Newcastle University

MMath Project

Modeling the Interactions Between Animals in a Group

Author:
Samuel ford

Supervisor:
Dr. Andrew fletcher

May 1, 2015

Abstract

Mathematically modelling animal behaviour in groups is becoming increasingly common. This project explores a research paper by Couzin et al. (2005) and the results given by them. An attempt is then made to recreate their results using code created in Matlab, before going on to modify the governing equations given by Couzin et al. (2005). Modifications are made by altering: the accuracy measure, the distance measure, which is used in determining how the individuals interact, from a metric distance model to a topological model and by implementing a variable speed. Using the results produced by these modifications, conclusions about how animals interact within a group and on the methods used to construct the models are then drawn.

Contents

Chapter 1

Introduction

How individual animals make collective decisions in a group is a phenomenon that has proved highly complex and baffled scientists for years. Indeed, moving animal groups provide some of the most intriguing and difficult to characterise examples of collective behaviour [1]. Several different classes of animals travel in large groups, for example many species of oceanic fish band together in large shoals that can span tens of kilometres and involve hundreds of millions of individuals [2] or anywhere from 25,000 to 40,000 honeybees seeking a new dwelling in the event of a new queen being born [3].

Figure 1.1: *Demonstration of collective behaviour in a large group of animals: fish in a tightly grouped shoal [4].*

For animals in a group, the ability to make decisions about the direction of travel depends on how interactions occur between neighbouring animals. Studies have shown that guppies, *Poecilia reticulata*, can learn the route to a food source by shoaling with knowledgeable conspecifics [5]. In many cases, information is transferred by a physical signal, for example scout honeybees use the full power of the waggle dance to inform their nest-mates about the distance and direction of a potential nest-site [6]. In some cases, however, information transfer via signalling is impractical, for example when migrating groups of fish, ungulates, insects and birds are considered, where crowding limits the range over which individuals can detect one another [7].

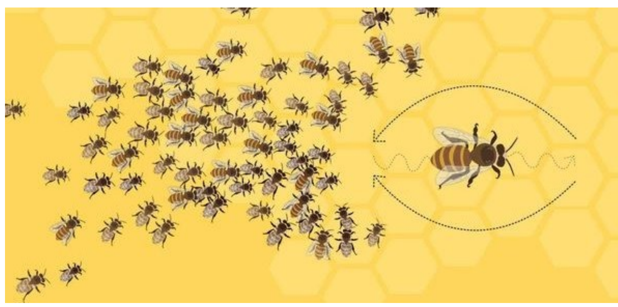

Figure 1.2: *Demonstation of how a single bee will use the waggle dance to influence the movement of a group[8].*

When a group of animals wish to reach a specific destination, the number of informed individuals (i.e. the number of individuals who know the location of the destination) within the group has an effect on the ability of the group to reach its desired location. Often the informed individual can be a single scout, responding to the diverse information they have personally obtained about the quality of a potential nest-site [6]. This raises the question of how, when the proportion of informed individuals is low in comparison to group size, a group is capable of travelling to its chosen destination. A mathematical model used to model animal interactions can help draw conclusions about collective animal behaviour. Mathematical modelling is becoming increasingly recognized as an important research tool when studying collective behaviour [9].

This project looks at a mathematical model devised by Couzin et al. (2005) which aims to model the behaviour of animals in a group when a select number have information about a desired location. In Chapter 2 the equations set out by Couzin et al. (2005) are discussed, describing the results produced, before creating a code that attempts to replicate their results. In Chapter 3 and 4 the code is altered to take into account some parameters not looked at by Couzin et al. (2005).

Chapter 2

The Reference Model

"Effective leadership and decision-making in animal groups on the move", by Iain D. Couzin, Jens Krause, Nigel R. Franks and Simon A. Levin is the foundation of this report. Couzin et al. (2005) investigate the social interactions between animals in a group, how this enables a group to move and the effect leaders have on the direction of motion. They then apply simple equations, using a self-propelled particle concept, to mathematically model behaviour observed in animal groups.

Couzin et al. (2005) show that a larger group needs a smaller proportion of informed individuals to move in the preferred direction. They also show that information can be communicated between individual members of the group without a need for signalling e.g. the waggle-dance of the honeybee [3], for a group to reach a desired location.

This Chapter presents the equations used by Couzin et al. (2005), describing how this effects the interaction between two neighbouring animals and how the equations work with each other to describe the movement of a group. The Chapter will also describe the methods used to measure accuracy, the parameters used and the results generated.

2.1 Governing Equations

Individuals in the modelled group of size N update their position vector $\mathbf{c}(x, y)$ according to two rules: collision avoidance and alignment with neighbours. An individual's first priority is to avoid collision with its neighbours. If two individuals get within a distance A of each other they will seek to move away, which is modelled using,

$$\mathbf{d}_i(t + \Delta t) = -\sum_{j \neq i} \frac{\mathbf{c}_j(t) - \mathbf{c}_i(t)}{|(\mathbf{c}_j(t) - \mathbf{c}_i(t))|}. \tag{2.1}$$

Each individual i has a 2-dimensional position vector $\mathbf{c}_i(t)$. Neighbouring animals within the collision distance A are represented by j, Figure 2.1, and have position vector $\mathbf{c}_j(t)$. The difference in positon between the individual in question i and its neighbour is $\mathbf{c}_j(t) - \mathbf{c}_i(t)$ and then normalised by $|(\mathbf{c}_j(t) - \mathbf{c}_i(t))|$. This creates a 2-dimensional direction vector for i which points directly to j. Each direction vector of i to a j is then summed to give a direction vector which would move i towards the average position of all js within A. Then the negative of this vector is taken ensuring i is moving away from each j. This gives us $\mathbf{d}_i(t + \Delta t)$, which is the direction vector at the next time step.

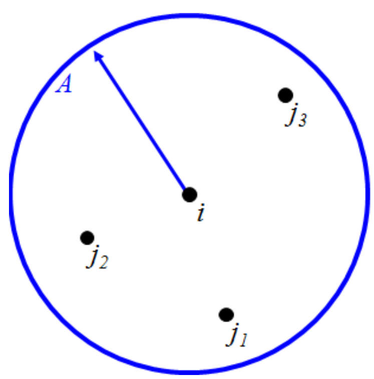

Figure 2.1: *Demonstration of the collision avoidance distance A for i and its neighbours j.*

Avoiding a collision is the first priority of any individual. If i does not find any individuals within A, it checks for any individuals within the alignment distance B, and aligns its direction vector with these neighbours. Individuals in a group want to stay together and move in the same direction, whether it is for protection from a predator [10] or for foraging purposes [7]. This interaction is governed by,

$$\mathbf{d}_i(t+\Delta t) = \sum_{j \neq i} \frac{\mathbf{c}_j(t) - \mathbf{c}_i(t)}{|(\mathbf{c}_j(t) - \mathbf{c}_i(t))|} + \sum_{j=1} \frac{\mathbf{v}_j(t)}{|\mathbf{v}_j(t)|}. \tag{2.2}$$

As before, i is the individual that is searching for others within the attraction distance. Now j represents all of the individuals within B, Figure 2.2. Again, the difference between the position vectors of i and each j are calculated and summed. Unlike in Equation 2.1, this value remains positive ensuring the individual moves towards the average positions of its neighbours. The second term in Equation 2.2 uses a 2-dimensional direction vector $\mathbf{v}_j(t)$, $\frac{\mathbf{v}_j(t)}{|\mathbf{v}_j(t)|}$, This term sums the direction vectors of all neighbours within B and calculates the average. This allows the individual i to align itself with the average direction of the interacting neighbours. Summing the two terms together gives $\mathbf{d}_i(t+\Delta t)$. This direction vector moves i towards its neighbours as well as travelling in the same direction as them.

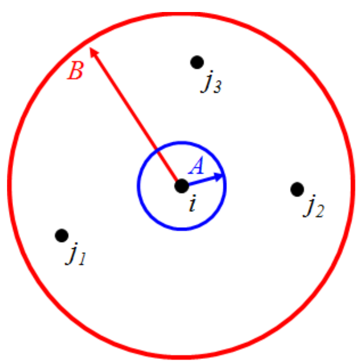

Figure 2.2: *Demonstration of collision avoidance distance A and the alignment distance B around i and the interaction with neighbours j.*

For both Equation 2.1 and Equation 2.2, the 2-dimensional direction vector, $\mathbf{d}_i(t+\Delta t)$, needs to be converted to a unit vector if it is to be used to calculate an individual's new position vector,

$$\widehat{\mathbf{d}}_i(t+\Delta t) = \frac{\mathbf{d}_i(t+\Delta t)}{|\mathbf{d}_i(t+\Delta t)|}. \tag{2.3}$$

Equation 2.1 and Equation 2.2 control how the animals interact with each other. To allow for informed individuals, who know about the preferred direction of travel, another equation is used,

$$\mathbf{d}'_i(t+\Delta t) = \frac{\widehat{\mathbf{d}}_i(t+\Delta t) + \omega \mathbf{g}_i}{|\widehat{\mathbf{d}}_i(t+\Delta t) + \omega \mathbf{g}_i|}. \tag{2.4}$$

