Characterizing heterogeneity in microscopic collectives

A Thesis

submitted to

Indian Institute of Science Education and Research Pune in partial fulfilment of the requirements for the BS-MS Dual Degree Programme

by

Balagopal R Nair

Indian Institute of Science Education and Research Pune

Dr. Homi Bhabha Road,

Pashan, Pune 411008, INDIA.

April, 2025

Supervisor: Dr. Danny Raj M

Certificate

This is to certify that this dissertation entitled "Characterizing heterogeneity in microscopic collectives" towards the partial fulfilment of the BS-MS dual degree programme at the Indian Institute of Science Education and Research, Pune represents study/work carried out by Balagopal R Nair at the Indian Institute of Technology, Madras under the supervision of Dr. Danny Raj M, Assistant Professor, Department of Applied Mechanics and Biomedical Engineering, during the academic year 2024-2025.

Balagopal R Nair

Committee:

Danny Raj M

Declaration

I hereby declare that the matter embodied in the report entitled "Characterizing heterogeneity in microscopic collectives" are the results of the work carried out by me at the Department of Applied Mechanics and Biomedical Engineering, Indian Institute of Technology, Madras, under the supervision of Danny Raj M and the same has not been submitted elsewhere for any other degree. Wherever others contribute, every effort is made to indicate this clearly, with due reference to the literature and acknowledgement of collaborative research and discussions.

Balagopal R Nair

20201135

Danny Raj M

Table of Contents

Table of figures

Abstract

Most collective systems in real life from animal herds, bird flocks and schools of fish to bacterial swarms, and cell migration are heterogenous in nature; they contain multiple types of individuals. These differences among the individuals can affect the emergent behaviour of the collective as a whole. There have been many techniques developed to infer the heterogeneity in these systems, but none of them deal with complex interactions between agents like that present in microscopic collectives. In these systems, long range hydrodynamic interactions have a considerable effect on the observed dynamics. By modelling a system that simulates a microscopic collective in a fluid medium and introducing heterogeneity into it, we can study the observed dynamics and their relation to the intrinsic properties.

Acknowledgement

I would like to thank my supervisor, Dr. Danny Raj M for his continuous guidance and support throughout the duration of the thesis. I learnt a lot from him even outside of the work I have done. I would also thank my parents and my sister for their unwavering support.

1. Introduction

Active matter refers to a system consisting of a large number of agents that are able to self-propel and interact with each other. Through these interactions, the agents are able to display new emergent behaviour on a system scale. Collective motion refers to the ordered movement of the system that emerges due to the interactions between agents. Most examples of collective systems are biological in nature across all scales of life, such as flocks of birds, schools of fish, herds of animals and swarms of bacteria. These systems exhibit self-organization, where global order emerges without centralized control, making them a subject of interest in physics, biology, and robotics. One of the main goals of studying collective motion is to understand the relation between the dynamics that we observe, the intrinsic properties of the system and the interactions between agents in the system.

Agent-based modelling is a technique used to mathematically model active matter systems. It involves modelling the rules of how agents interact with each other and obtaining the complex behaviour that emerges from the inter-agent interactions. The interactions between agents can be physical in nature, such as steric repulsion or hydrodynamic forces, or behavioural, such as alignment and attraction rules inspired by biological systems. By simulating these interactions, we can investigate how microscopic rules lead to macroscopic phenomena, including phase transitions and pattern formation.

A well-known example of an agent-based model of an active matter system is the Vicsek model (1). It consists of self-propelling agents that move with a constant speed in a 2D environment. The agents follow one simple rule; at a given time, an agent's direction of motion is given by the average of agents within its neighbourhood radius with some random noise added. Despite its simplicity, the model exhibits rich collective dynamics, including a transition from a disordered state, where agents move randomly, to an ordered state, where they move coherently in the same direction. This transition is analogous to phase transitions in condensed matter systems and has been widely studied in the context of non-equilibrium statistical physics.

Another example of an agent-based model is the 3D Couzin model (2). It is inspired by the Vicsek model and adds more rules to the agents in the system. Each agent has multiple concentric zones of different radii around it, the zone of repulsion, alignment, and attraction in that order. These zones define how an agent responds to its neighbours based on their relative positions. The zone of repulsion is the closest region to an agent. It ensures that the focal agent moves away when another agent enters this zone to avoid collisions. When an agent has no neighbours in this zone, it considers neighbours in the zone of alignment, where it adjusts its direction to match their average heading. If there are no neighbours in the alignment zone, the agent is instead influenced by the zone of attraction, moving toward nearby agents to maintain cohesion. Along with this, each agent also has a cone-shaped zone of blindness where the presence of neighbours does not affect the agent's movement.

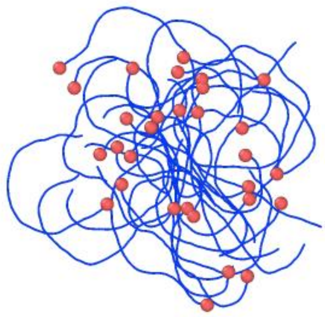

Figure 1. A snapshot from a simulation of a 2D Couzin model with agent trajectories. (16)

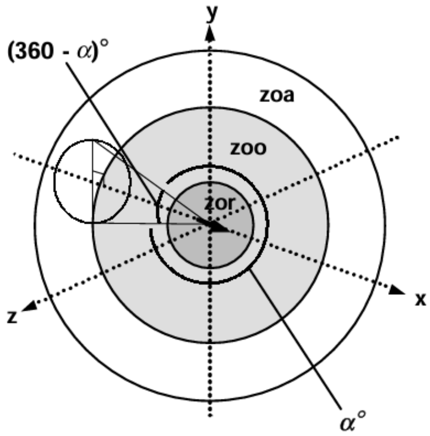

Figure 2. The zones of interaction around an agent in the Couzin model.

Image from Couzin et al. (2)

Varying the sizes of these zones leads to the emergence of different types of collective behaviour in the system. These behaviours were labelled as follows:

a) Swarm – occurs when the zone of alignment is minimal or absent while the zone of attraction is prominent and leads to an aggregate of the agents that don't show overall movement.

b) Torus – occurs when the zone of alignment is small compared to the zone of attraction. The agents rotate around an empty core in a random direction.

c) Dynamic parallel group – is seen when the size of the zone of orientation is intermediate, and the zone of attraction is intermediate to large. The agents move similar to the swarm or torus but as a group is more motile.

d) Highly parallel group – as the zone of orientation grows larger, the agents all align with each other and show linear movement.

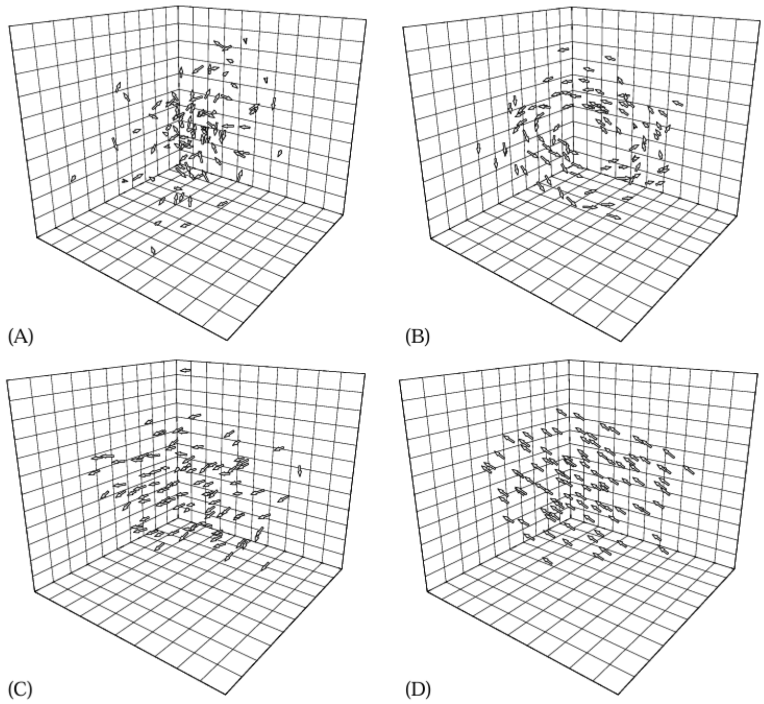

Figure 3. The different behaviours that we see in the Couzin model with different interaction radii. A) Swarm, B) Torus, C) Dynamic parallel group, D) Highly parallel group.

Image from Couzin et al. (2)

From a simple set of rules of interactions between agents, we are able to see a large variety in the emergent behaviour.

1.1 Heterogeneity

Most models of collective systems often assume the individuals in the system are identical, i.e., homogenous. In reality, this is not usually the case. Most real systems have agents with different properties, i.e., they are heterogeneous in nature. This heterogeneity can arise as discrete types within the collective or as a distribution of a certain property of the agents. In real systems, we see these heterogeneities in different forms. In bacterial systems, there may be variances in their size or speed. Animal herds may have multiple species within them, or same-species herds can have animals of different ages or sexes, which can lead to behavioural differences among them (3). In cell migration, we find that heterogeneity can arise dynamically in the form of leader-follower states that dictate the migration dynamics of these cells, especially during processes such as wound closure (4). Heterogeneity within the collective can affect the overall dynamics of the system. Hence, identifying it is essential to understanding how it impacts the system.

Classifying heterogeneity in collectives using only observed data has been done using many different techniques. In a paper by Nabeel et al. (5), the information of the system's heterogeneity was extracted from the velocities of the agents. The system that they studied was a 2D bidisperse collective. There were two types of agents, one whose intrinsic velocity (v_0) was in the +x direction (group 1) and one whose intrinsic velocity was in the -x direction (group 2). The agents were in a 2D periodic space. Different emergent behaviours could be observed depending on the agents' density and speeds.

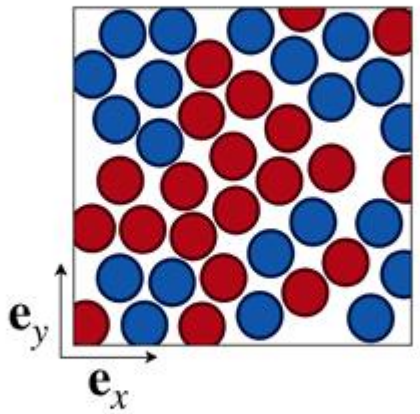

Figure 4. The bidisperse collective.

Image from Nabeel et al. (5)

Since the agents would all collide with each other and want to move in their desired directions, an agent's intrinsic velocity cannot be directly inferred from observing it as it could be pushed against its desired direction. The first approach to infer an agent's desired direction, the simple observer, involved an observer using information of the agents' instantaneous velocity (vi) averaged over a time window to obtain an estimate of the intrinsic velocity. The simple observer was not very accurate at classifying agents correctly

.

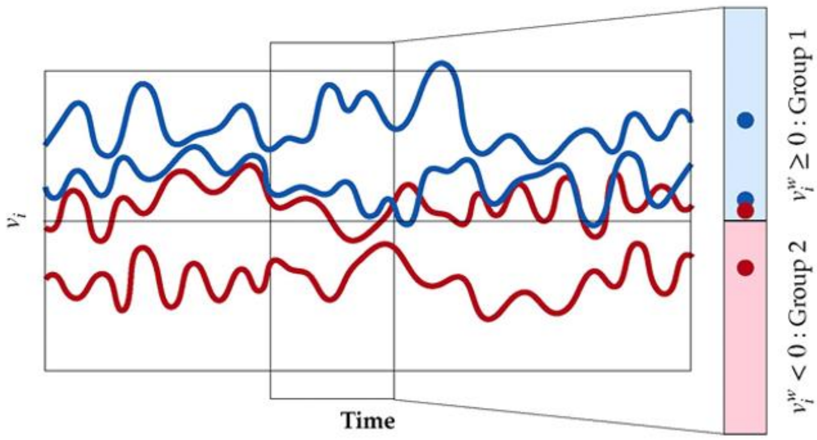

Figure 5. Visual representation of the simple observer. The time series of x-velocity is plotted for multiple agents belonging to both groups. The average of their x-velocity within a time window is taken and is used to determine their classification. We see that the observer wrongly classifies an agent from group 2 as group 1.

