CHALMERS
UNIVERSITY OF TECHNOLOGY

Characterisation of collective motion

Masters thesis in Engineering Physics

Cecilia Kjellman

Department of Energy and Environment
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2015

Thesis for the Degree of Master of Science

Characterisation of collective motion

Cecilia Kjellman

Department of Energy and Environment
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2015

Characterisation of collective motion

CECILIA KJELLMAN
© CECILIA KJELLMAN, 2015
Department of Energy and Environment
Chalmers University of Technology
SE-412 96 Gothenburg
Sweden
Telephone + 46 (0)31-772 1000

Gothenburg, Sweden 2015

Abstract

Self-organisation and emergence is a widespread and fundamental aspect of biological systems. Fish schools, insect swarms, bird flocks, colonies of bacteria and human crowds are familiar examples of systems of very different levels of complexity and scale. To determine what governs interaction in the many biological systems is of importance.

This thesis mainly focuses on comparing information usage for modelling collective motion, comparing using the distance to neighbouring individuals and time to collision. The thesis begins with analysing gathered fish school data in light of recent work in the field of human crowd behaviour. The method uses a pair distribution function and a possible interaction energy to compare the two characteristics distance to neighbouring individuals and time to collision. The result differs in an interesting way from the original article on human crowd behaviour.

The later part of the thesis describes existing collective behaviour models, and model adjustments, using either distance or time to collision as the most important attribute. Suitable existing measurements are touched upon. Finally simulations using the included models are discussed but no conclusion regarding any decisive variable is made.

Key words fish schools, model, interaction energy, dynamics, collective behaviour, decision making, individual-based model, behavioural rules

CHALMERS
UNIVERSITY OF TECHNOLOGY

Date: February 20, 2015
Supervisor: Kolbjørn Tunstrøm, Assistant Professor, Energy and Environment, Chalmers University of Technology

Acknowledgements

My thanks to: my partner Anton Mårtensson for support and love, Ingrid, Magnus och Erica Kjellman for distractions, Lena Lindblad for her entusiasm, Jonatan Kilhamn for his comments, my supervisor Kolbjørn Tunstrøm for his comments and providing me with data and a starting point, Ioannis Karamouzas, Brian Skinner, Stephen J. Guy for their article regarding pedestrian interactions.

Contents

Experiments and fish school data

The experimental data used in this thesis was collected by Tunstrøm et. al [Tunstrøm *et al.* , 2013], who used the same experimental setup as in Katz et. al. [Katz *et al.* , 2011]. Details regarding the experimental setup can be found there, a brief summary follows. The schooling fish used where golden shiner, *Notemigonus crysoleucas*, which school in shallow water. The fish (average length 5 cm) were placed in a shallow tank (2.1 m x 1.2 m x 0.05 m), making it possible to assume that the school mostly school in two dimensions. A digital camera (resolution 1920 x 1080 pixels, 30 fps) filmed free-swimming schools of different sizes, for 56 minutes[Katz *et al.* , 2011][Tunstrøm *et al.* , 2013].

Tracking software was used to produce positions and velocities for individual fish. First each frame was processed to identify the geometrical center of each fish, then fish in successive frames where pair-wise matched using Sbalzarini-Koumoutsakos algorithm [Sbalzarini & Koumoutsakos, 2005]. The resulting fish trajectories were smoothed using moving window average of one third of a second. [Katz *et al.* , 2011]

The tracking software does not always succeed with identifying fishes: "On average the percentage of frames with tracking accurancy above 80% were 88% for 30 fish, 91% for 70 fish, 80% for 150 fish and 71% for 300 fishs."[Tunstrøm *et al.* , 2013]. This assumes the number of false positives are small. The software sometimes falls short when stringing observations together, the result is that one fish in the experiment tank can be seen in the tracking data as hundreds of unique individuals.

Chapter 1

Statistical mechanical approach

In this chapter we take a look at a recent article by Karamouzas et al. [Karamouzas *et al.* , 2014a]. In the article it is proposed that time to collision (from here on denoted τ) provides an appropriate and sufficient description of the interaction between individuals in human crowds, while the distance between individuals (from here on denoted r) does not.

The method used by Karamouzas et al. [Karamouzas *et al.* , 2014a] are explained below. Later the results from the article are summarised. Results from repeating the method on the fish data, and results using a variant method, will be presented as well.

