tumor growth", Journal of mathematical biology **67**, 6-7, 1457–1485 (2013).

Hestenes, D., "Modeling theory for math and science education", in "Modeling students' mathematical modeling competencies", pp. 13–41 (Springer, 2010).

Hethcote, H. W., "The mathematics of infectious diseases", SIAM review **42**, 4, 599–653 (2000).

Horn, R. A. and C. R. Johnson, *Matrix analysis* (Cambridge university press, 2012).

Hubbard, J. H. and B. H. West, *Differential equations: a dynamical systems approach: ordinary differential equations*, vol. 5 (Springer, 2013).

Hughes, G. B. and M. Chraibi, "Calculating ellipse overlap areas", Computing and visualization in science **15**, 5, 291–301 (2012).

Judd, C. M., G. H. McClelland and C. S. Ryan, *Data analysis: A model comparison approach* (Routledge, 2011).

Kansal, A., S. Torquato, E. Chiocca and T. Deisboeck, "Emergence of a subpopulation in a computational model of tumor growth", Journal of Theoretical Biology **207**, 3, 431–441 (2000).

Kelly, P. J., "Computed tomography and histologic limits in glial neoplasms: tumor types and selection for volumetric resection", Surgical neurology **39**, 6, 458–465 (1993).

Kelly, P. J., C. Daumas-Duport, D. B. Kispert, B. A. Kall, B. W. Scheithauer and J. J. Illig, "Imaging-based stereotaxic serial biopsies in untreated intracranial glial neoplasms", Journal of neurosurgery **66**, 6, 865–874 (1987).

Kot, M., *Elements of mathematical ecology* (Cambridge University Press, 2001).

Levine, H., S. Pamuk, B. Sleeman and M. Nilsen-Hamilton, "Mathematical modeling of capillary formation and development in tumor angiogenesis", Evolution (2011).

Levine, H. A., S. Pamuk, B. D. Sleeman and M. Nilsen-Hamilton, "Mathematical modeling of capillary formation and development in tumor angiogenesis: penetration into the stroma", Bulletin of mathematical biology **63**, 5, 801–863 (2001).

Li, X., R. Balagam, T.-F. He, P. P. Lee, O. A. Igoshin and H. Levine, "On the mechanism of long-range orientational order of fibroblasts", Proceedings of the National Academy of Sciences **114**, 34, 8974–8979 (2017).

Lowengrub, J. S., H. B. Frieboes, F. Jin, Y.-L. Chuang, X. Li, P. Macklin, S. M. Wise and V. Cristini, "Nonlinear modelling of cancer: bridging the gap between cells and tumours", Nonlinearity **23**, 1, R1 (2009).

Mac Gabhann, F. and A. S. Popel, "Model of competitive binding of vascular endothelial growth factor and placental growth factor to vegf receptors on endothelial cells", American Journal of Physiology-Heart and Circulatory Physiology **286**, 1, H153–H164 (2004).

Menzel, A. M. and T. Ohta, "Soft deformable self-propelled particles", EPL (Europhysics Letters) **99**, 5, 58001 (2012).

Mersch, D. P., A. Crespi and L. Keller, "Tracking individuals shows spatial fidelity is a key regulator of ant social organization", Science **340**, 6136, 1090–1093 (2013).

Morale, D., V. Capasso and K. Oelschläger, "An interacting particle system modelling aggregation behavior: from individuals to populations", Journal of mathematical biology **50**, 1, 49–66 (2005).

Moussaid, M., E. G. Guillot, M. Moreau, J. Fehrenbach, O. Chabiron, S. Lemercier, J. Pettré, C. Appert-Rolland, P. Degond and G. Theraulaz, "Traffic instabilities in self-organized pedestrian crowds", PLoS computational biology **8**, 3, e1002442 (2012).

Norris, J. R., *Markov chains*, no. 2 (Cambridge university press, 1998).

Palsson, E. and H. G. Othmer, "A model for individual and collective cell movement in dictyostelium discoideum", Proceedings of the National Academy of Sciences **97**, 19, 10448–10453 (2000).

Perkins, D. N. and R. Simmons, "Patterns of misunderstanding: An integrative model for science, math, and programming", Review of educational research **58**, 3, 303–326 (1988).