Image from Nabeel et al. (5)

The second approach involves decoupling the agents' intrinsic motion from the component of motion that is caused by interactions with other agents. Thus, the observed velocity v_i is separated into its intrinsic velocity v_0 and its interactions. Since the main source of interactions is short-range collisions with an agent's neighbours, we can construct a neighbourhood parameter ϕ, using information of the velocities of the first Voronoi neighbours. This approach was called the neighbourhood observer. The component of the velocity of a neighbour along the line joining the focal agent and the neighbour was taken and averaged over the other neighbours to obtain this unknown neighbourhood parameter. This parameter, obtained over a window of time, was subtracted from the focal agent's instantaneous velocity to obtain an estimate of its intrinsic velocity for that time window.

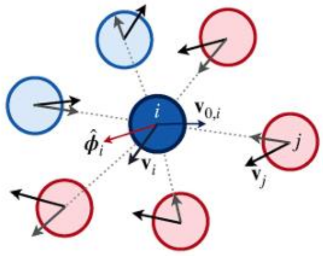

Figure 6. Obtaining the neighbourhood parameter.

Image from Nabeel et al. (5)

The neighbourhood observer was able to classify the agents as one group or the other more successfully than the simple observer. It was not able to quantify the exact velocities of the agents accurately.

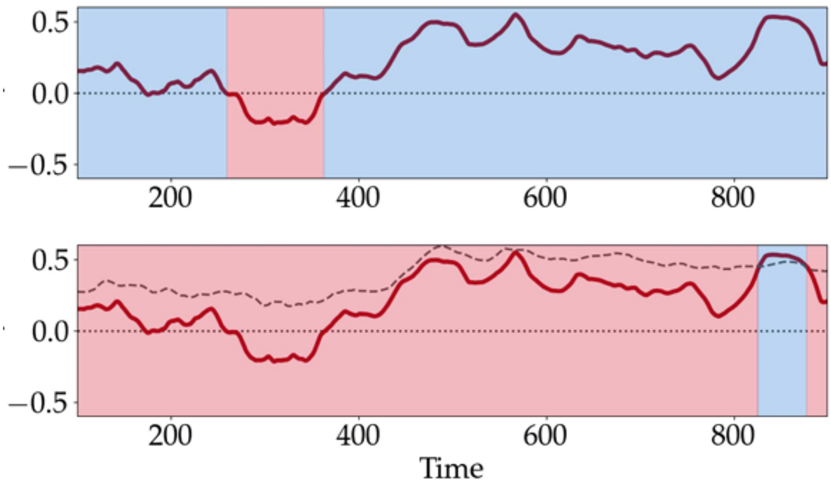

Figure 7. Comparing the accuracies of the simple and neighbourhood observers. The top panel is the simple observer and the bottom is the neighbourhood observer. The same agent (belonging to group 2) is observed by both observers, the red solid line being the observed x-velocity of the agent. The dotted line in the second panel is the neighbourhood parameter for that agent. The shading in the background is the group in which the observer classifies the agent, red being group 2 and blue being group 1. The neighbourhood observer is clearly more accurate than the simple one.

Image from Nabeel et al. (5)

In a paper by Tan et al., (6) principal component analysis (PCA) was used to classify more than two types of agents in a collective. PCA is an example of a linear dimensionality reduction technique. They used the Vicsek model as their collective system and introduced heterogeneity by having two or more types of agents with either differing intrinsic velocities or noise. The heading angles of the agents were extracted from the position data and used to construct the N × t data matrix on which PCA is performed to cluster the data into respective types. It was able to, with a sufficient amount of data, obtain distinct clusters for each agent type. This method requires a large amount of data to be accurate, which is not always available in realistic cases. PCA was also tested on another agent-based active matter model called the D' Orsogna model, yielding similar results.

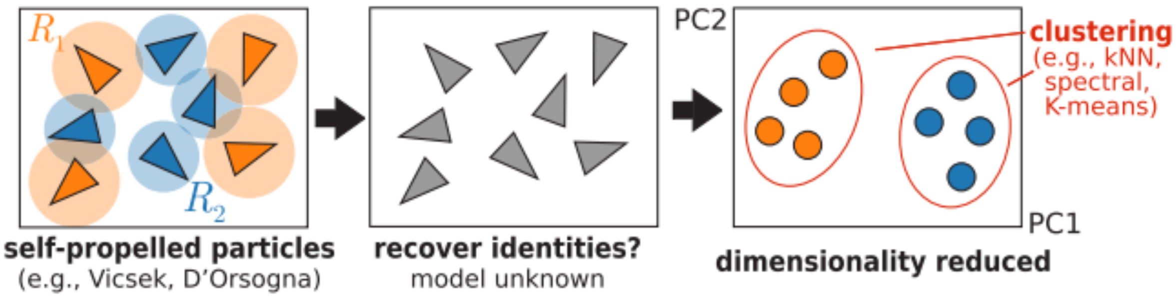

Figure 8. Schematic of the work done. Using only trajectory data of a heterogenous collective to classify the particles in the system

Image from Tan et al. (6)

In the above-discussed papers, the systems in consideration had fairly simple inter-agent interactions. The neighbourhood parameter could be used to classify agents more easily since the system is very dense and short-range collisions are a dominant interaction. PCA was only good at classifying agents when long-time datasets are available, which is usually not the case when tracking microscopic systems like cells, bacterial swimmers, etc.

In the work done by Schumacher et al. (7) ,it is discussed how, in the study of cell migration, heterogeneity in the system that is observed may not be an intrinsic property of the system and could be an emergent phenomenon or a statistical artefact. When developing a mechanistic description of collective cell migration, it becomes unclear whether or not to include heterogeneity. They show, using a minimal model of collective cell migration and using distribution of delay times in directional cross-correlation to measure heterogeneity, that non-zero heterogeneity can be measured in a homogenous population with interactions. The bias towards heterogeneity also increases with stronger intercellular forces.

Complex interactions between agents can affect how heterogeneity, if it exists, is perceived. Microscopic collectives in a fluid medium are an example of a system where these complex interactions are observed. We see microscopic collectives in nature in the form of bacterial swarms and cell migration. Recent work has also been done in developing artificial microswimmers, which are inspired by biological microswimmers and used in different biomedical applications, such as cancer treatment (8). Studying such systems is important because at the microscopic scale, hydrodynamic interactions differ from the macroscale. Hydrodynamic interactions along with heterogeneity in the system, can give rise to complex dynamics. Studying such systems will provide us with insight into how the heterogeneity in the presence of complex interactions can be extracted from real systems.

We model a microscopic collective with long-range hydrodynamic interactions and heterogeneity in the speed of the agents. How does the strength of these interactions affect the perceived heterogeneity of the system?

2. Methods

2.1 The model

I have constructed an agent-based model consisting of spherical agents of radius R immersed in a fluid medium in 2D periodic space. There are 3 main forces driving motion. The self-propulsion, hydrodynamic interactions and the excluded volume interaction which is modelled as a spring force. Along with this there is also a random noise involved in the rotation of each agent.

$$m\frac{d\vec{v}_i}{dt} = K_d\left(\vec{v}_{0_i} - \vec{v}_i\right) + \vec{F}_h + \vec{F}_r \tag{1}$$

F_h is the hydrodynamic force, and F_r is the excluded volume spring force interaction. We have chosen to model the interaction as a spring force because we assume that the agents are soft and allow overlapping to an extent, unlike hard spheres. v_0 is the intrinsic velocity of the agent in the absence of any interaction force ($v_0 = u_0 \cdot e_i^{\parallel}$, where u_0 is the speed of the particle and $e_i^{\parallel}$ is the direction of the agent), and v_i is the instantaneous velocity of the i^{th} particle. K_d is the viscous drag acting on the agent ($K_d = 6\pi\eta R$, η is viscosity).

Microscopic systems exist in the low Reynolds number regime, where viscous dominate over inertial forces. So, the inertial term goes to zero. This means that agents instantly change their velocity at each time step.

$$\vec{v}_i = \frac{d\vec{r}_i}{dt} = \vec{v}_{0_i} + \vec{U}_i + \vec{F'}_r \tag{2}$$

The instantaneous velocity of the i^{th} agent v_i is given by the sum of its intrinsic velocity v_{0i} with the interaction terms. F_h divided by K_d gives a velocity term U_i, which is the velocity field acting on agent i.

$$F_r' = \begin{cases} k(R_{min} - R_i), & if\ \ R_i \le R_{min} \\ 0, & otherwise \end{cases} \tag{3}$$

When dividing F_r by K_d, it is absorbed into the spring coefficient k. R_{min} is the minimum distance between the centres of two agents, after which the spring force activates. It is taken as twice the radius of an agent. Thus, the spring force activates when two agents overlap.

The angular velocity experienced by an agent consists of the hydrodynamic interaction term and a random noise term.

$$\frac{d\theta}{dt} = \Omega_i + \eta \tag{4}$$

In the absence of hydrodynamic interactions and the noise in the agents turning, the agents move at a constant speed in a straight line, and are only affected by collisions with other agents. The random noise at each time step for each agent is sampled from a uniform normal distribution.

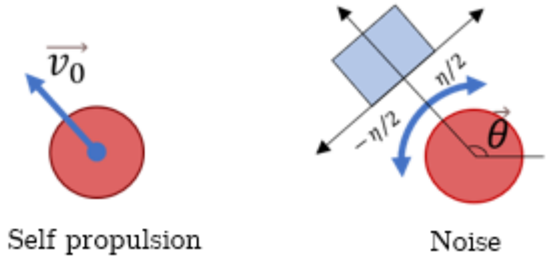

Figure 9. Intrinsic properties of the agent

2.1.1 The hydrodynamic field

We require a model that accurately represents the hydrodynamic flow field generated by a microscopic agent swimming in a fluid. At this scale, inertia is negligible, and viscous forces dominate the dynamics. To simplify the model, we use a far-field approximation and neglect vortices shed in the swimmer's wake.

Under these approximations, the hydrodynamic flow is represented as a dipole field. This approach is based on the work of Filela et al. (9), who used a similar hydrodynamic field to model fish dynamics. While their system differs from ours, they assume that agents have no inertia which is the same as our case. Hence, we have a simple representation hydrodynamic flow in microscopic scale that we can use to model the collective dynamics of the system. The equation of the velocity field acting on an agent is given as:

$$U_i = \sum_{j \neq i} u_{ji} \tag{5}$$

$$u_{ji} = I_f \, v_{0_j} \frac{e_j^{\theta} \sin\theta_{ji} + e_j^{\rho} \cos\theta_{ji}}{\rho_{ij}^2} \tag{6}$$

I_f is the tuneable strength of the dipole field and v_{0_j} is the intrinsic speed of the j^th^ agent. The strength of the dipole field produced by an agent is assumed to be linearly proportional to the intrinsic speed of that agent. θ_{ji} is the angle between the current direction of the j^{th} agent and the line joining i^{th} and j^{th} agent.