Pair distribution function: g

The approach of interest relies on the pair distribution function g. It describes the ratio between a probability distribution of interest and a reference distribution for pairwise measurements. The reference distribution should have no interactions between pairs, so that the pair distribution function show pairwise interactions for the distribution of interest. We define the pair distribution function as

$$g(x) = \frac{P(x)}{P_{NI}(x)} \tag{1.1}$$

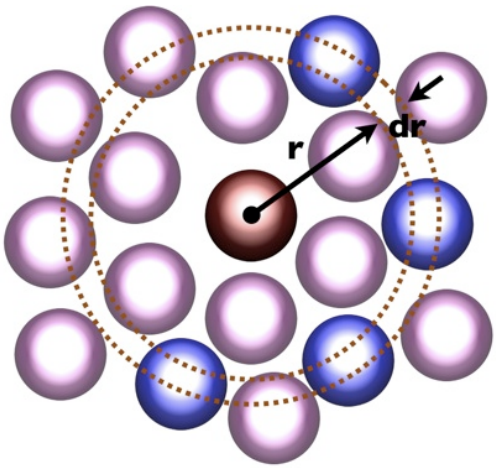
where x is an measurable variable between two individuals. $P_{NI}(x)$ is the probability density function for x if there was no interaction between individuals. $P(x)$ is the observed probability density function.

When interpreting $g(x)$, values higher than 1 indicates that the corresponding part of the probability distribution is more prevalent in the data of interest than in the reference. When $g(x)$ is lower than 1, it follows that

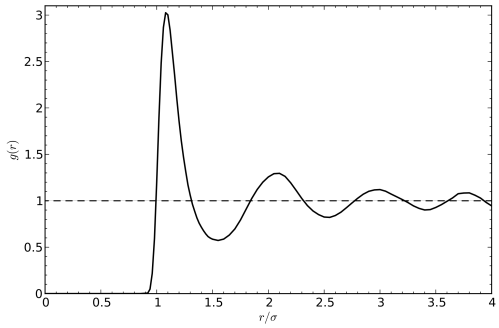
the corresponding part is less prevalent in the data of interest than in the reference. Both these cases indicates a pairwise interaction. If no interaction between individuals exists, there is no difference compared to the reference and $g(x) = 1$.

An example using r

Take a look at figure 1.1a. Using the red circle as the focal point all other circles can be used to form a pair, yielding a distance r. This is repeated using all the circles as the focal point. The data gathered are used to form the distance distribution $P(r)$.



(a) Example of how to determine the pairwise distribution. [Cri, n.d.]



(b) Example of a pair distribution function $g(r)$ with distinct highs and lows. [Sarnitskiy, n.d.]

Figure 1.1: Examples used for explaining $g(r)$.From the source for figure 1.1a: "A schematic of how the radial distribution function is calculated." From source for figure 1.1b:" Radial distribution function for Lennard-Jones fluid at $T^* = 0.71, n^* = 0.844$. Result of Monte-Carlo simulation performed by me [Grigory Sarnitskiy] for 128 particles in rhombic dodecahedron with periodic boundary conditions."

For pair-distance there is an theoretical $P_{NI}(r)$. Imagine a ring with thickness dr placed at distance r from the focal circle, as in figure 1.1a. The farther away from the focal circle the ring is placed the larger is the area covered by the circle. This means that the larger the distance r the more likely it contains other circles. The definition of the theoretical $P_{NI}(r)$ is the number

density multiplied by the area:

$$P_{NI}(r) = 2\pi r\rho dr \tag{1.2}$$

The two distributions can now form the $g(r)$ as described by (1.1). An example is shown in figure 1.1b. In the example we can deduct a few things about the structure that was used to create $P(r)$. We see that it is structured since it has defined highs and lows. For $r/\rho < 1$ we see that $g(r) = 0$, so it follows that the minimum distance between two circles is 1.

Compared to the example above, there is a noticeable difference when constructing $g(\tau)$. There is no theoretical $P_{NI}(\tau)$, but we need one to construct $g(\tau)$. Next we will look at alternative ways to define $P_{NI}(r)$ that will work for $P_{NI}(\tau)$ as well.