Peruani, F., A. Deutsch and M. Bär, "A mean-field theory for self-propelled particles interacting by velocity alignment mechanisms", The European Physical Journal Special Topics **157**, 1, 111–122 (2008).

Pletjushkina, O. J., O. J. Ivanova, I. N. Kaverina and J. M. Vasiliev, "Taxol-treated fibroblasts acquire an epithelioid shape and a circular pattern of actin bundles", Experimental cell research **212**, 2, 201–208 (1994).

Politopoulos, I., "Review and analysis of agent-based models in biology", University of Liverpool (2007).

Prabhakar, B., K. N. Dektar and D. M. Gordon, "The regulation of ant colony foraging activity without spatial information", PLoS computational biology **8**, 8, e1002670 (2012).

Ribba, B., T. Alarcón, K. Marron, P. K. Maini and Z. Agur, "The use of hybrid cellular automaton models for improving cancer therapy", in "International Conference on Cellular Automata", pp. 444–453 (Springer, 2004).

Robinson, E. J., O. Feinerman and N. R. Franks, "Flexible task allocation and the organization of work in ants", Proceedings of the Royal Society of London B: Biological Sciences p. rspb20091244 (2009).

Saut, O., J.-B. Lagaert, T. Colin and H. M. Fathallah-Shaykh, "A multilayer grow-or-go model for gbm: effects of invasive cells and anti-angiogenesis on growth", Bulletin of mathematical biology **76**, 9, 2306–2333 (2014).

Serini, G., D. Ambrosi, E. Giraudo, A. Gamba, L. Preziosi and F. Bussolino, "Modeling the early stages of vascular network assembly", The EMBO journal **22**, 8, 1771–1779 (2003).

Sherratt, J. A. and J. D. Murray, "Models of epidermal wound healing", Proc. R. Soc. Lond. B **241**, 1300, 29–36 (1990).

Swanson, K. R., "Mathematical modeling of the growth and control of tumors.", (2000).

Swanson, K. R., C. Bridge, J. Murray and E. C. Alvord Jr, "Virtual and real brain tumors: using mathematical modeling to quantify glioma growth and invasion", Journal of the neurological sciences **216**, 1, 1–10 (2003).

Swanson, K. R., R. C. Rockne, J. Claridge, M. A. Chaplain, E. C. Alvord and A. R. Anderson, "Quantifying the role of angiogenesis in malignant progression of gliomas: in silico modeling integrates imaging and histology", Cancer research (2011).

Udiani, O., N. Pinter-Wollman and Y. Kang, "Identifying robustness in the regulation of collective foraging of ant colonies using an interaction-based model with backward bifurcation", Journal of theoretical biology **367**, 61–75 (2015).

Verhulst, F., *Nonlinear differential equations and dynamical systems* (Springer Science & Business Media, 2006).

Vicsek, T., A. Czirók, E. Ben-Jacob, I. Cohen and O. Shochet, "Novel type of phase transition in a system of self-driven particles", Physical review letters **75**, 6, 1226 (1995).

Xiong, Y., P. Rangamani, M.-A. Fardin, A. Lipshtat, B. Dubin-Thaler, O. Rossier, M. P. Sheetz and R. Iyengar, "Mechanisms controlling cell size and shape during isotropic cell spreading", Biophysical journal **98**, 10, 2136–2146 (2010).

Yao, Z., J.-S. Wang, G.-R. Liu and M. Cheng, "Improved neighbor list algorithm in molecular simulations using cell decomposition and data sorting method", Computer physics communications **161**, 1-2, 27–35 (2004).

APPENDIX A

VERLET VS CELL LINKED FOR CHAPTER 2 AND 3

An improved method to calculate the number of neighbors in an algorithm is necessary to avoid some unnecessary distance interaction calculations and to reduce the computational time to perform simulations (Yao *et al.*, 2004). By assuming the number of particles is N, in a squared box of length $L = 1$, we will chose random numbers and compare the time needed for the two methods denoted as Verlet list and Cell linked list. At a fixed time t, the Verlet method is construct such as for a particle i, it will compute all the neighbors which are within the radius R, and will stock them in a neighboring list. On the other hand, the Cell linked method will partition the domain into several artificial cells of length dx and locate the ants in numbered cells at a fixed time. Then, for a particle i in cell m, it will list all the neighbors that are in the cells around cell m which are within the radius $R \leq dx$.