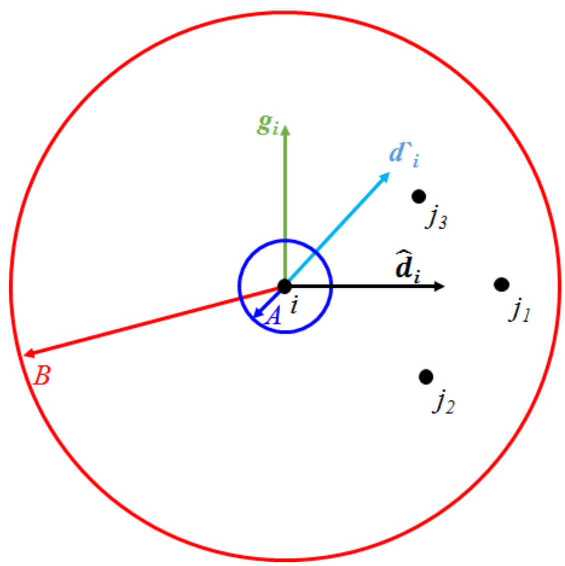

Figure 2.3: *Demonstration of Equation 2.4.* $\boldsymbol{g}_i$ *is the preferred direction,* $\widehat{\boldsymbol{d}}_i$ *is the direction of travel towards the neighbours of* i *and* $\boldsymbol{d}'_i$ *is the direction of travel taken.*

Within any group there will be a proportion of informed individuals, p, who are aware of the preferred direction of travel for the group. For these individuals, $\mathbf{g}_i$ is the preferred direction as a unit vector, for uninformed members of the group $\mathbf{g}_i$ is zero. ω is a weighting term, which controls the influence that the preferred direction has on the direction of travel taken. If $\omega = 0$ then $\mathbf{g}_i$ has no influence over direction. When $\omega = 1$ the preferred direction has an effect equal to that of the interaction Equations 2.1 and 2.2. For values of ω greater than 1 the preferred direction dominates travel direction, Figure 2.3.

Finally, a new position vector must be calculated. This is done simply by taking the new direction vector $\mathbf{d}'_i(t+\Delta t)$, multiplying it by speed s_i and adding the result to the current position vector $\mathbf{c}_i(t)$, shown in Equation 2.5.

$$\mathbf{c}_i(t + \Delta t) = \mathbf{c}_i(t) + s_i \Delta t \mathbf{d}'_i(t + \Delta t). \tag{2.5}$$

2.2 Methods and Initial Parameters

To allow for uncertainty in the motion of the animals, the direction vector $\mathbf{d}'_i(t + \Delta t)$ is altered by rotating it by an angle generated from a circular-wrapped Gaussian distribution, with mean = 0 and standard deviation = 0.01 radians.

$$Uncertainty = N(0, 0.01). \tag{2.6}$$

To prevent individuals turning through an angle that would be too great for any animal to manage, a maximum turn angle is set. This is given by $\theta \Delta t$.

$$Maximum = \theta \Delta t. \tag{2.7}$$

If the change in the angle between the old direction vector $\mathbf{v}_i$ to $\mathbf{d}'_i(t+\Delta t)$ is less than $\theta \Delta t$ then the new direction vector remains unchanged. If, however, the angle between the two is greater than $\theta \Delta t$ then it moves $\theta \Delta t$ from the old direction vector towards the new vector.

Each individual begins at a random positon and with a random orientation generated from a uniform distribution from -π to π. The parameters of the standard model are: $A = 1, B = 6, \omega = 0.5, \Delta t = 0.2, \theta = 2$ and $s_i = 1$.

Using Equations 2.1 to 2.7, together with the given parameters, setting group size to $N = 2$, proportion of informed individuals to $p = 0.5$ and preferred direction to $\mathbf{g}_i = (0, 1)$ for informed individuals results in Figure 2.4.

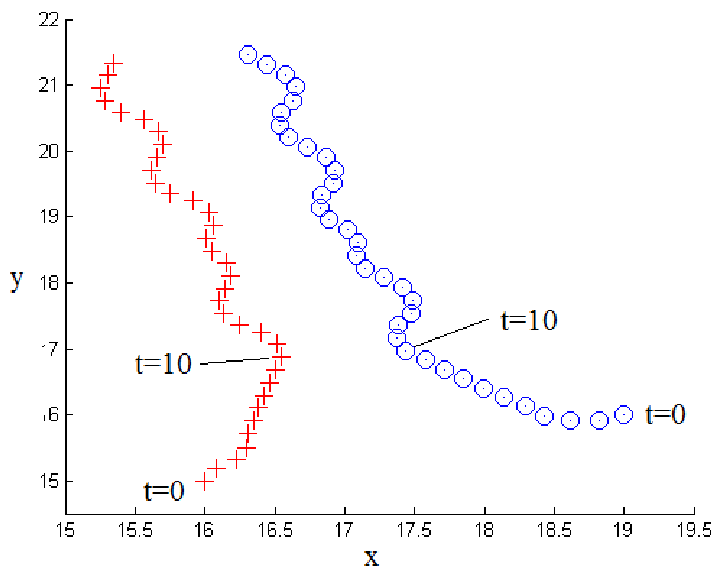

Figure 2.4: *Demonstration of governing equations and methods. Each symbol is one time step apart, with red being the informed individual and blue uninformed.*

The individuals initially start a distance apart that is greater than A but less than B. They will therefore be both subject to Equation 2.2 and will attempt to align. They both begin to move towards each other whilst also attempting to travel in the same direction. The informed individual is subject to Equation 2.4 and therefore attempts to maintain the preferred direction $\mathbf{g}_i$, which in this case is directly up. This results in the informed individual traversing less towards the uninformed individual and more towards the top of the graph.

After around 10 time steps, the two individuals get within A of each other. At this point they attempt to move apart. If there is no limit to their change in angle, $\theta\Delta t$, they would move directly apart. For this model however, with a limit imposed they gradually move away from each other. A handful of time steps later, both individuals have moved out of A but remain within B, resulting in them once again being drawn together.

This process repeats itself, with the informed individual attempting to maintain its preferred direction. Despite the fact that neither individual is able to signal whether it has knowledge of the preferred direction, or what the preferred direction might be, both individuals travel in roughly the same direction. This reveals that direct information transfer between individuals is not necessary to achieve the preferred direction of travel.

To understand how the number of informed individuals and group size affects the ability of the group to achieve its preferred direction, Couzin et al. (2005) introduce a measure for accuracy. This determines how well the group is orientated towards their preferred direction given by $\mathbf{g}_i$. A vector is taken from the group's centroid 50 time steps before the end, extending to the groups centroid at the final time step. The angle between this vector and the preferred direction is then determined.

For each N and each p the angle is calculated 400 times. The 400 runs are then normalized so that the minimum value of the runs is 0, corresponding to no information transfer (groups move in random directions), and the maximum value of the runs is 1, corresponding to the motion of the simulated groups always being exactly aligned with $\mathbf{g}_i$. A mean of the 400 normalised runs is then calculated giving the accuracy for that N and p [7].

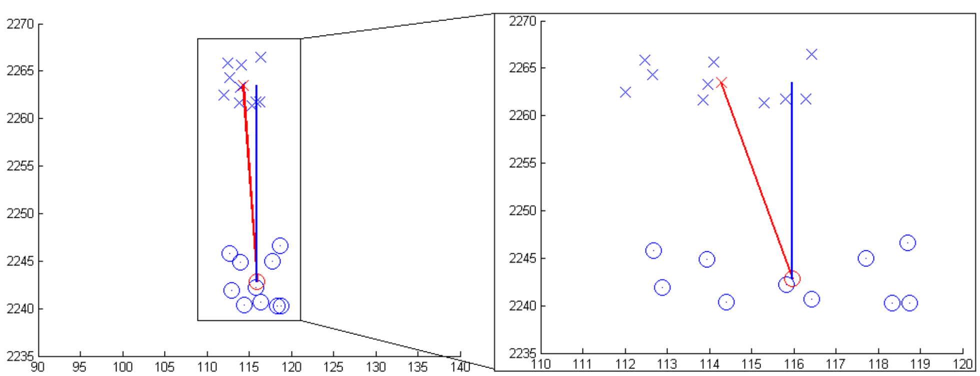

Figure 2.5: *Demonstration of accuracy measure, X the individuals at the final time step, O the individuals 50 times step before the end and red representing average position. The blue line represents the preferred direction of travel and the red line the actual direction of travel.*

Figure 2.5 demonstrates the accuracy measure. Here, $N = 10$, $p = 0.7$ and the previous methods and parameters remain in place. One line is drawn between the two average positions (red) and another drawn from the average position of the group in the direction of $\mathbf{g}_i$ (blue). For this case, an angle deviation of 0.06 radians is obtained. The code would be ran 400 times and an overall average of the angle deviations would be taken.

2.3 Reference model results

Using all the equations set out in Section 2.1 and the methods and parameters in Section 2.2, Couzin et al. (2005) obtained Figure 2.6.