Alternative probability distribution $P_{NI}(x)$

The first thing we have to realise is that every characteristic we give to the reference, that is not connected to the non-interacting property, we have to give to the data as well. If we do not do this we will get a false sense of structure in the data, produced by the characteristic, not interaction.

In figure 1.2 we see that the $g(r)$ placed on the diagonal fares best, matching each type to itself. Deviations from $g(r) = 1$ in the diagonal plots is from noise, and deviations in non-diagonal plots is mainly from difference in shape. The exception to the rule above is that "Torus" and "Theoretical" are similar for small values, but since "Theoretical" is infinite and "Torus" id finite, the similarity does not last.

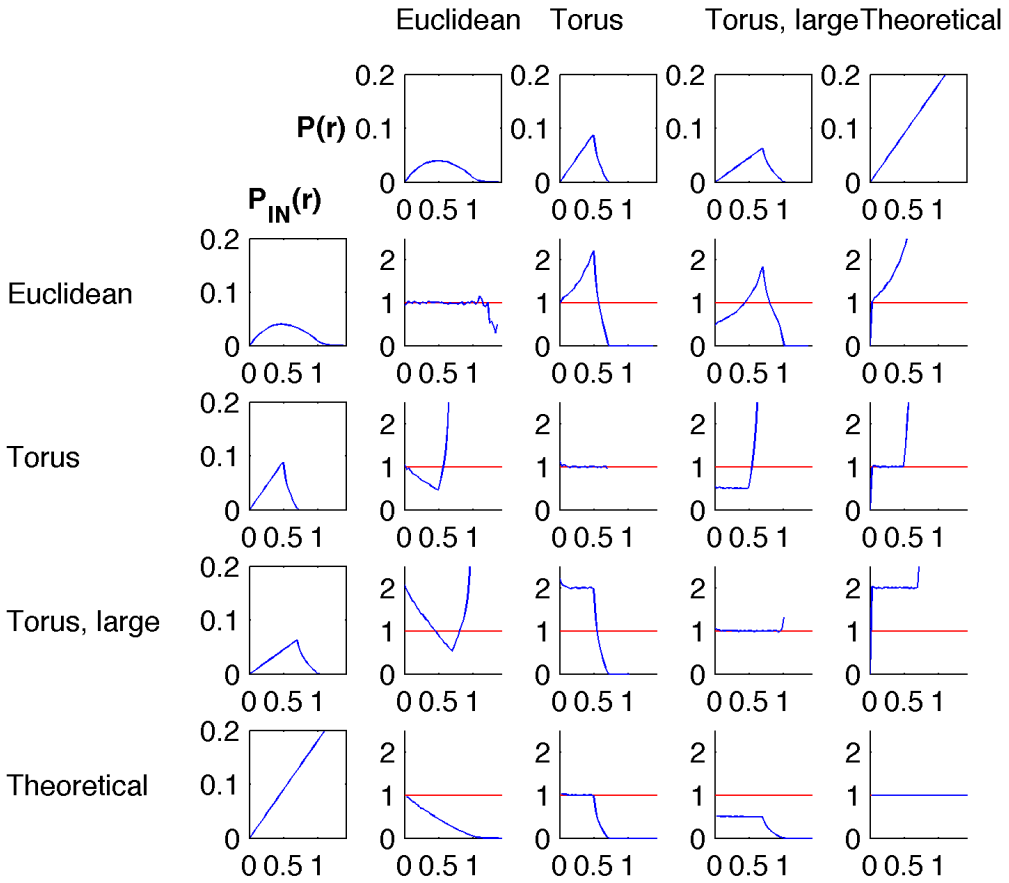


Figure 1.2: Example of how the nature of the reference must match the nature of the data. All data used is unstructured, using 900 points per area unit, uniformly placed on a square. "Euclidean" means all distances are measured as is. "Torus" means all sides are connected left-right, top-bottom, making the square a torus. All distances for the "Torus" is measured as the shortest distance following the surface of the tours. "Torus, large" means a larger torus was used. "Theoretical" uses the theoretical distribution (1.2). Since all points is unstructured, uniform random placed, all $g(r)$ should be 1. Each seperate column uses the same $P(r)$, and each seperate row uses the same $P_{NI}(r)$. The plots in the grid show a red line at $y = 1$ and in blue: the $g(r)$ from using the $P(r)$ and $P_{NI}(r)$ of the matrix element.