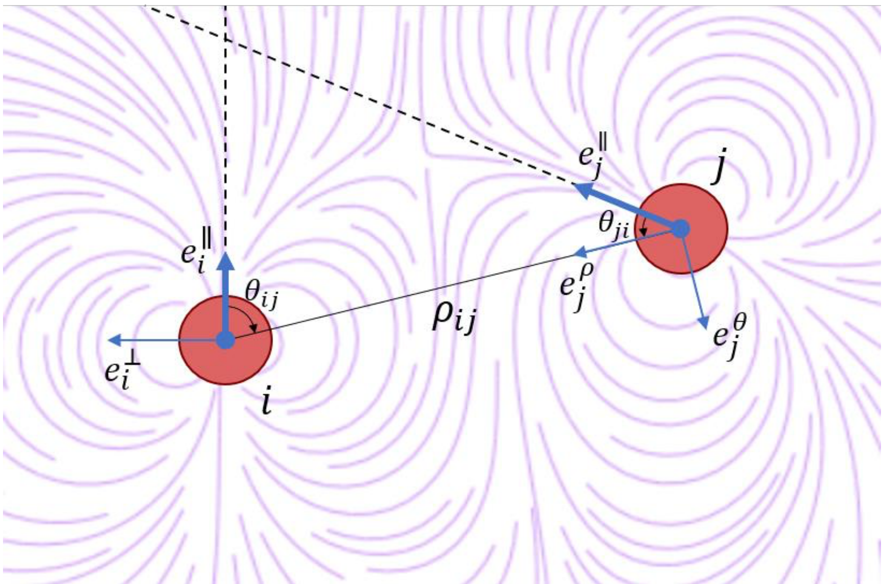

Figure 10. Hydrodynamic interactions between agents

We assume the potential flow approximation to neglect vorticity to simplify the system. The shear effect of the hydrodynamic flow affects the angular velocity of an agent. It is modelled as the term Ω_i.

$$\Omega_i = \sum_{j \neq i} e_i^{\parallel} \cdot \nabla u_{ji} \cdot e_i^{\perp} \tag{7}$$

2.1.2 Introducing heterogeneity

Heterogeneity is introduced in the model via the intrinsic speeds (v_0) of the agents. The intrinsic speeds are sampled from a normal distribution with a mean of 1 and standard deviation σ. Increasing σ increases the heterogeneity within the system. The normal distribution from which the intrinsic speeds are sampled is truncated at 0 to prevent negative values.

2.1.3 Parameters used in the model

The model is coded entirely in Python using NumPy for calculations (10). Graphs were plotted using Matplotlib. Animations and snapshots were done using OVITO (11)

- The agents are in a 2D periodic square arena with a length of 3.
- The agents themselves have a radius of 0.1. The intrinsic speed of the agent in the homogenous case is set to 1, which is 10 times the length scale of the agent. This is typically the case for bacteria like E.coli, which has typical lengths of 1-2μm and speeds of around 30μm/s (12) (13).
- The angular noise term is obtained from a uniform distribution of size η, which is set as 0.1.
- The size of the timestep (dt) is set to 0.01.
- The tuneable parameters in our system are the hydrodynamic strength (I_f) and the standard deviation of the intrinsic speeds (σ).
- For each simulation, the agents are given random initial positions and orientations. The simulation is initialized for some time steps to allow the system to stabilize and prevent the artefacts caused by randomly positioning agents.
- The values of these parameters and the range of values for tuneable ones were chosen for convenience and after some trial and error.

2.1.4 Features of the model

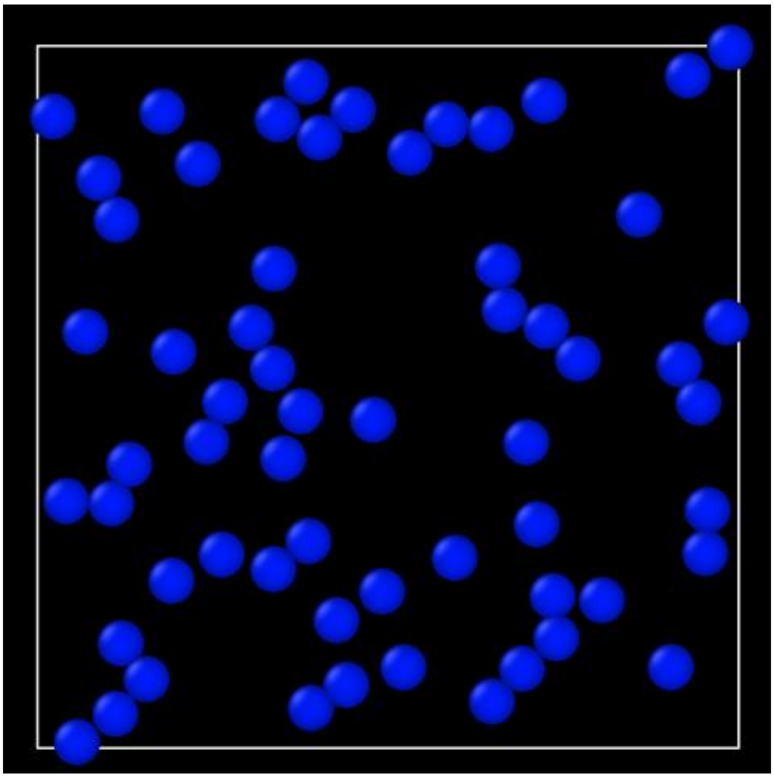

Figure 11. A snapshot of the simulation

The hydrodynamic interactions bring a lot of interesting dynamics to the system. Agents are pushed and pulled by the fields produced by other agents. These pushes and pulls cause a change in the velocities of the agents, bringing large variability to the observed speed.

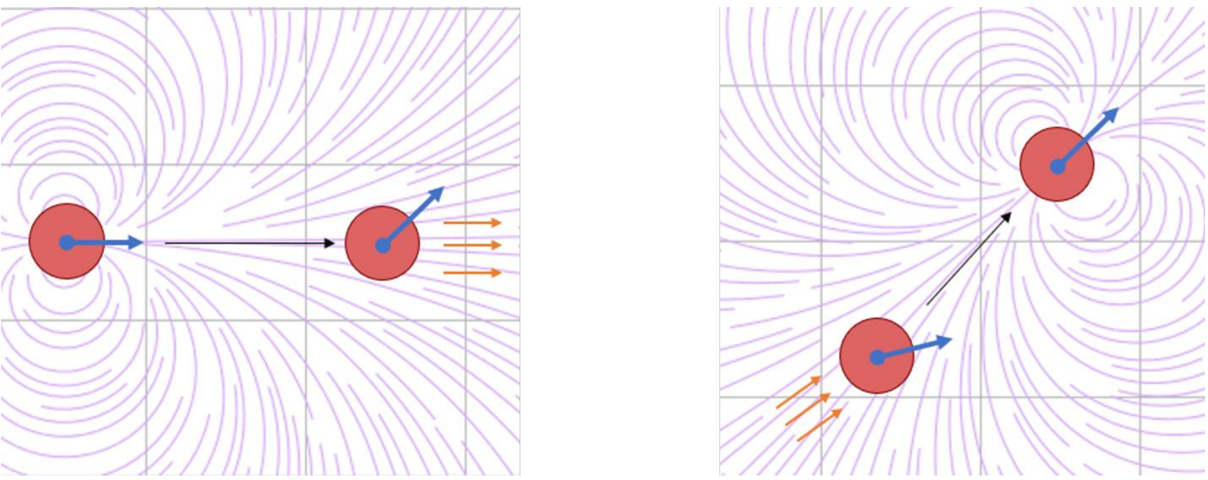

Figure 12. The pushing and pulling effect of the hydrodynamic field

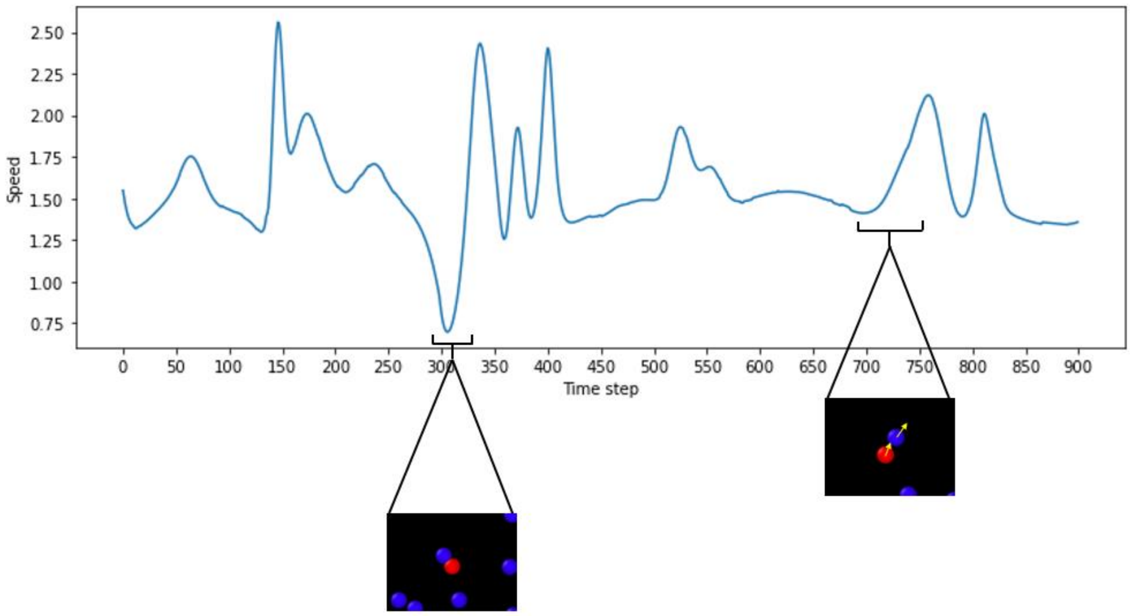

Figure 13. The variability in observed speed of an agent due to the presence of hydrodynamic interactions. The snapshots show the collision and speeding up events happening

Due to the shape of the field, agents will affect other agents that are not on their path and cause a change in their trajectories.

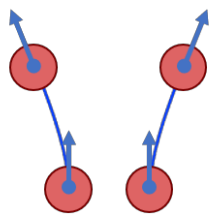

Figure 14. Trajectory of two parallel moving agents altered due to hydrodynamic interactions

3. Results and Discussion

3.1 Preliminary results

3.1.1 Effect of hydrodynamic interactions on a homogenous system

To study the effect of hydrodynamic interactions, we consider a homogenous system. The system itself is not very dense, and agents can move freely, with collisions happening semi-frequently. Long-range hydrodynamic interactions will be dominant and will affect the speeds of the agents. An observer will only have the speeds and positions of the agents available to them. Using only this information, what can an observer infer about the system?

To study this, we took 60 agents in a 3x3 length box. The agents' intrinsic speeds are all set to 1, a homogenous system. The hydrodynamic strength I_f is set to 0.02. The simulation is run for 1000 time steps, and the speeds of every agent at every time step are stored to create a histogram. We get the resulting histogram in Figure 15.

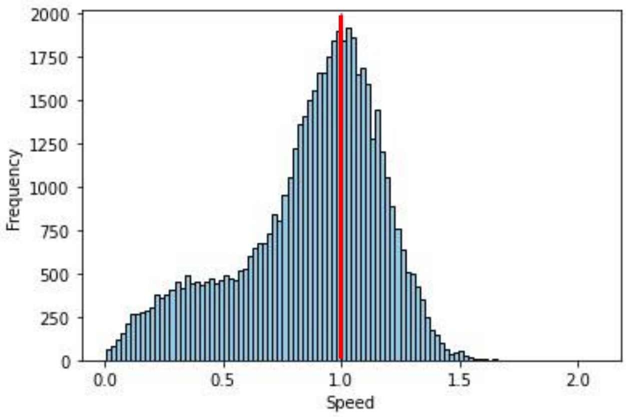

Figure 15. The distribution of observed speeds in a system with hydrodynamic interactions and collisions. The red line represents the intrinsic speed of all the agents

We see that the observed speed forms a distribution around the intrinsic speed. The distribution is not Gaussian and has a bump at lower speeds. From this, we can infer that an observer looking at speeds will assume that the system is heterogeneous in speed since there is a significant variance in agent speed from its intrinsic value. The initial bump can be attributed to collisions slowing down the agents.

To properly understand the effects of hydrodynamic interactions we must also consider the case where they are not present. Collisions will be the only interactions present in the system. How will what an observer sees be affected by the absence of hydrodynamic interactions?

We use the same system: 60 agents in a 3x3 length box, agents' intrinsic speeds are set to 1, hydrodynamic strength I_f set to 0.02, and the simulation is run for 1000 time steps.