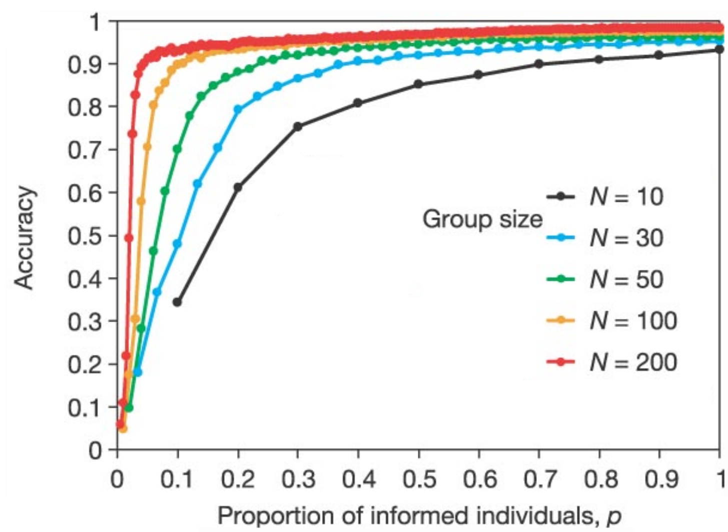

Figure 2.6: *Group accuracy as a function of the proportion of informed individuals p, for different group sizes N [7].*

Group sizes here are comparable to the size of schools, flocks or herds of many species, but smaller than large aggregates such as honeybee colonies [7]. This owes to the nonlinear increase in computer processing time required as N increases [7]. Immediately apparent for all group sizes is that an increase in p results in an increase in accuracy. This is expected, as if there is an increase in the number of individuals who know the preferred direction, more will reach the chosen destination regardless of where the uninformed individuals travel.

With a larger group, a smaller proportion of informed individuals are required to achieve a specific level of accuracy. This appears counterintuitive at first, but for a larger group it is harder for any individual to move out of the range of B and therefore leave the group and travel randomly. Any informed individual is also likely to have a greater number of individuals surrounding it, and they in turn will have a greater number of individuals around them. This allows for information to be transferred among the group with greater ease and therefore an informed individual has a greater effect on the group.

2.4 A new measure of accuracy

This section deals with replicating the results from the information provided by Couzin et al. (2005). This is done to assess whether the code described in Appendix A is performing correctly. Matlab is the software package used. The equations and parameters set out previously are used, however a change to the accuracy measure is implemented. This is done to preserve the information about the range of values that the runs give for accuracy and therefore give us more information about the variability at each N and p. A centroid is still drawn from 50 time steps before the end to the final time step. The angle between the two is still calculated. Now however, a score from 0 to 1 is assigned to this angle: 1 if it is perfectly aligned to the preferred direction and 0 if it is π away from the preferred direction. A mean is then taken of all runs for a given group size and proportion.

Due to the limitations of the technology available, only models containing group sizes of $N = 10, 30, 50$ were possible. It took approximately 18 hours for 50 runs with $N = 10$ and approximately 132 hours for 6 runs with $N = 50$. Therefore, due to time constraints, there are 200 runs for each value of p with $N = 10$, 11 runs for each value of p with $N = 30$ and 6 runs for each value of p with $N = 50$. This resulted in random variation having a large effect when $N = 50$ and to a lesser extent when $N = 30$. Therefore the decision is taken for polynomial regression to be implemented on all 3 group sizes, in an attempt to produce a smooth curve. This is done using the software package R.

2.4.1 The effect changing the proportion of individuals has on accuracy, for different group sizes

Group size $N = 10$

When $N = 10$, 200 results were produced for each proportion of informed individuals p. At each p the mean is taken and plotted on the raw data, with a line drawn between each mean shown in Figure 3.1.

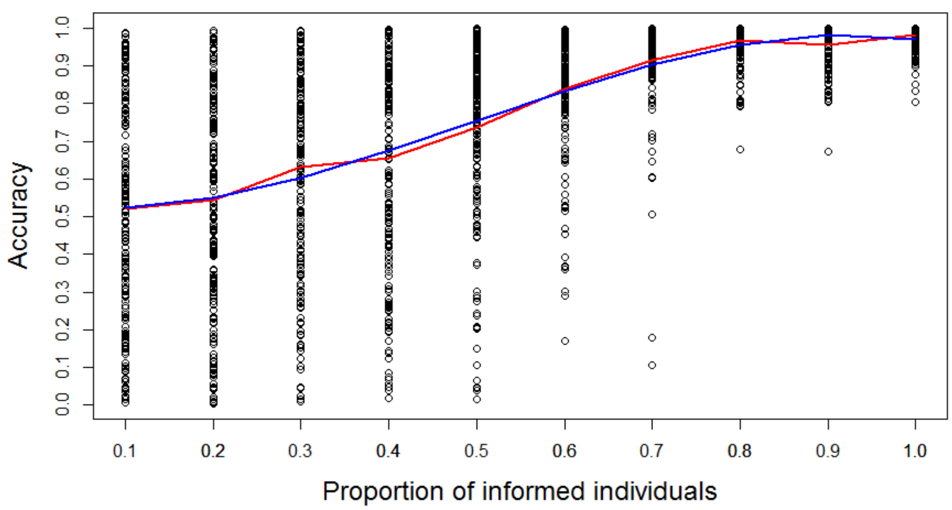

Figure 2.7: *Accuracy against the proportion of informed individuals with a polynomial regression line (blue) and a line between the means (red) when* $N = 10$.

As discovered by Couzin et al. (2005), as the proportion of informed individuals p is increases there is also an increase in accuracy. There are a few examples where the accuracy is higher in a lower proportion of informed individuals p. This is probably caused by the limited number of runs. Looking at a box plot of the data will reveal outliers that may be having an adverse effect.

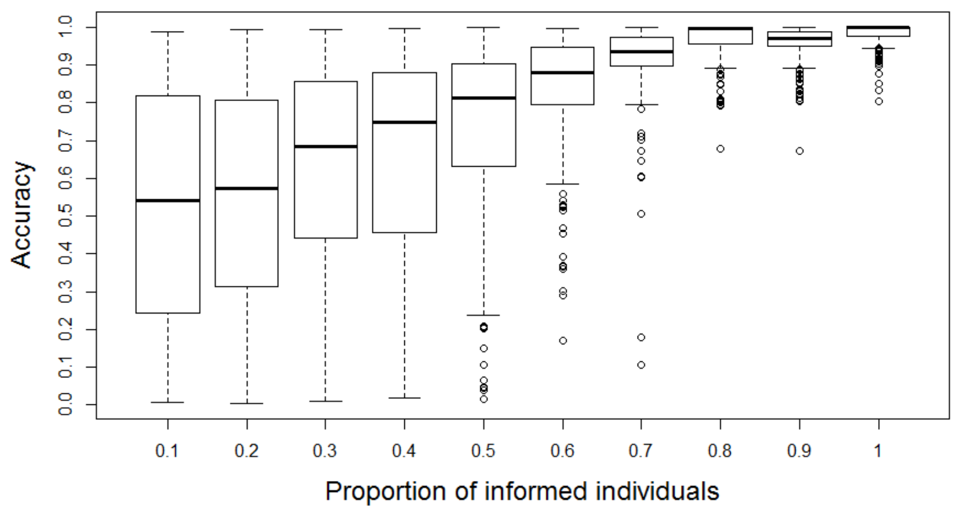

Figure 2.8: *Box plot of accuracy against the proportion of informed individuals when* $N = 10$.

Figure 2.8 shows how the variation in the values of accuracy produced by the runs decreases, as the proportion of individuals increases. The median increases at every proportion with the exception of 0.9. This may be due to an insufficient number of runs to suppress the effects of an outlier. On the other hand, this may be due to a group of informed individuals leaving the one uninformed individual to travel randomly. The uninformed individual can therefore have a large effect on the accuracy measure, especially if the uninformed individual were to travel in a direction at an angle of $\pi/2$ from the informed group.

For $p = 0.8$, if a group of informed individuals is to break away and leave two uninformed individuals behind, they could potentially travel in the same direction or in different or even opposite directions cancelling out the effect they would have on the accuracy measure. This is likely why for $p = 0.9$ a lower accuracy is observed than for $p = 0.8$. There are a couple of unusual points at $p = 0.7$, which may be the result of an error. Again this shows, as with Figure 2.7, that for $N = 10$ a general increase in the accuracy for an increase in the proportion of informed individuals p is seen. This is consistent with the results found by Couzin et al. (2005). Although when $N = 10$ it is possible to have a large amount of runs, variability is still a factor. In Figure 2.7 a polynomial regression line is fitted to the data.

Using polynomial regression, the best fitting line is a cubic polynomial given by $Ac = 0.53 - 0.27p + 2.2p^2 - 1.45p^3$, with accuracy Ac and the proportion of informed individuals p. A cubic polynomial is chosen over other polynomials as a cubic term is the highest power which is significant at the 5% level of a t-test. The polynomial regression line is a close fit to the mean line but smoother. It does dip at $p = 1$, but this is caused by the unusual result at $p = 0.9$. Before it can be decided if an increase in proportion results in an increase in accuracy, $N = 30$ $N = 50$ must be looked at.