Simulating $P_{NI}(r)$ for distance

Instead of comparing alternative $P_{NI}(r)$ to the theoretical (1.2), they are evaluated on how well they handle 3 scenarios: uniform blob, uniform square and square grid, see figure 1.3. Of these scenarios the "Blob" is supposed to represent a unconstrained swarm, and the aim is to find a $P_{NI}(r)$ for which this blob $g(x) = 1$. Secondary we look at the "Square", which represents a swarm very constrained by walls, and that peaks and dips do not get smoothed out for "Square grid".

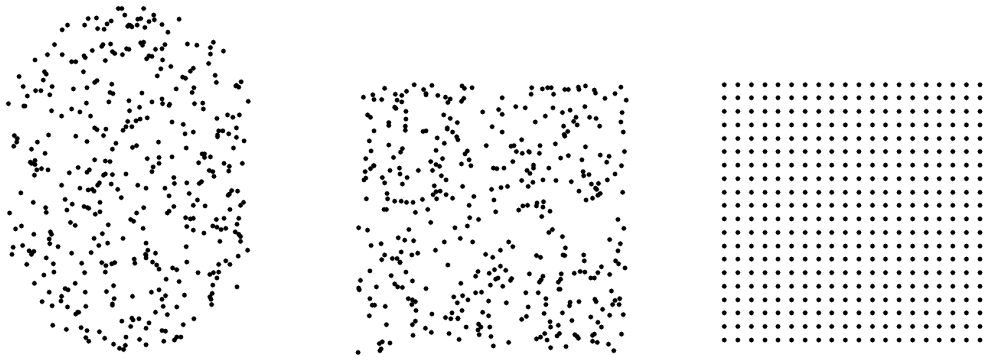


Figure 1.3: Illustrations of the different scenarios tested. Left: "Blob" points uniformly distributed over the area of two overlaping circles. The overlap is so that $^1/_3$ och the total area is unique to each circle, and $^1/_3$ is shared. This blob is supposed to represent a school not constrained by topology or walls. Since the points are placed uniformly and randomly, the desired $g(r) = 1$. Middle: "Square" points uniformly distributed over tha area of an square. Since the points are placed uniformly and randomly, the desired $g(r) = 1$. Right: "Square grid" Since the points are ordered and evenly, theoretical $g(r)$ fluctuates around 1, with clear peaks and dips.

Figure 1.4 show a graphical example from simulations made, using different $P_{NI}(r)$, numbers from repeated simulations is seen in table 1.1. The bottom three $P_{NI}(r)$ is the same as in figure 1.2. The other $P_{NI}(r)$ uses the euclidean distance, with all points uniform random inside a desired shape.