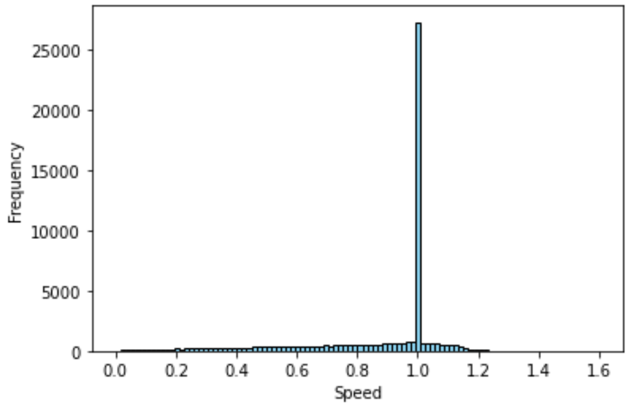

Figure 16. The distribution of observed speeds when there are no hydrodynamic interactions (only collisions)

We see that the observed speed is considerably different compared to when hydrodynamic interactions are present. We see a prominent spike in frequency at the intrinsic speed. An observer would correctly assume that the system is homogenous due to the lack of any variance in the observed speed. There are small frequencies of non-intrinsic speeds that can be attributed to the collisions occurring among agents. This small change in the speed of an agent quickly recovers to its intrinsic speed.

We saw the effects of hydrodynamic interactions on the system scale, but how are individual agents affected? Will an observer be able to infer homogeneity more easily by looking at individual agents instead of the whole system? We can answer this by plotting the time series of the speed of a singular agent and see how it evolves with time compared to its intrinsic speed.

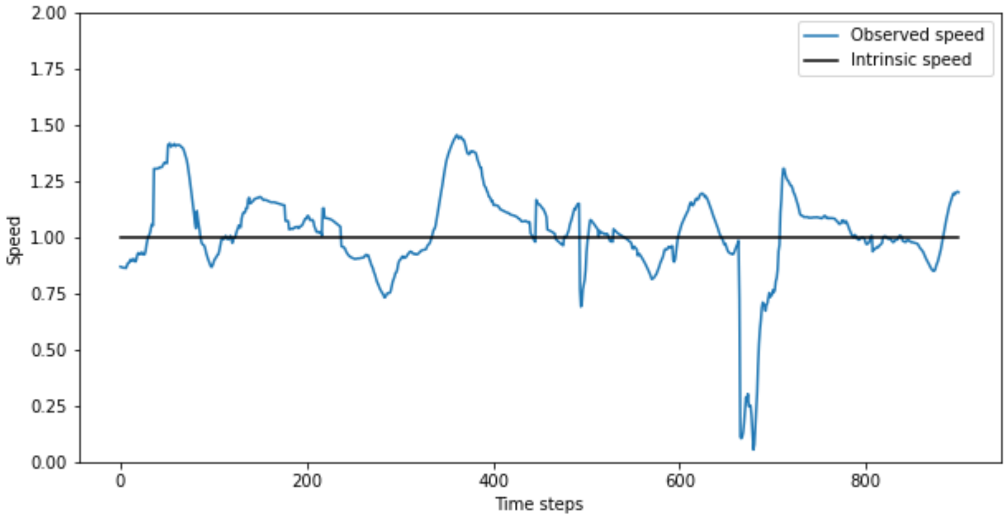

Figure 17. Time series of speed of a single agent, with hydrodynamic interactions

The speed of an agent varies significantly from its intrinsic speed throughout the entire simulation. There are prominent increases and decreases in speed compared to its intrinsic speed throughout. The decreases can be attributed to both long-range hydrodynamic forces slowing down agents and short-range collisions between agents. The increases in speed are due to the hydrodynamic interactions pushing or pulling agents to speeds higher than their intrinsic speed.

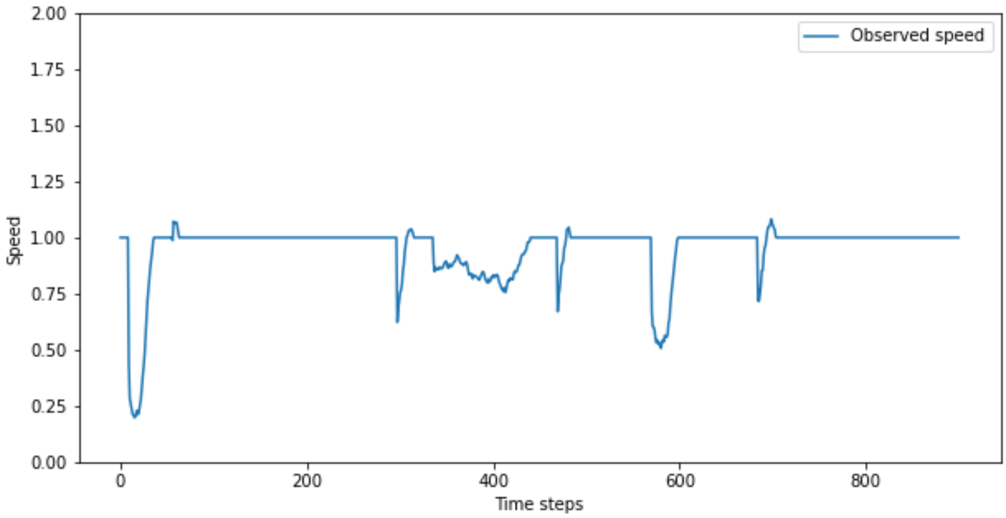

Figure 18. Time series of speed of a single agent, without hydrodynamic interactions

From the time series of a single agent's speed without any hydrodynamic interactions, we see that an agent is able to maintain its intrinsic speed for longer times. The decreases in speed are due to collisions with other agents. The small peaks can be attributed to the collision spring force causing an extra boost of speed.

An observer would still be unable to tell if a system is homogenous or heterogenous from this data. We need to explore methods to extract the intrinsic speed of the system to correctly characterize its homogeneity.

3.1.2 Using the neighbour parameter to obtain intrinsic speed of a homogenous system

In the absence of any external interactions, an agent would always move at its intrinsic speed. An agent's movement is influenced by interactions with other agents in the system. By accounting for the influence of these neighbours on an agent's speed, we should be able to get information about the intrinsic speed of the focal agent. For our system, an agent's observed speed can be decoupled into its intrinsic speed plus the interactions.

If we can quantify the effect of the interactions in a parameter, we can calculate an estimate for intrinsic velocity by subtracting the parameter from the observed velocities at each time step. From this, we can calculate the estimated intrinsic speed of each agent for every time step. Plotting the distribution of the estimated intrinsic speed should give us a significant peak at the actual intrinsic speed, similar to the distribution obtained in Figure 8b, when there are no hydrodynamic interactions.

We construct this parameter by considering the effect of the first Voronoi neighbours of the focal agent. We take the component of the velocities of the neighbours acting along the line joining the neighbours and focal agent and take its average to obtain the neighbourhood parameter. This neighbourhood parameter was constructed the same way as Nabeel et al. in their work with the bidisperse collective (5).

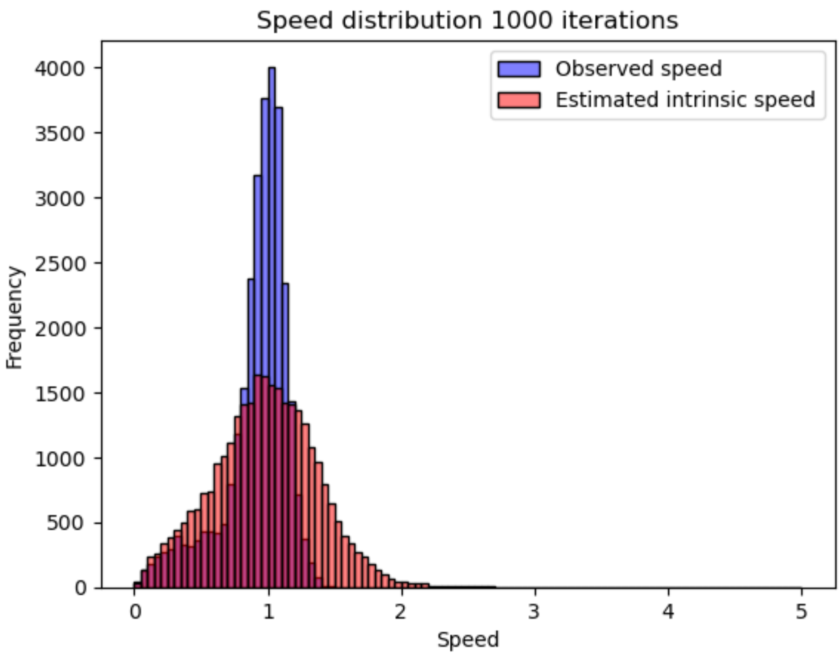

Figure 19. Distributions of the estimated intrinsic speed and the observed speed. The intrinsic speed was estimated using the neighbour parameter

The neighbourhood parameter is clearly unable to extract the intrinsic speed from the observed speeds. It was originally used in a much denser system where short-range collisions were dominant. In our system, the agents are not very closely packed, so the first Voronoi neighbourhood of an agent is not an accurate representation of the actual agents interacting with it. It is unable to capture the effect of long-range interactions that are present in our system. Moreover, the orientation of the neighbours can also affect the direction of the hydrodynamic force acting on an agent, which is not captured by the neighbourhood parameter.

We were able to construct a parameter that borrowed the hydrodynamic interaction directly from the model and obtain the expected distribution.

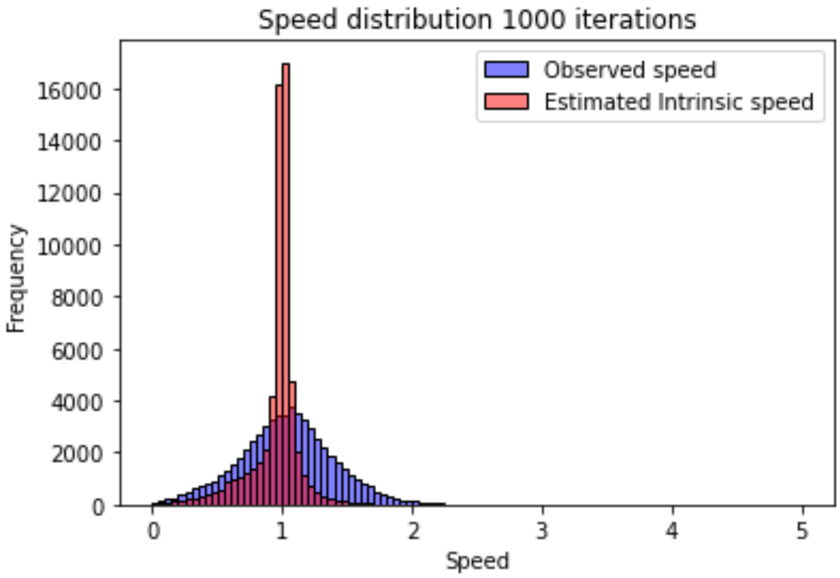

Figure 20. Distributions of the estimated intrinsic speed and the observed speed. The intrinsic speed was estimated using a parameter that borrowed information directly from the model

This shows that this method of obtaining the intrinsic speed is viable. However, the main feature of this parameter is that it must be agnostic to the model we use it on. Constructing a parameter that is able to accurately quantify the complex long-range interactions in our system and account for the orientation of agents without directly borrowing from the model itself proved difficult.

3.2 Heterogenous systems

We saw how the presence of hydrodynamic interactions in the system caused the observed speeds of a homogenous system to be perceived as heterogeneous. Trying to extract the intrinsic speed from this observed speed proved to be a difficult task.

So, we now ask how heterogeneity within the intrinsic speed itself affects the observed speeds. We introduce this heterogeneity in speed by sampling each agent's intrinsic speed from a normal distribution with mean 1 and varying standard deviation. The distribution from which we sample the speeds is truncated at zero since speed cannot be negative. Since the speeds are truncated at zero, for large values of standard deviation, sampled intrinsic speeds are more likely to be lower than the mean than higher. This can affect how the dynamics play out.