Group size $N = 30$

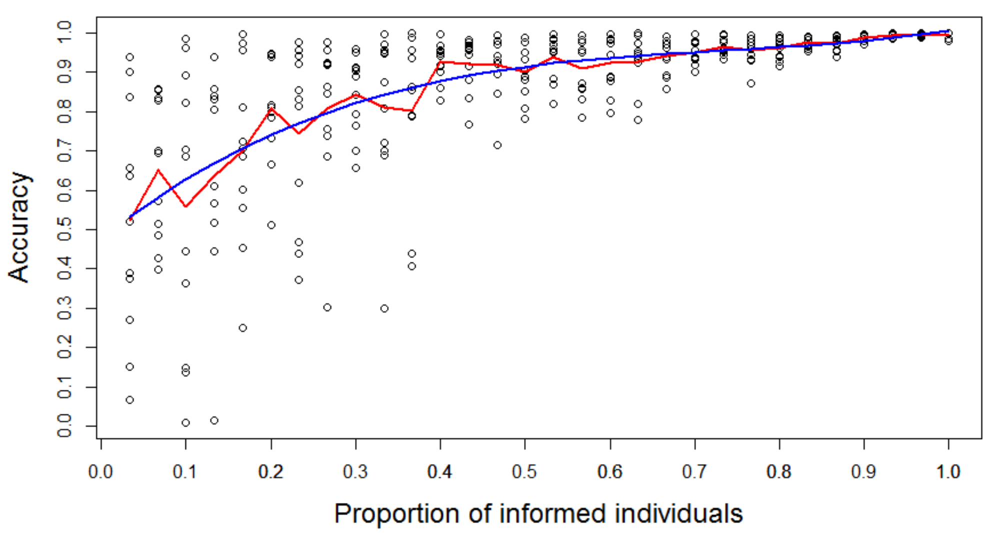

Figure 2.9: *Accuracy against the proportion of informed individuals with a polynomial regression line (blue) and a line between the means (red) for* $N = 30$.

Due to time constraints, there are only 11 runs for each p. As before, for each value of p the mean is taken and plotted along with a polynomial regression line.

As before, an increase in the proportion of informed individuals results in an increase in accuracy. Due to the vastly reduced number of runs erratic changes in mean between neighbouring values of p is observed, however, this doesn't affect the overall trend.

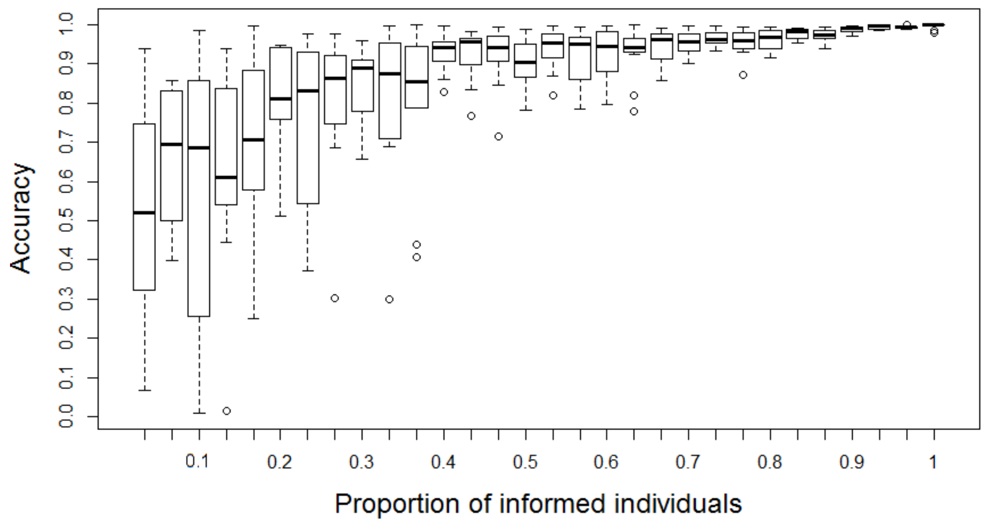

Figure 2.10: *Box plot of accuracy against the proportion of informed individuals for* $N = 30$.

With a smaller number of runs, it is harder to make conclusions on any outliers. It is possible however, to say that variation decreases for larger values of p and the median increases for larger values of p. It appears that the variation decreases faster when $N = 30$ than when $N = 10$, although it is hard to guarantee this is the case due to the reduced runs for $N = 30$ compared with $N = 10$. To help accommodate for the now considerably large random variation between different values of p, a polynomial regression line is again fitted.

The polynomial line removes a lot of the random variability whilst clearly maintaining the general trend. Again with a cubic polynomial as the best fit, $Ac = 0.48 + 1.7p - 2.1p^2 + 0.973p^3$. It is again clear that this replicates

Couzin et al.'s (2005) findings, increasing p increases accuracy, for $N = 30$.

Group size $N = 50$

Plotting the raw data with a mean line over the top results in Figure 2.11.

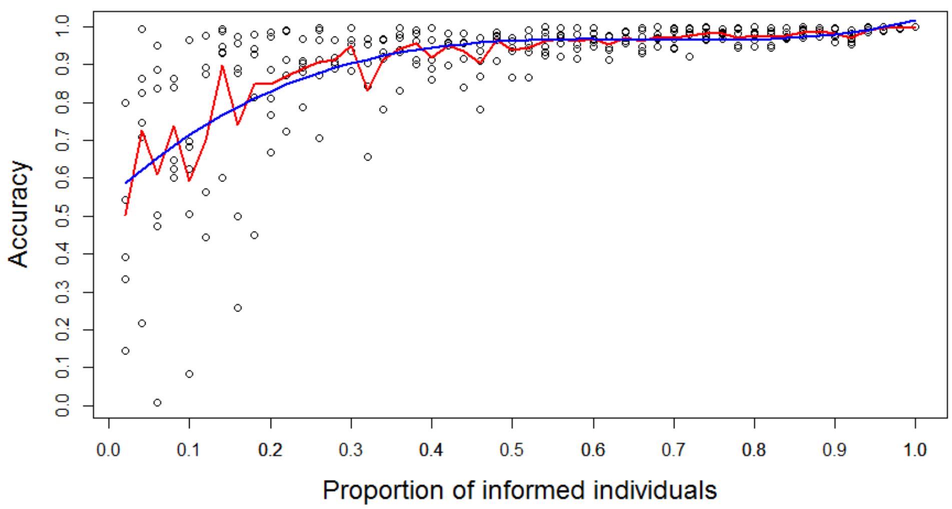

Figure 2.11: *Accuracy against the proportion of informed individuals with a polynomial regression line (blue) and a line between the means (red) for* $N = 50$.

Now that simulations for 3 different values of N have been completed, it can be concluded that for any N an increase in the proportion of individuals will result in increased accuracy. As before, for the lower values of p the random variation between each value of p is dramatic. The variation between different values of p does decrease for higher values, which is due to the decreased variability in the values of accuracy at higher values of p, shown in Figure 2.11.

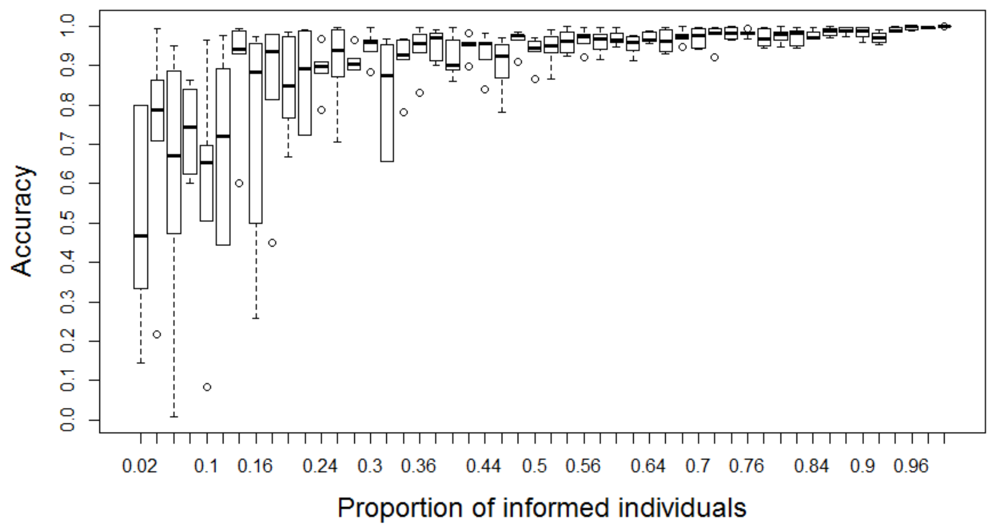

Figure 2.12: *Box plot of accuracy against the proportion of informed individuals for* $N = 50$.

Due to the limited number of runs, it is impossible to make any conclusions about outliers or individual levels of variation. It is possible, however, to comment on the general trend of the variation. As previously noted there is a decrease in variation as p increases. It can therefore be concluded that for any N, as p increases variation decreases. Variation also appears to decrease faster when $N = 50$ than $N = 30$. The same affect is visible when $N = 30$ to $N = 10$. It can be concluded that variation decreases faster for larger values of N. Using the accuracy measure, set out by Couzin et al. (2005), it would have been impossible to draw any conclusions about the variability of results for accuracy at a specific N and p. It would therefore be impossible to conclude how variation changes across N and p. Changing the accuracy measure has not lost information about the level of accuracy relative to p, as the same conclusions have been drawn.

As before, fitting a polynomial regression line removes a large amount of the random variability. Again, a cubic polynomial provides the best fit with equation $Ac = 0.55 + 1.94p - 2.96p^2 + 1.5p^3$. Due to marginally higher than expected accuracy for lower values of p, the polynomial regression line decreases slightly for values of p between 0.6 and 0.8, with an increased number of runs this problem would be reduced or even disappear.

In conclusion, changing the accuracy measure results in the same first conclusion given by Couzin et al. (2005), this being that an increase in p results in an increase in accuracy. In addition, changing the accuracy measure has resulted in new information being presented about the variability in accuracy for a particular N and p and the median values of N and p.

2.4.2 What effect does group size have on accuracy?

The second conclusion Couzin et al. (2005) drew from their results is that an increase in the group size results in an increase in accuracy for p. Plotting all 3 mean lines on the same graph results in Figure 2.13 and all 3 polynomial regression lines in Figure 2.14.