Comparison of Verlet List vs Cell Linked

The idea of Verlet algorithm is to create a list of neighboring for each particle in the system at a fixed time. In the algorithm, for each particle i, it will construct a list of all the indices j, where the distance between i and j is less or equal to the radius R, i.e $d(i, j) \leq R$. We know for N particles, the computational time operations for the Verlet method will be of order N^2, since the distances of all the pairs need to be calculated. While varying the number of particles N between $[10, 1000]$, and fixing the radius $R = \frac{1}{64}$, we observe that the time elapsed vs N for Verlet list and O(N^2) are identical (see figure A.1).

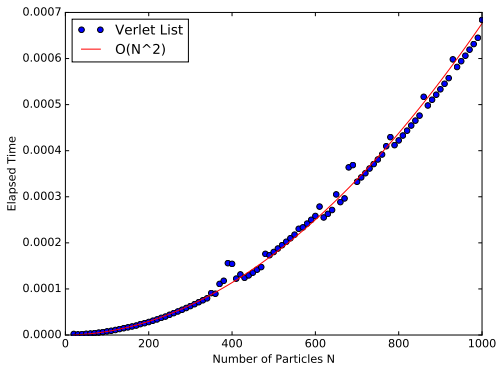


Figure A.1: Computational Time of Verlet and $O(N^2)$ vs the Number of Particles N with Radius $R = \frac{1}{64}$. One can Observe that the Time Elapsed for Both are Identical, which Satisfies the Order of Verlet List

The concept behind the Cell linked list algorithm is that, during a fixed time at the beginning of a simulation, the domain will be partitioned to indicate cells of length dx. After that, it will assign based on the position of each particle i, the number of the cell in which this particle is placed in it. Therefore, the list of cell neighbors will be created for each particle i located in cell m, by finding the distance of i and all the other particles located in the cells around cell m to be less than or equal to the radius R. Note that R should always be $\leq dx$. The number of operations for the Cell linked list is of order N times the average number of particles inside the cells, i.e c, around particle i within radius R, hence the scale is of $O(cN)$.

By comparing the computational time of Verlet List vs Cell Linked List with

different number of particles N, we can see that for small numbers of particles $N < 100$), the Verlet list algorithm is more efficient. On the other hand, for larger N, the time consuming for the Cell linked list will be more efficient (see figure A.2). Hence, for large number of particles, Cell Linked method will save more time to calculate the neighboring list.

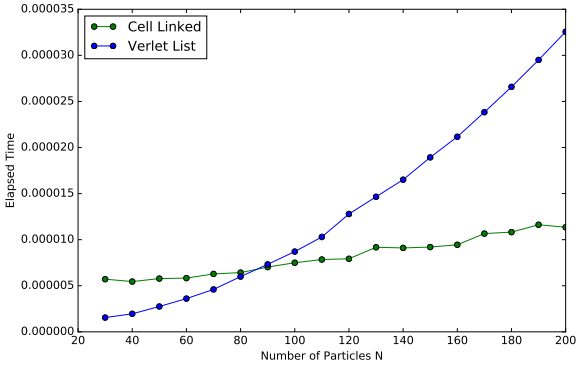


Figure A.2: Computational Time Verlet and Cell Linked vs the Number of Particles N with Radius $R = \frac{1}{64}$. When the $N < 100$ Verlet List is more Efficient, whereas for $N > 100$, Cell Linked is

In addition, we can see that these two methods also depend on the radius R. If we vary $N \in [10, 1000]$ in the box, and take different values for R, such as $\frac{1}{64}, \frac{1}{4}, \frac{1}{2}$, one can deduce that Cell Linked method becomes slower as we increase the value of R (see figures A.3, A.4 -**left**). Ultimately, this is happening because the grid is not small enough when R is increasing, so that every particle is a neighbor of the others, hence, we come back to the same reasoning of figure A.2, where in this case the Verlet method will be more efficient.