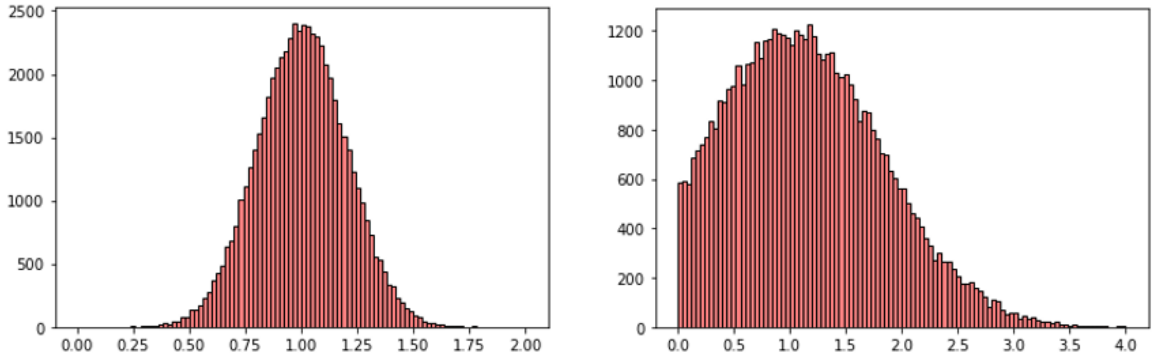

Figure 21. The truncated normal distributions for standard deviation 0.2 (left) and 0.8 (right). The bias for lower values sampled at higher standard deviation is visualised

When observing these systems, the scale at which data is taken from can affect the conclusions that can be drawn. For our analysis, we consider the system at the ensemble scale which consists of multiple short time realizations combined, the realization scale where a single instance of the system's evolution is analyzed, and the agent scale, where we focus the agents' dynamics.

3.2.1 Ensemble scale

In real experiments involving tracking microscopic swimmers, getting long time-tracking data is usually not feasible. Usually, the data is obtained for short amounts of time. Multiple of these short-time realizations are taken together to form an ensemble. When the analysis being done does not depend on which realization a data point is from, it is all pooled together to get large, averaged-out data to work with.

We take a similar approach and run the simulation of the system of 60 agents for 1000 time steps, then repeat it for an ensemble of 100 realizations. The data obtained is independent of which realization or time it was obtained from and is all pooled together. This is done for varying levels of heterogeneity and hydrodynamic strength (I_f). The heterogeneity is increased by increasing the standard deviation (σ) of the normal distribution from which intrinsic speed is sampled.

We plot distributions of the intrinsic speed (v_0), which is hidden in the model and the observed speed, which is data available to us. The plots are obtained for σ varying from 0 to 0.8 and I_f varying from 0 to 0.04

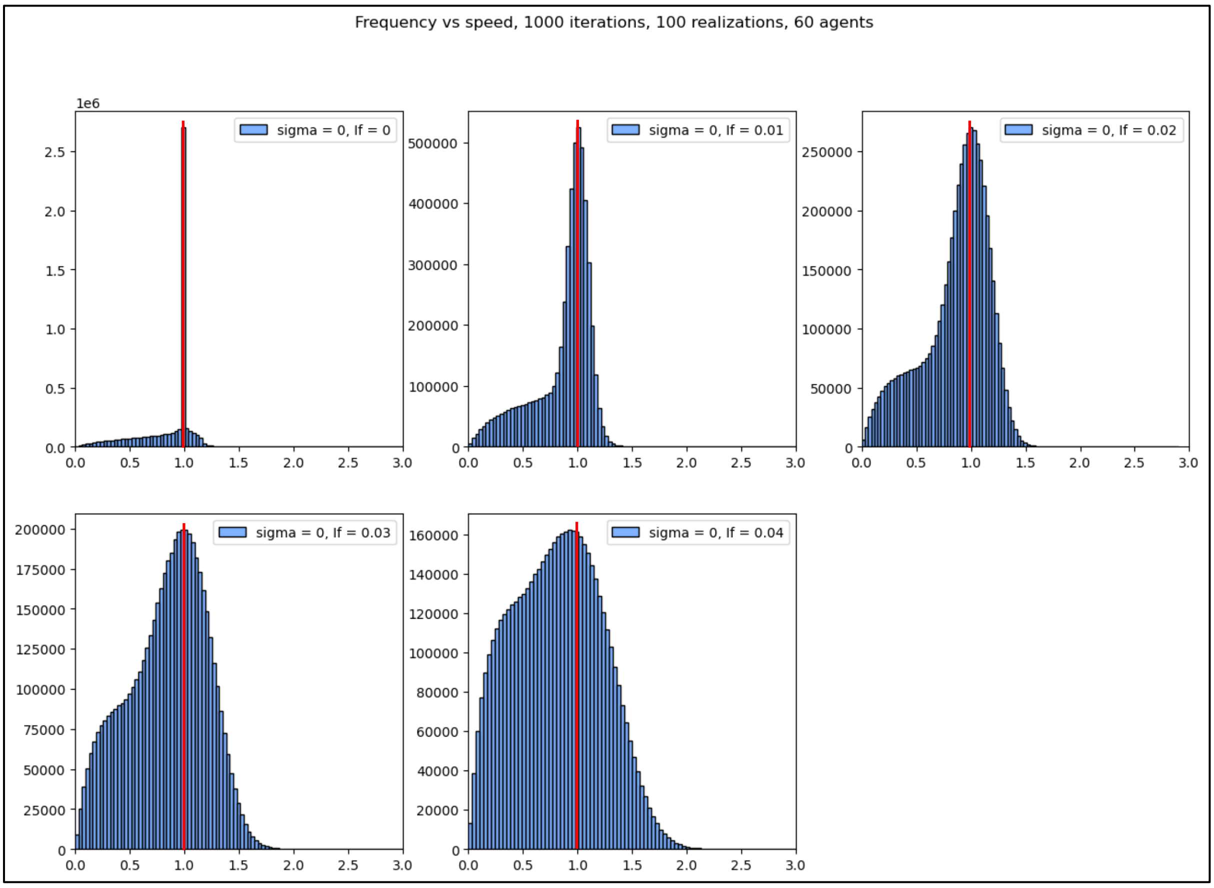

Figure 22. Distribution of observed speeds at $\sigma = 0$ (homogenous) for $I_f = 0$ to $I_f = 0.04$. The red line represents the intrinsic speed of the agents.

At $\sigma = 0$, the system is homogenous with an intrinsic speed of 1 (Figure 22). The speed distribution's width increases as we increase the strength of the hydrodynamic interaction. In the zero and low I_f cases, there is a small bump at low values of speeds. The bump is more prominent as we look at higher values of I_f. Also, as I_f increases, there are more agents at lower speeds than at higher speeds.

An observer looking at these speed distributions for $I_f > 0$ would wrongly infer that the systems are heterogeneous due to the spread in observed speeds, like in the case of a single realization. This spread increases with stronger hydrodynamic interactions, implying higher heterogeneity. The bump that is noticed can be attributed to collisions in the system for low I_f cases. This is because collisions are more likely to cause agents to slow down than hydrodynamic interactions. But as we take higher values of I_f the hydrodynamic interactions also slow down the agents on average, despite there being higher peak speeds. As the interactions get stronger, an agent interacts with more neighbours and along with collisions, this causes an overall decrease in the observed speeds.

For $\sigma > 0$, the system is heterogenous and intrinsic speeds are sampled from a distribution around 1. For low spread in heterogeneity ($\sigma = 0.2$, Figure 23), we notice again that there is a bump at lower speeds that isn't as prominent as in the homogenous case. The speed distribution again becomes wider as I_f increases. At high values of I_f the distribution loses its clear peak and becomes smooth. Overall, the mode of the distribution shifts towards the left of 1 as I_f increases.

The bump that we attributed to collisions becomes less prominent because of the heterogeneity. Due to intrinsic speed being sampled from a distribution around 1, there are agents that have low intrinsic speed that contribute to the bump that we see due to collisions. Similar to the homogenous case, at higher values of I_f, there are more agents with lower speeds. However, the heterogeneity additionally causes the mode of the distribution to shift to the left.

At intermediate spread in heterogeneity ($\sigma = 0.4$, Figure 24), some of the features that were prominent in the low heterogeneity case vanishes. The bump attributed to collisions is no longer visible. The increase in the width of the distribution is no longer as significant as in earlier cases. The mode of the distribution shifts to the left of 1 even for low values of I_f.

The changes that we see can be attributed to the increase in heterogeneity. More agents with lower intrinsic speeds cause the bump to disappear and the mode of the distribution to shift left.

At high spread in heterogeneity ($\sigma = 0.6$ and 0.8, Figures 25 and 26), the distribution's mode shifts significantly to the left. The shape of the distribution barely changes when the values of I_f increase. As I_f increases, the peak speeds observed also keep increasing, but it is barely visible.

At higher heterogeneity, the distribution from which intrinsic speeds are sampled gets truncated at 0 to prevent negative speeds, due to which more lower speeds are sampled. This is what causes the significant shift in the modes of the observed distribution. The agents with high intrinsic speeds are unable to maintain their speeds due to an increase in interactions. The hydrodynamic interactions may cause increases in speeds, but constant interactions with surrounding agents cause it to be short-lived.

A feature that is common across all degrees of heterogeneity is that in the absence of hydrodynamic interactions, the observed speed distribution is significantly less smooth.

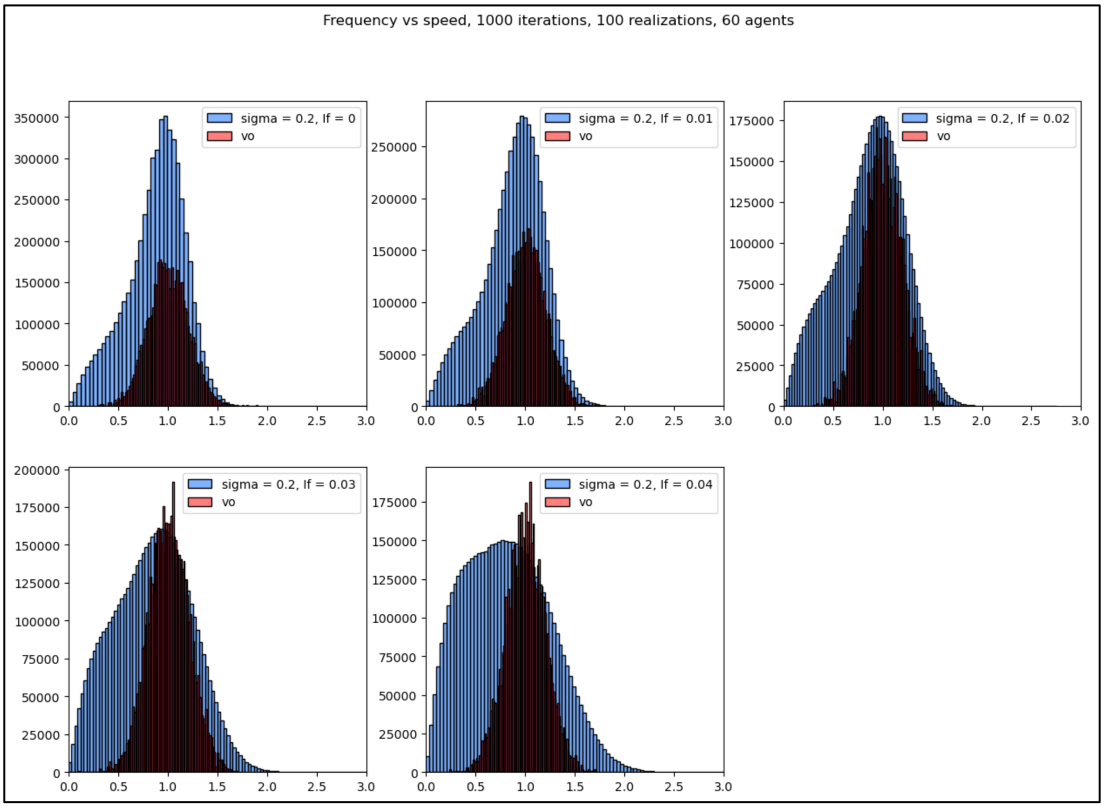

Figure 23. Distribution of observed speeds at $\boldsymbol{\sigma = 0.2}$*, for If = 0 to If = 0.04, along with intrinsic speed distribution* v_0