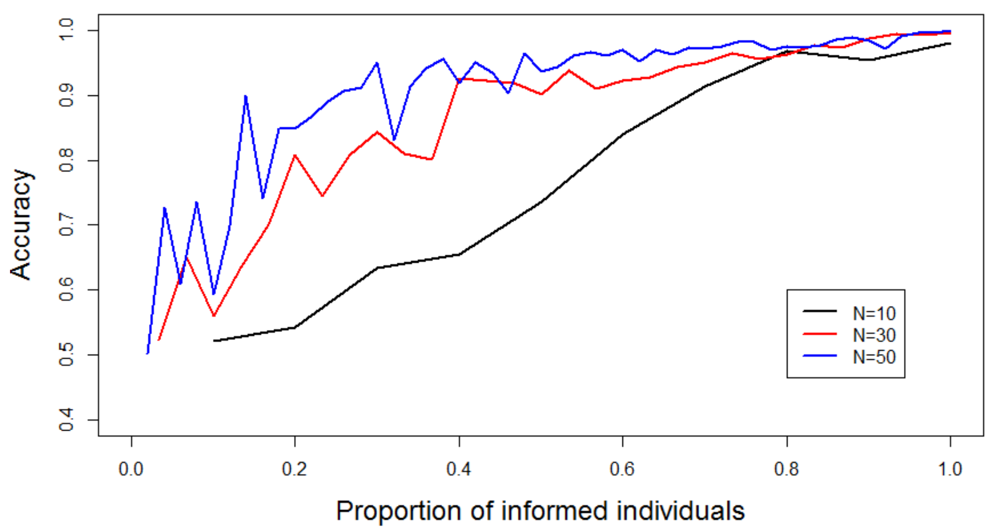

Figure 2.13: *Mean lines for* $N = 10, 30, 50$.

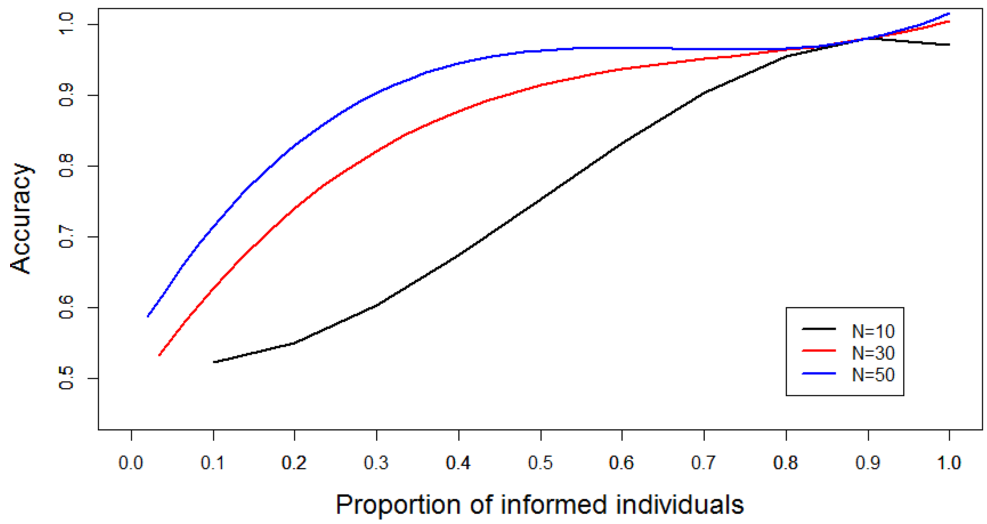

Figure 2.14: *Polynomial regression lines for* $N = 10, 30, 50$

It is immediately apparent in both figures that a larger N results in a greater degree of accuracy. This is the same result as produced by Couzin et al. (2005). It is reasonable to assume that for larger N than has been modelled here an increase in accuracy would again be seen. It is therefore possible to conclude that the results of the research paper have been replicated.

Chapter 3

An updated model incorporating topological distances

Couzin et al. (2005) assume that animals within a group use metric distance (i.e. the neighbours within a fixed distance of an individual) to determine who to follow. In Chapter 2, an arbitrary distance B is used as the metric distance. A study published by Ballerini et al. (2008) came to the conclusion that interaction does not depend upon the metric distance for starling flocks, but rather on the topological distance (i.e. interactions between a fixed number of individuals) [11]. In order to investigate the effects of a topological distance, the model is altered to account for this.

To implement a topological model, it needed to be decided how many individuals an animal would interact with. In the case of starlings, it was discovered that each bird interacts on average with six or seven neighbours [11]. Therefore, six shall be the number of animals an individual will interact with within a group (Figure 3.1). It is decided that avoidance of collision would still be an individuals highest priority, therefore A will remain metric and an individual will first check to see if any neighbours are within A. If they are, then, as previously, the individual will seek to move away from them using Equation 2.1. If they are not, then an individual will seek to align itself with its six nearest neighbours using a modified version of Equation 2.2,

$$\mathbf{d}_i(t+\Delta t) = \sum_{j=1}^{6} \frac{\mathbf{c}_j(t) - \mathbf{c}_i(t)}{|(\mathbf{c}_j(t) - \mathbf{c}_i(t))|} + \sum_{j=1}^{6} \frac{\mathbf{v}_j(t)}{|\mathbf{v}_j(t)|}. \tag{3.1}$$

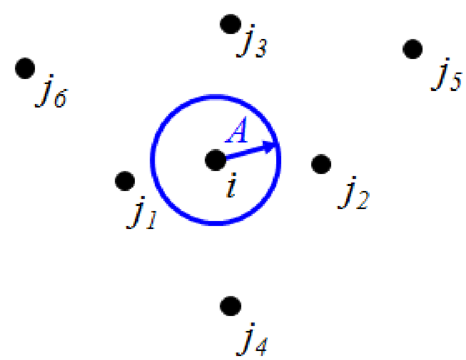

Figure 3.1: *Graphical representation of a topological distance model.*

The group size is fixed as $N = 10$ and there will be 200 runs for each p. Ballerini et al. (2008) found that a topological interaction grants significantly higher cohesion of the aggregation compared with a standard metric interaction [11]. Therefore, it would be expected that there would be a higher accuracy for each value of p in a topological model than found in metric model. Using a Matlab function to identify each individual's nearest 6 neighbours [12], a simulation is run to produce Figure 3.2 and a polynomial regression line is fitted to create Figure 3.3.

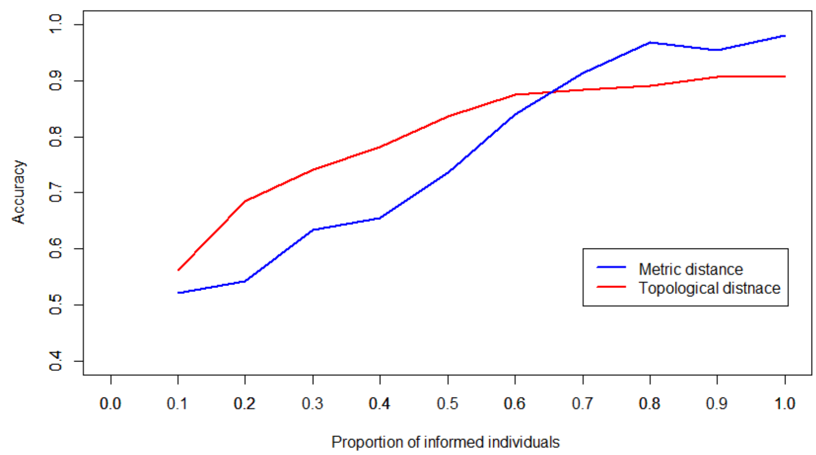

Figure 3.2: *Differences between a metric model (blue) and a topological model (red).*

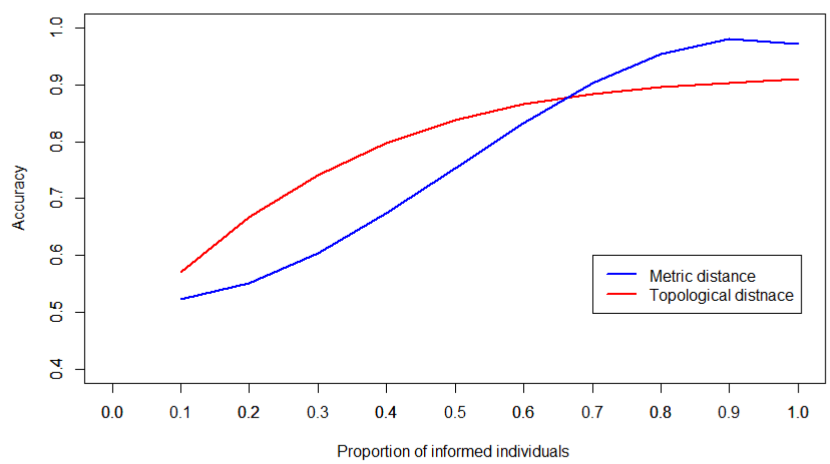

Figure 3.3: *A polynomial regression line fit for the differences between a metric model (blue) and a topological model (red).*

For p values of 0.1 to 0.6, a topological model has a higher accuracy than a metric distance. For values of p greater than 0.7 however, the accuracy is lower. This may be because in a metric distance model, it is possible for informed individuals to form a breakaway group, with the uninformed individuals allowed to drift randomly. The informed individuals traverse directly north, unaffected by the uninformed individuals. In the case of a topological model, informed individuals will still have an effect on uninformed individuals, even over large distances. This will have a negative effect upon the group direction, as the informed individuals will be influenced by the uninformed individuals random movements. Another possible explanation for the lower accuracy at high values of p is a result of the group remaining together longer. Individuals have to re-adjust more to ensure they do not collide, which results in the group drifting slightly off from the preferred direction.