Moreover, if we fix the number of particles to $N = 1,000$, and plot the computational time of Verlet list and Cell linked when the radius $R \in [0, 0.6]$, we can see that, as R increases, Cell-linked method becomes less efficient than Verlet Method (see figure A.3 - **right**.). Also, we can see in figure A.4 - **right**, that at $R = \frac{1}{4}$, Verlet and Cell linked method overlapped which satisfies that intersection point shown in figure A.3 - **right**.

In conclusion, for large number of particles N, and for small radius R, we can see that the computational time of the Cell linked method is more efficient than the Verlet Method. In our studies in the previous chapters, since the number of ants and cancer cells is large but the radius of interaction is small, then the neighbor list of the Cell linked method will be much better and faster to perform all the calculations needed.

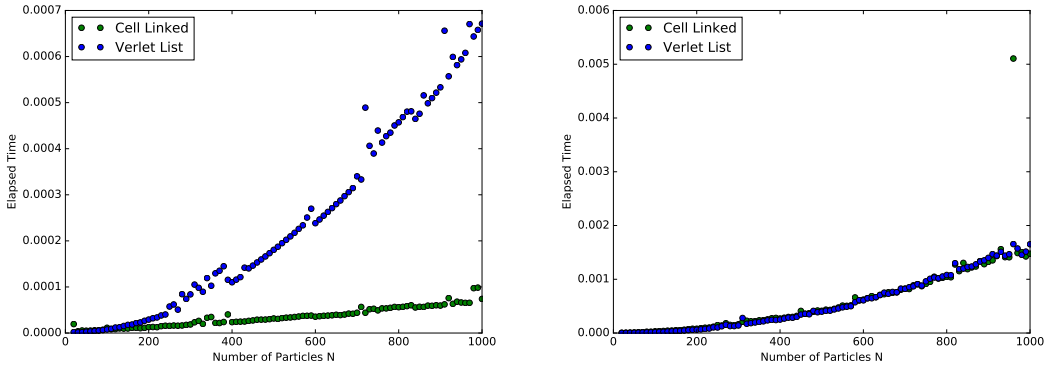


Figure A.3: Computational Time of Verlet and Cell Linked vs the Number of Particles N with Radius $R = \frac{1}{64}$ (Left) and with $R = \frac{1}{4}$ (Right). When the Radius is Close to Zero, one can Observe that Cell Linked Method is more Efficient, whereas by Increasing $R = 0.25$, the Two Method will have the Same Elapsed Time

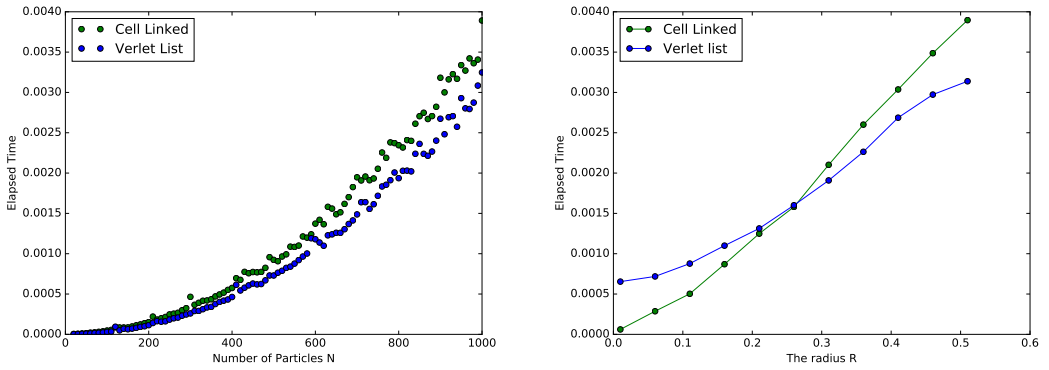


Figure A.4: Computational Time of Verlet and Cell Linked vs Number of Particles N with Radius $R = \frac{1}{2}$ (Left). Computing the Time of the Two Methods while Varying the Radius R Between $[0, 0.6]$ with Number a Fixed Number of Particles $N = 1000$ (Right). When we increase R to 0.5, Verlet List Method will be more Efficient in this Case. By Fixing the Number of Particles and Comparing the Two Method vs R, it is Observed that at $R \approx 0.25$ where the Elapsed Time Cross Between the Two