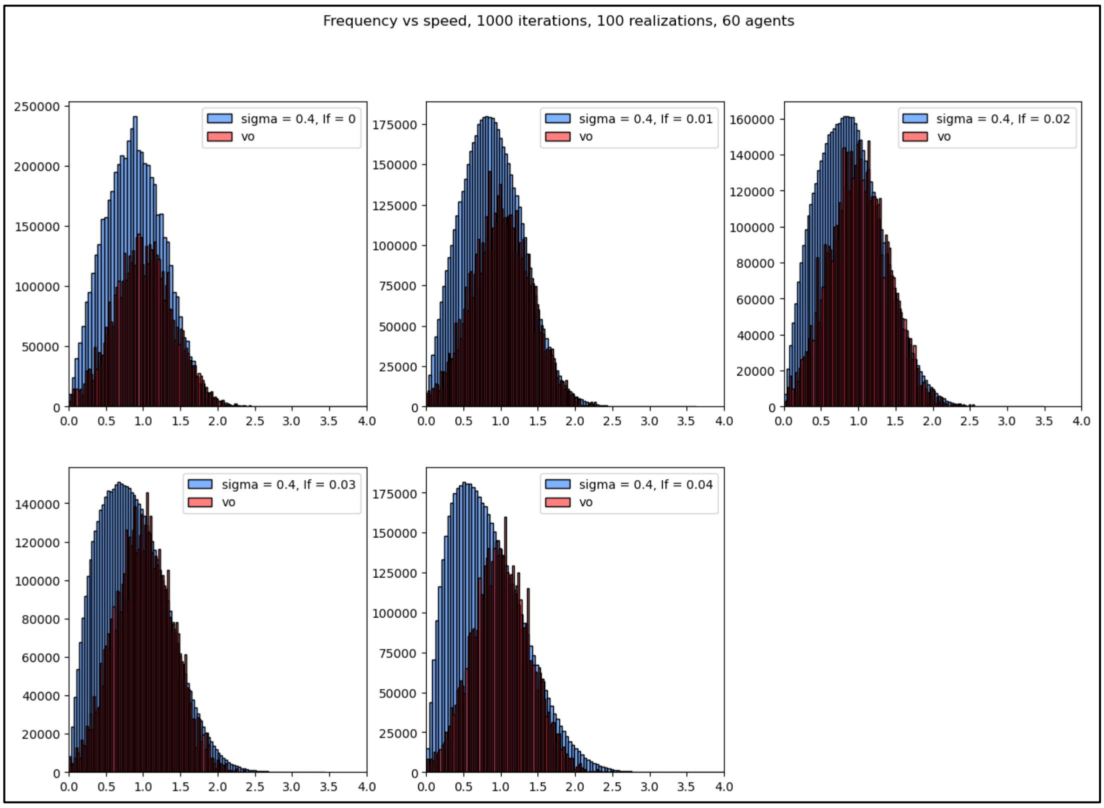

Figure 24. Distribution of observed speeds at $\boldsymbol{\sigma = 0.4}$*, for If = 0 to If = 0.04, along with intrinsic speed distribution* v_0

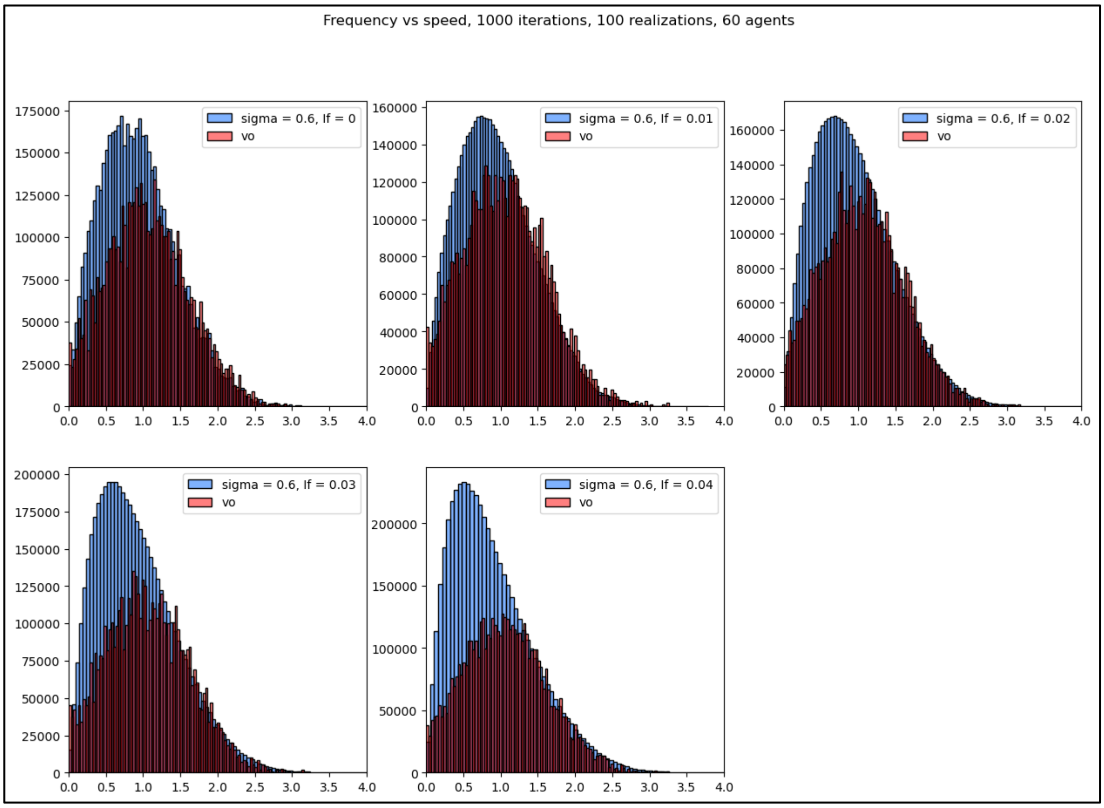

Figure 25. Distribution of observed speeds at ***σ = 0.6****, for If = 0 to If = 0.04, along with intrinsic speed distribution* v_0

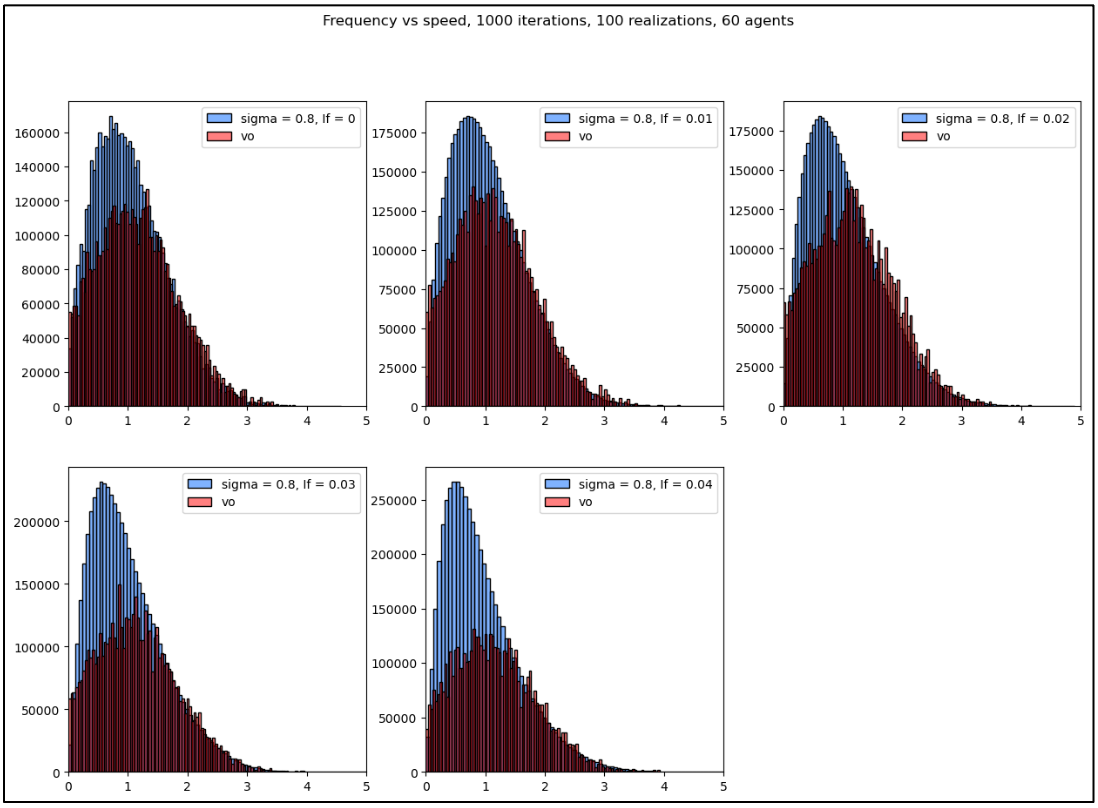

Figure 26. Distribution of observed speeds at ***σ = 0.8****, for If = 0 to If = 0.04, along with intrinsic speed distribution* v_0

3.2.2 Realization scale

At the realization scale we analyze the data obtained by an observer for one realization. This perspective allows us to study the system without data getting averaged or pooled over multiple datasets.

We have a look at a system of 60 agents run for a single realization of 1000 timesteps. The heterogeneity is set to $\sigma = 0.2$ and the hydrodynamic strength, $I_f = 0.02$.

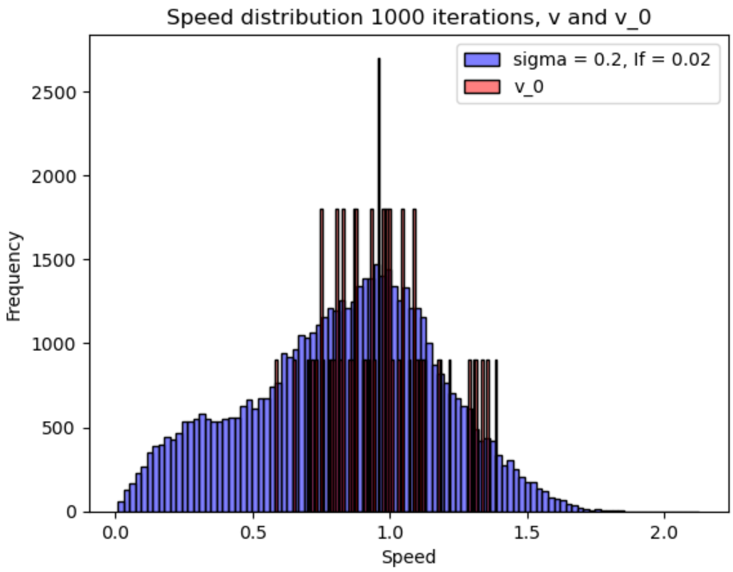

Figure 27. The distribution of intrinsic speed and observed speed for a single realization.

We notice that the intrinsic speeds are not sampled enough to get a smooth normal distribution since we have only 60 agents. The observed speed distribution, however, is relatively smooth due to the hydrodynamic interactions. The discreteness of the intrinsic speeds is lost, and the observed speed makes the system seem more heterogeneous than it actually is.

3.2.3 Agent scale

In the ensemble case, the agents' observed speeds were pooled across every time step across multiple realizations. The intrinsic speeds, however, remain the same for all time steps in a realization and only change across realizations. To compare the observed speeds with the intrinsic speeds directly, we took the time average of the observed speed in a realization and obtained it for multiple realizations to get more data points. In realistic systems, intrinsic speeds are not known to the observer. This analysis is done to understand the relation between what is observed and what is intrinsic to the model.

The average speeds were plotted against their corresponding intrinsic speeds, with intrinsic speeds on the x-axis and average speeds on the y-axis. This was plotted along with an x = y reference line. If a point was above the line, that agent's average speed over the single realization was above its intrinsic speed. Being below the line indicates that the agent's average speed was lower than its intrinsic speed. If a point lies on the line, that implies that the agent was able to maintain its intrinsic speed on average. The spread of points gives us an idea about the observed and intrinsic heterogeneity of the system. The spread of points along the x-axis indicates the range of the intrinsic heterogeneity of the system. The spread of points along the y-axis indicates the variation of average speeds that is observed.