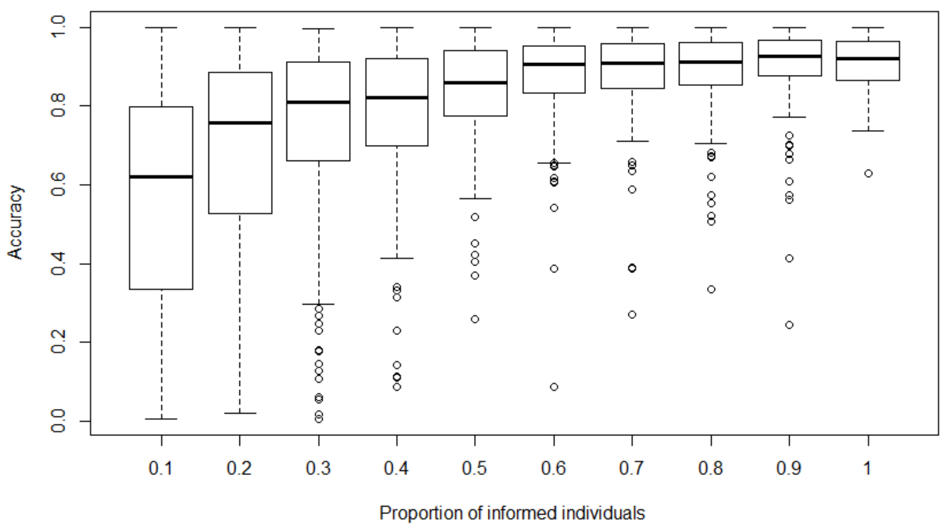

Figure 3.4: *A box plot showing the variability for a topological model.*

Figure 3.4 compared with Figure 2.2 shows the median value of accuracy increase quickly for a topological model but are lower after $p = 0.7$ compared to the metric model. The variance also decreases quicker in the topological model, although after $p = 0.7$ variability is higher in the topological model. The same explanation for the lower mean for $p > 0.6$ can be applied to a lower median and higher variance.

For Figures 3.2, 3.3 and 3.4, an individual interacts with its 6 nearest neighbours. This raises the question of how the accuracy measure would be affected if it only interacted with 4 or 8 individuals.

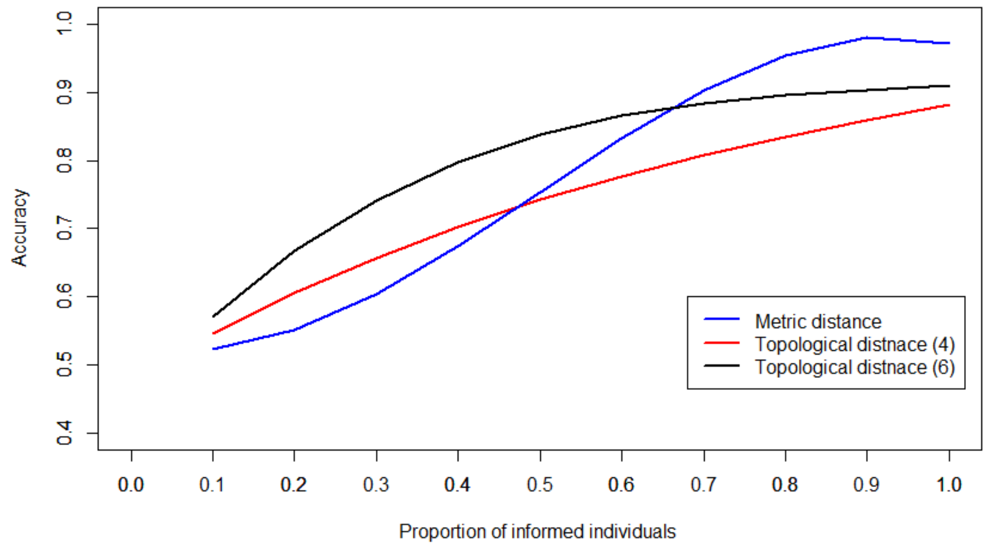

Figure 3.5: *A polynomial regression line fit for the differences between a metric model (blue), a topological model with 4 nearest neighbours (red) and a topological model with 6 nearest neighbours (black).*

When only interacting with four nearest neighbours, the group achieves a consistently lower accuracy than with six nearest neighbours. It achieves a higher accuracy for p values from 0.1 to 0.4 than a metric distance model, but lower thereafter. Curiously, for both topological models, the last p value that produces a higher accuracy than the metric model is the number of interacting individuals divided by ten. This is probably due to the fact that the group stays together longer, resulting in more uninformed individuals remaining in the group. This means more re-adjustments being made to avoid collisions. Also, for $p = 0.5$, it is possible for two groups to form, one with informed individuals travelling north and one comprised of uninformed individuals travelling in a random direction. In a metric distance model for $p = 0.5$, one group of informed individuals would still form but a group of uninformed individuals may not form. Instead, the uninformed individuals travel randomly, independent of each other and therefore their mean positions could cancel in the final accuracy measure.

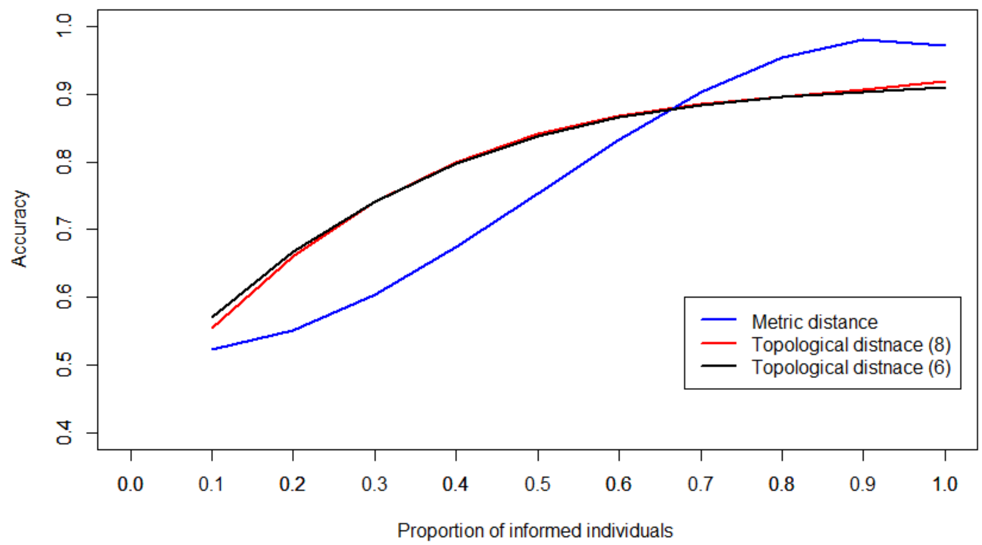

Figure 3.6: *A polynomial regression line fit for the differences between a metric model (blue), a topological model with 8 nearest neighbours (red) and a topological model with 6 nearest neighbours (black).*

There is almost no difference between a model with eight interacting neighbours than with six. Eight interacting individuals do not produce any different characteristics than six. For six and eight individuals there will still be an interaction between the informed and uninformed for any value of p, unlike with four individuals.

In a realistic example, the number of informed individuals is likely to be low. The higher accuracy increase seen for lower p values in a topological model therefore reflects the findings by Ballerini et al. (2008) that a topological interaction grants significantly higher cohesion of the aggregation compared with a standard metric one [11]. If time allowed, it would be fascinating to see the effects that a topological model would have on larger values of N, especially given that Ballerini et al. (2008) modelled large aggregations of animals [11].

A higher accuracy is seen for values of p equal to or less than 0.6 in a topological model compared with a metric distance model. A lower accuracy however, is seen for values of greater than 0.6. This is not a contradiction with the results obtained by Ballerini et al. (2008). It is in fact highlighting a flaw with the measure of accuracy. In a topological model, the individuals constantly interact with each other, giving the uninformed individuals a greater influence over the direction of the informed individuals. In a metric model, the informed individuals can leave the uninformed individuals behind. Without the influence of the uninformed individuals, the informed individuals can travel along the preferred direction easier and the uninformed individuals can travel randomly. Often, the average positions of the uninformed individuals will cancel in the accuracy measure, having so having little effect. For this reason, it is difficult to draw any conclusions on the effectiveness of a topological model.

Chapter 4

An updated model incorporating variable speeds

Another factor that may have an effect on the abilities for animals to group is speed. Couzin et al. (2005) assume all individuals have a fixed speed of $s = 1$, however individuals in a group will have a range of speeds. The model is altered to take into account this factor.

Each individual is given a speed from a normal distribution with mean=1 and a standard deviation=0.2. These are arbitrary values which do not reflect any real species. At each time step, a new speed is generated from a normal distribution to allow an animal to change speed. If the new speed is 0.05 greater or lower than its current speed, it can only increase or decrease its current speed by 0.05 respectively. To account for uncertainty in the speed change of an animal, the speed is altered by a random normal distribution with mean=0 and a standard deviation=0.001. Using the same accuracy measure as previously, and the metric distance model described in Chapter 2, Figure 4.1 is produced.

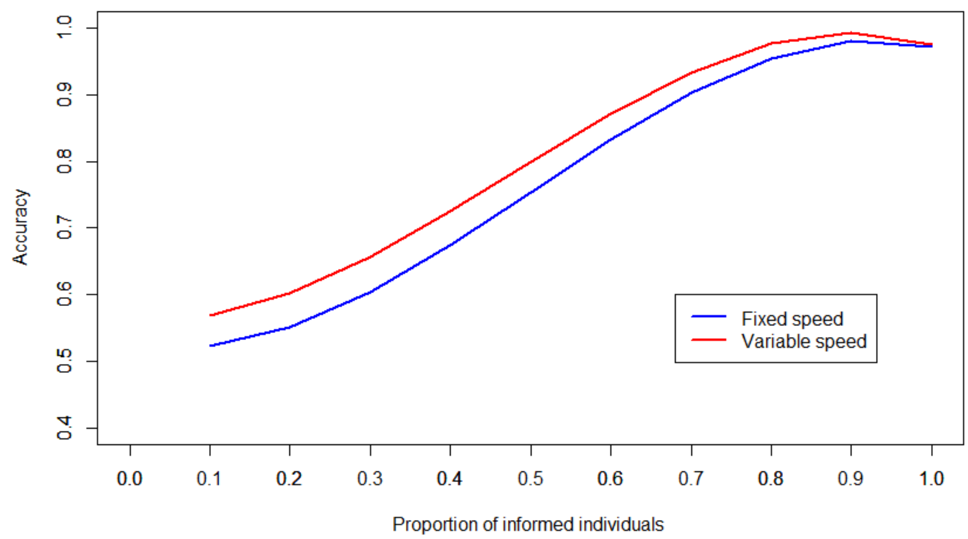

Figure 4.1: *A polynomial regression line fit for the differences between a fixed speed model (blue) and a variable speed model (red).*

A variable speed results in an increased accuracy throughout. This is largely due to the informed individuals having a greater speed and therefore breaking away from the group at a faster rate. They are therefore unaffected by the uninformed individuals and can travel in the preferred direction, whilst the average direction of travel of the uninformed individuals cancels to zero. Alternatively, the informed individuals have a smaller speed than the uninformed individuals, therefore being left behind by the group. This leaves them to travel in the preferred direction without hindrance. This model is probably unrealistic, however, as animals in a group would likely align speeds with each other. This speed adjustment can be modelled using,

$$\mathbf{s}_i(t+\Delta t) = \frac{\sum\limits_{j\neq i} \mathbf{s}_j(t)}{\sum\limits_{j\neq i} 1}. \tag{4.1}$$

Equation 4.1 sums the speeds of all neighbours within B using $\sum\limits_{j\neq i} \mathbf{s}_j(t)$ before dividing by the number of neighbours $\sum\limits_{j\neq i} 1$ to give the speed at the next time step $\mathbf{s}_i(t+\Delta t)$. Using Equation 4.1 to allign speeds whilst capping any speed change at 0.05 and having a random variation in speed of mean=0 and standard deviation=0.001 produces Figure 4.2.

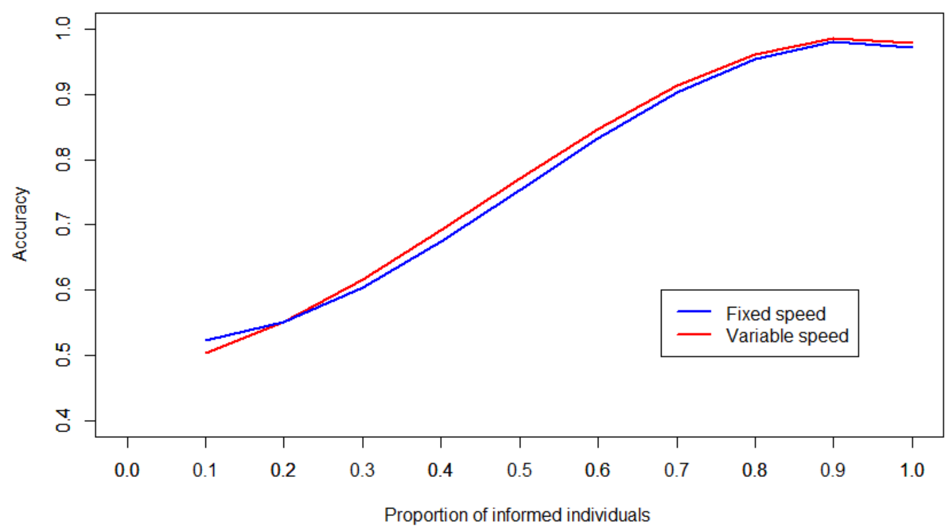

Figure 4.2: *A polynomial regression line fit for the differences between a fixed speed model (blue) and a variable speed model with alignment(red).*

When individuals align speeds, there is very little difference between accuracy levels. Therefore, it can be concluded that variable speed, in this case with mean=1 and standard deviation=0.2, makes no difference to the accuracy measure. It is therefore unnecessary in a model predicting animal behaviour in groups to have a variable speed(with mean=1 and standard deviation=0.2) and so it is perfectly acceptable to have a fixed arbitrary speed of 1.

Changing the speed has once again exposed issues in the accuracy measure, particularly when it comes to group coherence. For a variable speed that does not align, the group breaks up very quickly, but, because this allows the informed individuals to travel without influence from the uninformed individuals longer, a higher accuracy is obtained. It would be useful therefore to develop a new accuracy model that takes into account whether the uninformed individuals remain in the group or are travelling independently.

Chapter 5

Conclusions

The reference model shows that an increased number of individuals N results in a desirable accuracy for a smaller proportion of informed individuals p. For a fixed number of individuals, an increase in the proportion of informed individuals p also results in a higher accuracy. Increasing p resulting in an increase in the accuracy is intuitive, given that if more individuals are aware of the preferred direction, there will be an increase in the number of individuals travelling in the preferred direction. A larger N requiring a smaller p is counter-intuitive. It could be expected that a larger group would need the same proportion of informed individuals, or perhaps even a greater proportion. With a larger group however there are a greater number of interactions between individuals, resulting in information being shared between neighbours more quickly. This results in more individuals travelling in the preferred direction.

Redefining the accuracy measure resulted in the same conclusions given by Couzin et al. (2005) but presented new information about the variation and median values at each p. The median values behaved in the same way as the mean values, increasing for higher values of p and being higher in larger values of N. The variation, however, decreases as p increases for a fixed N and for a fixed p, the variation decreases for larger N. These conclusions would have been impossible to draw from the accuracy measure given by Couzin et al. (2005).

Due to time constraints and computer limitations it is not possible to produce results for large values of N and not possible to run each value of p for when $N = 50$ more than six times. This produced significant random variation in the results, therefore it is useful to apply polynomial regression to produce a smooth curve, from which conclusions can be drawn with greater ease. Given more time and computing power, it would have been possible to draw more conclusions, particularly when $N = 50$.

Chapter 3 implements the recommendation given by Ballerini et al. (2008), suggesting a topological model over a metric distance model. Ballerini et al. (2008) state that a topological interaction grants significantly higher cohesion of the aggregation compared with a standard metric one [11]. While this is seen for values of p less than or equal to 0.6, the topological model actually results in a lower accuracy for values greater than 0.6. Rather than this being due to a flaw in the conclusions drawn by Ballerini et al. (2008), it is due to the accuracy measure used here. In a metric model the informed individuals can break away while the uninformed individuals drift in different directions, which cancel on averaging. In a topological model however, individuals are always interacting with each other, resulting in uninformed individuals having an effect on the informed individuals, especially for large values of p. It is therefore difficult to draw any conclusions about a topological model whilst using the current measure of accuracy.

The final change to the reference model is to implement a variable speed instead of using a fixed speed. When an individual is given a random speed at each time step, accuracy is higher at each value of p. This exposed another flaw in the accuracy measure; individuals were able to break away from the group and travel uninterrupted. For the leaders, this resulted in them travelling in the preferred direction while the uninformed individuals travelled randomly. This results in a higher accuracy but they were no longer travelling as a coherent group. When individuals were allowed to align their speeds, any change in the accuracy measure is minimal and could almost certainly be put down to random variability. Therefore, for the conditions outlined in Chapter 4 and with the accuracy measure used, variable speed produced no change in the results.

In conclusion, improving the accuracy measure had benefits when dealing with the reference model, but ultimately is shown to have weaknesses when changes to the model were introduced. It may be sensible to implement another measure of accuracy, which gives weight to those individuals travelling along the preferred direction over those travelling randomly. Alternatively, a measure of group size could be used to determine whether the group remains cohesive or breaks up. For a group that breaks up, an accuracy measure would be irrelevant as the group has failed to travel as one and therefore discounted.

Acknowledgements

I would like to give thanks to Dr. Andrew Fletcher of Newcastle University for his support. Also to Bethany Andrews for use of her computer and support. Thanks to all family and friends, whom without their continued support none of this would have been possible.

Bibliography

[1] David Sumpter, Jerome Buhl, Dora Biro, Iain Couzin, Information transfer in moving animal groups. Theory Biosci 127. 177—186 (2008).