The agents' average speeds and intrinsic speeds were stored for a system of 60 agents for 1000 time steps across 100 realizations. This was done for varying heterogeneity (σ) and hydrodynamic strength (I_f).

Taking σ = 0 case is trivial; all the data points would lie above $v_0 = 1$. We start from the low heterogeneity case at σ = 0.2 (Figure 28).

We see that for both the $I_f = 0$ and $I_f = 0.04$ cases, the spread of points along the x-axis is nearly the same. This is because they have the same extent of heterogeneity. At $I_f = 0$, there are a few points at lower values of v_0 lying above the line. For higher values of v_0, they tend to be below the line. At $I_f = 0.04$, the number of points above the line increases for higher v_0 but the points that were below the line at $I_f = 0$ go even lower and causes a larger spread of points.

More points below the x = y line than above indicate that most agents do not maintain their intrinsic speed and move slower on average. This is true even when there are no hydrodynamic interactions, and the effect is more pronounced when there are hydrodynamic interactions, with agents having even slower average observed speeds. The few points above the line can be attributed to the slower-moving agents getting dragged along by colliding with other agents for the $I_f = 0$ case. For $I_f = 0.04$, hydrodynamic interactions along with collisions cause slower-moving agents to move at a higher average speed. The range of observed speeds is roughly similar for both cases, implying that the extent of observed heterogeneity is similar for both, with If = 0.04 case having more agents that move slower.

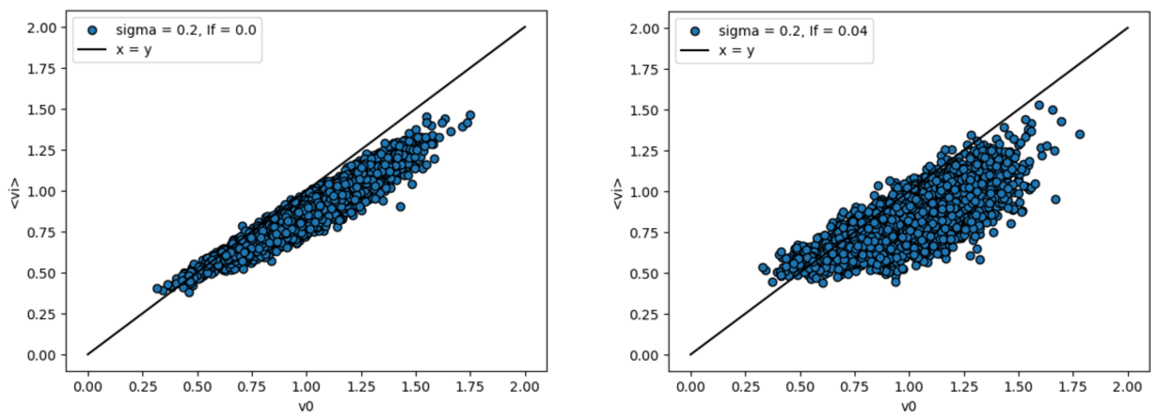

Figure 28. Average speed vs intrinsic speed, at σ = 0.2 , I_f = 0 and 0.04

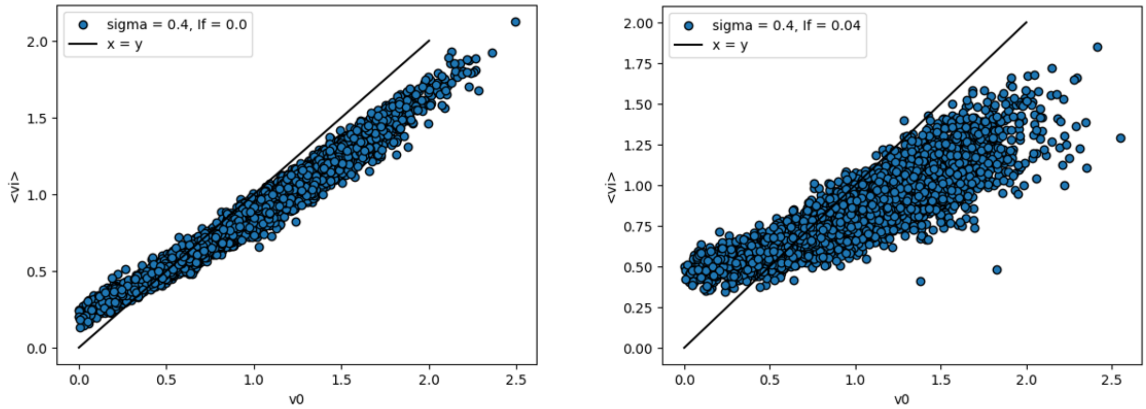

Figure 29. Average speed vs intrinsic speed, at σ = 0.4 , If = 0 and 0.04

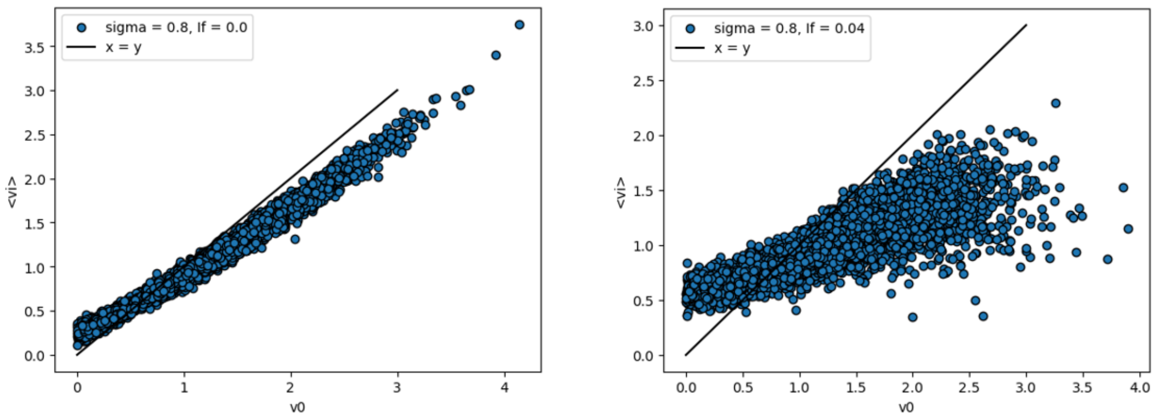

Figure 30. Average speed vs intrinsic speed, at σ = 0.8 , If = 0 and 0.04

For the higher heterogeneity cases at σ = 0.4 and 0.8 (Figures 29 and 30), we see that the spread of agents along the x-axis is similar for $I_f = 0$ and 0.04 cases because the extent of heterogeneity is the same for both. More points lie above the line for $I_f = 0$ at lower intrinsic speeds when compared to σ = 0.2 case. For $I_f = 0.04$, there is significantly more agents above the line at low intrinsic speeds. At higher speeds, the points below the line at $I_f = 0$ move much lower for $I_f = 0.04$. We also notice that at $I_f = 0.04$, there seems to be a short region at lower speeds where the points are arranged relatively flat.

Compared to the σ = 0.2 case, there are more agents above and below the line at $I_f = 0.04$. Due to the higher strength of the hydrodynamic interactions, more interactions between agents occur. As a result, agents with lower and higher intrinsic speeds struggle to maintain their intrinsic speeds and get pushed away from the x = y line. The flat region we see at lower intrinsic speeds is because of the lower bound of speeds at 0. At higher heterogeneity due to the intrinsic speeds being sampled from a truncated distribution, we have more agents with lower speeds. These lower-speed agents get pushed to higher speeds due to hydrodynamic interactions and form the flat region that is observed.

The range of observed speeds decreases compared to the $I_f = 0$ case. This is more noticeable at $\sigma = 0.8$. This implies that an observer sees less heterogeneity in speeds than there actually is in the system, and the average speed observed is slower than the intrinsic.

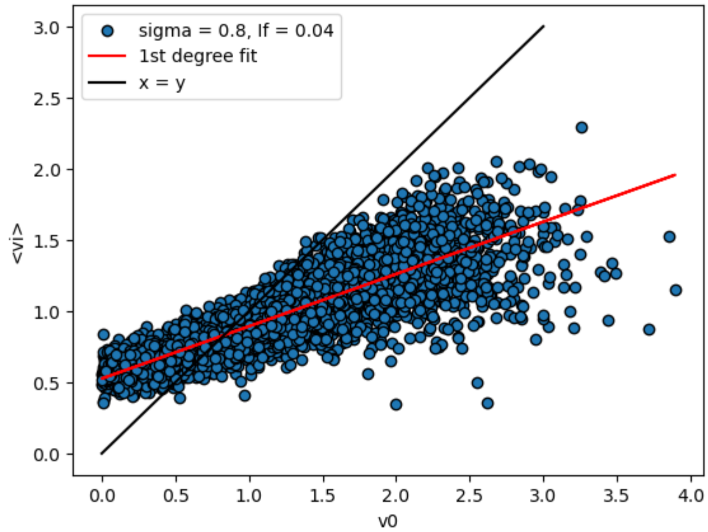

Figure 31. Average speed vs intrinsic speed, at σ = 0.8 , If = 0.04, along with a linear fit line

For $\sigma = 0.8$ and $I_f = 0.04$ case, the slope of the agents is low compared to the x = y line. The spread along the x-axis is always fixed since that is the intrinsic velocity. So, when the spread along y-axis changes, the slope of the points will also change accordingly. Having a lower spread along the y-axis means the slope of the line fit to the points will also be lower. So, a slope lower than the x = y line means that the perceived heterogeneity of speeds is lower than the actual heterogeneity of the system. The slope is a good indicator of what the perceived heterogeneity is like.

4. Conclusion and Outlook

4.1 Summary

We have modelled and investigated heterogeneity in collective systems with long-range hydrodynamic interactions to better understand how such heterogeneity may look like in realistic microscopic systems. We have demonstrated that these interactions significantly influence the observed properties of the system. A homogenous system would appear heterogeneous in speed due to hydrodynamic interactions. We attempted to extract the intrinsic speed of the system by constructing a parameter from observed data that would quantify the interactions acting on an agent. We were unable to construct a parameter that could capture the complex interactions in our system while remaining agnostic to the model. We then pivoted to introducing heterogeneity to the system by sampling each agent's intrinsic speed from a normal distribution and studying how the observed speeds were affected. We studied it at three scales; the ensemble, realization, and agent scale. Across these scales, we studied how the heterogeneity is perceived externally compared to the intrinsic heterogeneity present within the system. By varying the strength of these long-range interactions and heterogeneity in the system, we were able to characterize how the observed dynamics shift and how the relationship between intrinsic properties and observed behaviour changes.

4.2 Discussion

The hydrodynamic interactions are a key part of what makes the study of microscopic collectives an interesting topic. They are able to significantly affect the observed properties of the agents within the system. A system with homogenous intrinsic speed looked heterogenous to an observer due to the variations in speeds of the agents caused by the hydrodynamic interactions. As discussed in Schumacher et al., complex interactions can cause an observer to wrongly assume heterogeneity when the system is homogenous.

The hydrodynamic interactions also affect how heterogenous systems are perceived. A main observation of the ensemble scale analysis was that observed speeds of the agents tended to be lower than the intrinsic speeds. As the strength of the hydrodynamic strength increased, there was significant decrease in observed speeds. The hydrodynamic interactions can aid the speeds of the agents as we notice an increase in peak speeds, but mostly end up hindering them. At stronger levels of hydrodynamic interactions, even the extent of heterogeneity in the system gets masked; a highly heterogeneous system would seem less heterogeneous to an observer looking at their speeds.

So, the presence of hydrodynamic interactions in a system can cause an observer to assume heterogeneity in a system that is homogenous and can also do the near opposite by underestimating the extent of heterogeneity in a highly heterogeneous system. We need to improve observation methods and equip observers with more information to correctly characterize the extent of intrinsic heterogeneity in these systems.

4.3 Future directions

Since microscopic systems are low inertia in nature, the observed speed of an agent can be exactly decoupled into its intrinsic speed and the interactions acting on it. Therefore, working on developing a parameter capable of accurately capturing the complex nature of these long-range interactions—and using it to extract the intrinsic speed from the distribution of observed velocities—remains a possible avenue for future work.

As for the model itself, future work could involve introducing heterogeneity in different ways, such as differences in response to environmental cues or chemotactic behaviour, to further simulate the complexity found in biological systems. Such efforts could improve the model's applicability to a broader range of real-world systems and deepen our understanding of how heterogeneity is shaped and masked by interaction dynamics.

5. Appendix

5.1 Using PCA to classify heterogenous Vicsek model

As part of reading about different methods of classification of heterogenous collectives, I recreated the model used by Tan et al. (15), in which principal component analysis (PCA) is the method used to classify heterogeneity in the system.

Heterogeneity is introduced in the Vicsek model by having agents with differing intrinsic speeds or noise. The data matrix on which PCA is done contains tangents of the agents' orientation angles. I used the sine and cosine of angles instead and generated a dataset twice as large to do PCA on. I was able to replicate the results they obtained using this new dataset (16).

Additionally, in the paper, they only tested for heterogeneity in interaction radius and intrinsic speed. I further tested it on the heterogeneity of the interaction radius using the same dataset. PCA was unsuccessful in separating the two speeds. The orientation data, which is what we use may not be enough to separate information about interaction radius.

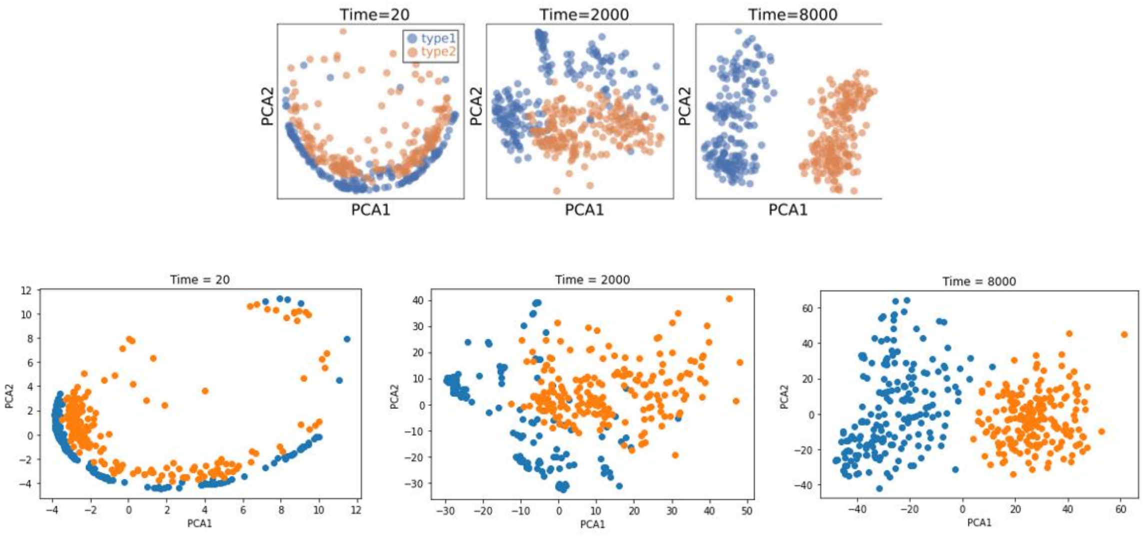

Figure 32. PCA on Vicsek model with heterogeneity in noise. The images on top are from the original paper (6) and the ones on the bottom are recreated by me. The formation of clusters is clearer when longer time datasets are used

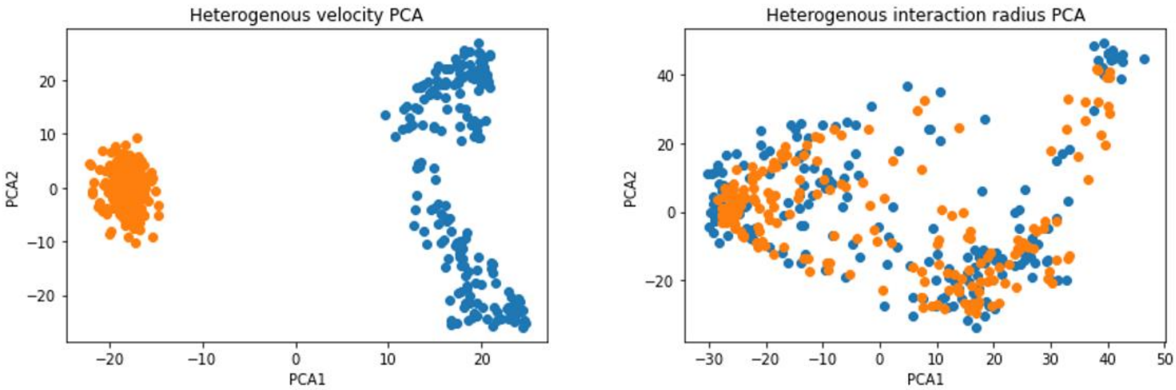

Figure 33. PCA on Vicsek model with heterogeneity in velocity (left) and interaction radius (right). PCA is unable to separate the agents with different interaction radii

Algorithm of the code (Full code is in (15)):

Input – the simulation time, the cutoff time (time after which observations are stored), length of time step, length of square arena, number of particles of each type (two types total), the intrinsic speed, interaction radius and intrinsic noise of each type,

- Initialise positions and orientations of all agents.
- Simulate particle dynamics:
  - Loop over number of time steps
  - Update velocity of the agents
    - Loop over number of agents
    - Calculate distance between focal agent and every other agent
    - Get list of neighbours within focal agent's interaction radius
    - Get average orientation of the neighbours that the agent must align with
    - Add noise to obtained angle
    - Multiply cosine and sine of the with intrinsic speed to get x and y velocities of the agents
  - Update positions of the agents
    - Velocity multiplied with timestep gives new position
    - Adjust for periodic boundaries
  - When t > cutoff time:
    - Store data matrix containing cosine and sine of angles to do PCA on.
    - Store position data to animate simulation

- Print position data onto a file to animate
- Obtain principal components of data

5.2 The 2D Couzin model

The Couzin model was used to simulate animal group behaviour. It was originally done in 3D. Each agent has a zone of repulsion (ZOR), orientation (ZOO) and attraction (ZOA), which affects how an agent reacts to its neighbours. These interactions lead to interesting dynamics. I have coded the system in 2D, still maintaining the key features of the system.

Algorithm of the code (Full code is in (16)):

Input: Simulation time, number of agents, radius of repulsion, width of region of orientation, width of region of attraction, noise factor, angle of vision.

- Initialize positions and angles of all agents
- Simulate agent dynamics:
  - Loop over number of time steps
  - Obtain matrix of neighbours for each zone (repulsion, orientation attraction)
  - Checking for zone of repulsion:
    - Loop over number of agents
    - If an agent has neighbours in ZOR, obtain the change in angle needed to repel from these neighbours
  - Checking for zone of orientation:
    - Loop over number of agents
    - First check if an agent has neighbours in zone of repulsion. If yes, skip.
    - Only for agents with no neighbours in ZOR, identify the neighbours in the ZOO and obtain angle needed to orient with these neighbours
  - Checking for zone of attraction:
    - Loop over number of agents
    - First check if an agent has neighbours in zone of repulsion. If yes, skip.
    - Only for agents with no neighbours in ZOR, identify the neighbours in the ZOA and obtain angle needed to move towards these neighbours
  - Get the average angle each agent needs to adjust and add noise
  - If the new angle is larger than the turning rate, replace new angle of the agent with turning rate. Else keep new angle.
  - Update velocity by multiplying cosine and sine of angles with intrinsic speed
  - Update position by multiplying velocity with time step
- Store position data to animate

5.3 The heterogeneous microscopic collective

This model was created using experience gained from coding other agent-based models and was used to create the results discussed in this thesis.

Algorithm for the model (Full code in (10)):

Input variables: Number of time steps, length of square arena, number of agents, number of initialization steps, mean of the intrinsic speed distribution, standard deviation of the intrinsic speed distribution, noise factor, hydrodynamic strength, spring constant, radius of the agent.

- Initialize the agent positions and orientations
- Update the velocity of each agent according to initial conditions
- Run simulation:
    - Loop over number of time steps
    - Calculate collision force between agents:
        - Loop over each agent
        - Check if every other agent is within focal agent's range to collide with
        - Calculate the force each agent applies on the focal agent
    - Update velocity of each agent
        - Calculate the hydrodynamic field acting on each agent by every other agent
        - Also obtain the rotational field
        - To the angle of each agent add the rotation due to the hydrodynamic field and a random noise
        - Calculate velocity of each agent by adding intrinsic speed times cosine and sine of new angles, the force acting on each agent, and the velocity field.
    - From updated velocities, calculate current speed
    - Update position (v*dt)
    - When t > number of initialization steps:
        - Store current speed into an array
        - Store intrinsic speed into an array
        - Store position data to animate
    - Write stored position data onto a file to animate

6. References

1. **Vicsek, Tamas, et al.** PH YS ICAL REVIEW LETTERS 7 AUGUs~ 1995 Novel Type of Phase Transition in a System of Self-Driven Particles.

2. Collective Memory and Spatial Sorting in Animal Groups. **Couzin N W Iain D, Krausew Jens,Jamesz Richard, Ruxtony Graeme D, Franksz Nigel R.** 2002, J. theor. Biol, Vol. 218, pp. 1-11.

3. The importance of individual variation in the dynamics of animal collective movements. **Delgado Maria del Mar, Miranda Maria, Alvarez Silvia J.,Gurarie Eliezer, Fagan William F., Penteriani Vincenzo, di Virgilio Agustina, Morales Juan Manuel.** 1746, s.l. : Royal Society Publishing, 19 5 2018, Vol. 373.

4. Roles of leader and follower cells in collective cell migration. **Qin Lei, Yang Dazhi, Yi Weihong,Cao Huiling, Xiao Guozhi.** 14, s.l. : American Society for Cell Biology, 1 7 2021, Vol. 32, pp. 1267-1272.

5. Disentangling intrinsic motion from neighborhood effects in heterogeneous collective motion. **Nabeel Arshed, Masila Danny Raj.** 6, 1 6 2022, Chaos, Vol. 32.

6. Intrinsic statistical separation of subpopulations in heterogeneous collective motion via dimensionality reduction. **Tan Pei, Miles Christopher E.** 1, 1 1 2024, Physical Review E, Vol. 109.

7. **Schumacher Linus J.** 23 8 2017, Cell Systems, Vol. 5, pp. 119-127.

8. Advances of medical nanorobots for future cancer treatments. **Kong Xiangyi, Gao Peng, Wang Jing, Fang Yi, Hwang Kuo Chu.** 1, s.l. : BioMed Central Ltd, 1 12 2023, Journal of Hematology and Oncology, Vol. 16.

9. Model of Collective Fish Behavior with Hydrodynamic Interactions. **Filella Audrey, Nadal François, Sire Clément, Kanso Eva, Eloy Christophe.** 11 5 2018, Physical Review Letters, Vol. 120.

10. Heterogenous microscopic collective. **Balagopal.** s.l. : Github, 2025, Vols. https://github.com/Balagopal-R-Nair/Heterogenous-Vicsek-Model.

11. Visualization and analysis of atomistic simulation data with OVITO-the Open Visualization Tool. **Stukowski, Alexander.** 2010 : Modelling and Simulation in Materials Science and Engineering.

12. **Rob Phillips, Theriot, Jane Kondev and Julie.** “Physical Biology of the Cell”. 2009, p. 29.

13. **Maeda, Kayo, et al.** Effect of Temperature on Motility and Chemotaxis of Escherichia coli. 1976. pp. 1039-1046.

16. Heterogenous Vicsek model. **Balagopal.** s.l. : Github, 2025, Github repository, Vols. https://github.com/Balagopal-R-Nair/Heterogenous-Vicsek-Model.

17. Couzin model 2D. **Balagopal.** 2025, Github repository, pp. https://github.com/Balagopal-R-Nair/Couzin-Model.