[2] Nicholas C. Makris, Purnima Ratilal, Srinivasan Jagannathan, Zheng Gong, Mark Andrews, Ioannis Bertsatos, Olav Rune God, Redwood W. Nero, J. Michael Jec, Critical Population Density Triggers Rapid Formation of Vast Oceanic Fish Shoals. SCIENCE 323. 1734—1737 (2009).

[3] Lindauer, M. Communication in swarm-bees searching for a new home. Nature 179, 63—66 (1957).

[4] safe4work. 2012. Amazing Fish Ball. [ONLINE] Available at: http://safe4work.org/tag/large-shoal/. [Accessed 24 April 2015].

[5] Swaney,W., Kendal, J., Capon, H., Brown, C., Laland, K. N. Familiarity facilitates social learning of foraging behaviour in the guppy. Anim. Behav. 62, 591—598 (2001).

[6] Franks, N. R., Pratt, S. C., Mallon, E. B., Britton, N. F., Sumpter, D. J. T. Information flow, opinion polling and collective intelligence in house-hunting social insects. Phil. Trans. R. Soc. Lond. B 357, 1567—1583 (2002).

[7] Iain D. Couzin, Jens Krause, Nigel R. Franks4, Simon A. Levin. Effective leadership and decision making in animal groups on the move. Nature 433, 513—516 (2005).

[8] behance. 2010. Waggle Dance Poster. [ONLINE] Available at: https://www.behance.net/gallery/644892/Waggle-Dance-Poster. [Accessed 24 April 2015].

[9] Couzin, I. D., Krause, J. Self-organization and collective behaviour in vertebrates. Adv. Study Behav.32, 1-75 (2003).

[10] Michael S. Mooring, Benjamin L. Hart. Animal Grouping for Protection from Parasites: Selfish Herd and Encounter-Dilution Effects. Behaviour 123,173—193 (1992).

[11] M. Ballerini, N. Cabibbo, R. Candelier, A. Cavagna, E. Cisbani, I. Giardina, V. Lecomte, A. Orlandi, G. Parisi, A. Procaccini, and M. Viale, V. Zdravkovic. Interaction ruling animal collective behavior depends on topological rather than metric distance: Evidence from a field study. PNAS 105, 1232—1237 (2008).

[12] MathWorks. 2007. NEARESTNEIGHBOUR Demonstration. [ONLINE] Available at: http://uk.mathworks.com/matlabcentral/fileexchange/12574-nearestneighbour-m. [Accessed 2 April 2015].

Appendix A

The Matlab code devised to repliacte the results of Couzin et al.(2005) and produce the figures in Chapter 2. Changes were made to this code to produce the results in Chapter 3 and 4.

```
N=50; number of individuals
A=1; Alpha-Minimum collision distance
B=6; Beta-Maximum attraction distance
F=zeros(N,2); Initial sizing of 2-Dimensional direction vector
S=1; Speed vector, randomly generated form a uniform distribution
T0=0; Initial time step
n=12500; Number of time steps
H=0.2; Size of time step
W=0.5; How inclined to follow the preferred direction of travel the object is
G=[0,1]; Preferred direction of travel
Z=2; Theta

for i=1:6

    for p=1:N
    P=(p)/N; Proportion of informed individuals
    C=unifrnd(-5,5,N,2); 2-Dimenisional position vector, randomly generated
                         from a uniform distribution
    v=unifrnd(-pi,pi,N,1); Direction angle, randomly generated from a
                           uniform distribution

        for J=1:round(P*N)
        v(J)=pi/2;
        end
```

```
F=zeros(N,2); Inital sizing of 2-Dimensional direction vector

    for I=1:N Calculation of direction vector
    F(I,:)=[cos(v(I)),sin(v(I))]; 2-Dimensional direction vector, calculated
                                      from the direction angle
    end End I

fold=v; Initial sizing of angle vector

    for I=1:n Calculating movement over time

        for J=1:N Each individual object
        Fnew=zeros(N,2); Initial sizing of 2-Dimensional new direction
                  vector

            for K=[1:J-1,J+1:N] Every object but the one we are
                  measuring from
            D = sqrt(bsxfun(@plus, sum(C(K,:).^2,2), sum(C(J,:).^2,2)') -
                  2 * (C(K,:) * C(J,:)')); distance calcualtor

                if D<=A Is object K within alpha of object J
                Fnew(J,:)=Fnew(J,:)-((C(K,:)-C(J,:))/D);Summation of
                                          directional inputs
                Fnew(J,:)=Fnew(J,:)/norm(Fnew(J,:));
                end End if D<=A

            end End K

            if Fnew(J,:)==0 Assessing if any objects did end up in Alpha

                for K=[1:J-1,J+1:N] Every object but the one we are
                  measuring from

                    if D<=B Is object K within beta of object J
                    Fnew(J,:)=Fnew(J,:)+((C(K,:)-C(J,:))/D)+F(K,:);
                  Summation of directional input
                    Fnew(J,:)=Fnew(J,:)/norm(Fnew(J,:));
```

```
        end End if (D<=B)

      end End K

    end End if F(J,:)==0

    if Fnew(J,:)==0
    Fnew(J,:)=F(J,:);
    end End if F(J,:)==0

Fnewnew(J,:)=Fnew(J,:)/norm(Fnew(J,:)); Vector normalization
end End J

for J=1:N
F(J,:)=Fnewnew(J,:);Vector normalization
end

for J=1:round(P*N) Calculating new position of leaders
f=zeros(N,1); Initial sizing of angle vector
F(J,:)=F(J,:)+W*G; Update direction vector
F(J,:)=F(J,:)/norm(F(J,:))+normrnd(0,0.01); Vector normalization

f(J)=atan2(F(J,2),F(J,1));

    if abs(f(J)-fold(J))>H*Z Calculating if angle is more than theta*time
            step away from current trajectory

      if f(J)<fold(J) If no, this calculates where the two
            trajectories are relative to each other

        if abs(f(J)-fold(J))<pi Calculating which way the old
            trajectory should be shiffted
        f(J)=fold(J)-(H*Z); Calculating the new angle
        else The Else to f(J)-fold(J)<pi
        f(J)=fold(J)+(H*Z); Calculating the new angle
        end End if f(J)-fold(J)<p

      else The else to f(J)<fold(J)
```

```
        if abs(f(J)-fold(J))<pi Calculating which way the old
                trajectory should be shiffted
        f(J)=fold(J)+(H*Z); Calculating the new angle
        else The else to f(J)-fold(J)<pi
        f(J)=fold(J)-(H*Z); Calculating the new angle
        end End if f(J)-fold(J)<pi

      end End if f(J)<fold(J)

    end End if abs(f(J)-fold(J))<H*Z

    if f(J)>pi Calculating if the angle of trajectory has excided the
                limits
    f(J)=f(J)-2*pi; Repositioning the angle of trajectory
    end End if f(J)>pi

    if f(J)<-pi Calculating if the angle of trajectory has excided the
                limits
    f(J)=f(J)+2*pi; Repositioning the angle of trajectory
    end End if f(J)>pi

F(J,:)=[cos(f(J)),sin(f(J))]; conversion back to a vector
fold(J)=f(J); saves f(J) for comparison at next pass
C(J,:)=C(J,:)+S*H*F(J,:); Update position vector
end End J

for J=round(P*N)+1:N Calculating the position of the uninformed
f(J)=atan2(F(J,2),F(J,1))+normrnd(0,0.01); Conversion back to an
                angle + Simulation of random behaviour

    if abs(f(J)-fold(J))>H*Z Calculating if angle is more than theta*time
                step away from current trajectory

      if f(J)<fold(J) If no, this calculates where the two
                trajectories are relative to each other

        if abs(f(J)-fold(J))<pi Calculating which way the old
```

```
                trajectory should be shiffted
            f(J)=fold(J)-(H*Z); Calculating the new angle
            else The Else to f(J)-fold(J)<pi
            f(J)=fold(J)+(H*Z); Calculating the new angle
            end End if f(J)-fold(J)<p

        else The else to f(J)<fold(J)

            if abs(f(J)-fold(J))<pi Calculating which way the old
                trajectory should be shiffted
            f(J)=fold(J)+(H*Z); Calculating the new angle
            else The else to f(J)-fold(J)<pi
            f(J)=fold(J)-(H*Z); Calculating the new angle
            end End if f(J)-fold(J)<pi

        end End if f(J)<fold(J)

    end End if abs(f(J)-fold(J))<H*Z

    if f(J)¿pi Calculating if the angle of trajectory has excided the
                limits
    f(J)=f(J)-2*pi; Repositioning the angle of trajectory
    end End if f(J)>pi

    if f(J)<-pi Calculating if the angle of trajectory has excided the
                limits
    f(J)=f(J)+2*pi; Repositioning the angle of trajectory
    end End if f(J)>pi

  F(J,:)=[cos(f(J)),sin(f(J))]; conversion back to a vector
  fold(J)=f(J); saves f(J) for comparison at next pass
  C(J,:)=C(J,:)+S*H*F(J,:); Update position vector
  end End J

T=T0+I*H; Calculate new time step
  if T==2490 50 time steps beofe the end

  x=mean(C); mean of all positions 50 times steps before the end
```

```
        end end if T==2490

    end End I

  y=mean(C); mean of all positions at the end
  O=y-x; difference in positions at 50 time steps apart
  O=O/norm(O); normalisation
  X=atan2(O(1,2),O(1,1)); conversion into angle
  Y=abs(pi/2-X); difference from g
  M(i,p)=abs(1-(Y/pi)); linear scale 0 to 1
  end

end
```