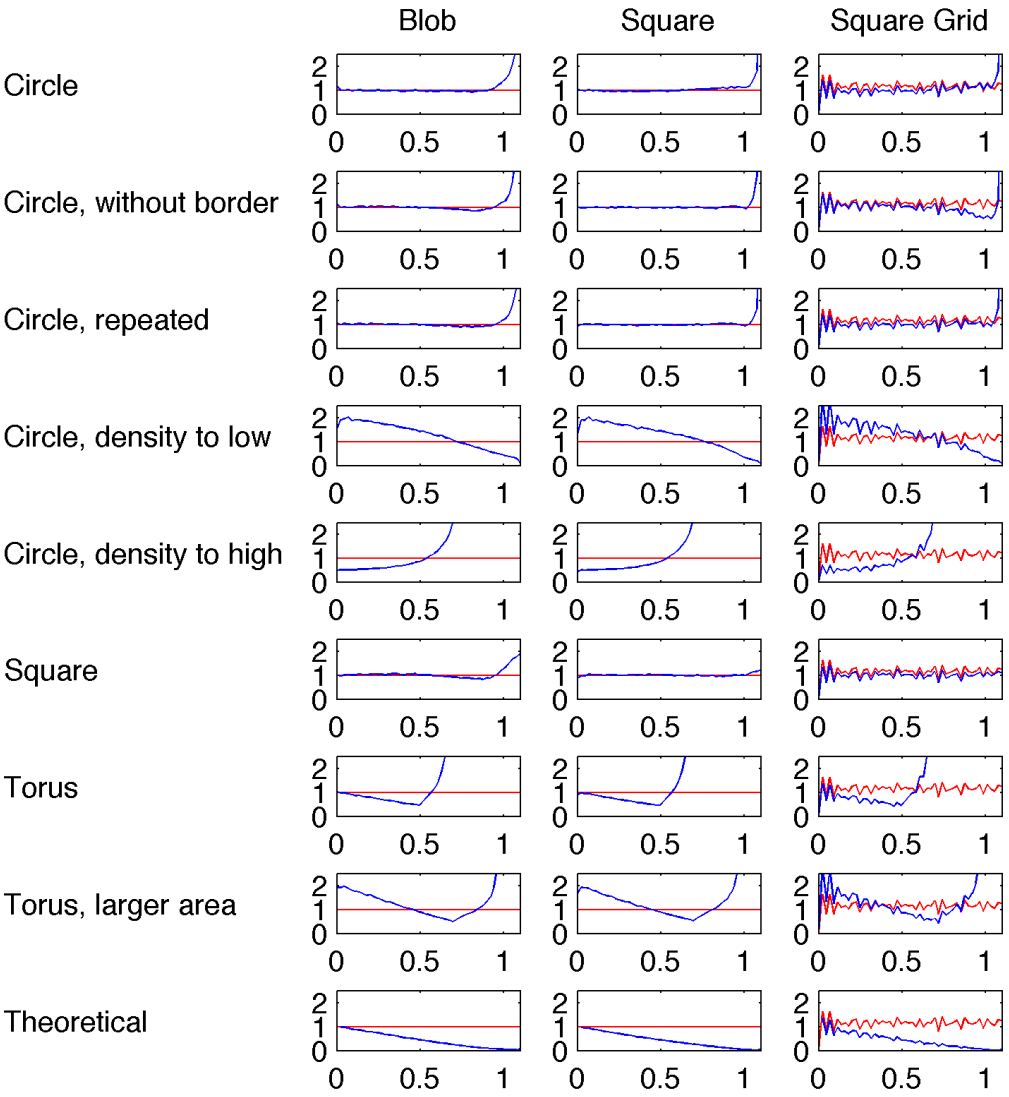


Figure 1.4: Pair distribution function $g(r)$ for the senarios in 1.3 using different $P_{NI}(r)$. If not indicated otherwise 900 points per area unit where used, giving more than $4 \cdot 10^5$ pairs and distance measurements.The pair distances were binned into 50 bins. Each row has a single $P_{NI}(r)$ and each column has a single $P(r)$. In red are the expected $g(r)$ for each $P(r)$, and in blue the resulting $g(r)$ using the $P_{NI}(r)$ of the row. Also see table 1.1.

	Blob	Square	Square Grid
Circle	0.1514 (0.342)	0.1516 (0.332)	0.1590 (0.339)
Circle, without border	0.1650 (0.459)	0.1661 (0.471)	0.1389 (0.440)
Circle, repeated	0.1496 (0.323)	0.1490 (0.323)	0.1597 (0.327)
Circle, density to low	0.5383 (0.559)	0.5371 (0.558)	0.3985 (0.475)
Circle, density to high	15.588 (15.59)	20.162 (20.16)	15.613 (15.61)
Square	0.1516 (0.140)	0.1510 (0.140)	0.1508 (0.155)
Torus	0.9347 (0.935)	0.9335 (0.934)	0.9622 (0.962)
Torus, larger area	0.3877 (90.54)	0.3874 (94.72)	0.3335 (78.61)

Table 1.1: Mean deviation of 100 runs from expected values for $g(r)$ as seen in figure 1.4. Values exclude large r to avoid the exploding tail. Values in parenthesis includes all values of r, even the tail. When testing different values for the number of points used, area and random seed, the mean deviations do vary, but still follow the general pattern above.

The conclusions from figure 1.2 stil stands. The last $P_{NI}(r)$ are infinite or loop around, and since we have made no effort to adjust the scenarios to this, "Torus", "Torus, larger area" and "Theoretical" performs poorly. Both table 1.1 and figure 1.4 confirms this.

The $P_{NI}(r)$ that have the wrong density both perform poorly. Density is a characteristic we want to keep from the data when we choose our $P_{NI}(r)$-method. As seen in table 1.1, density have a bigger effect than shape.

The two shapes tested; "Circle" and "Square" both perform well. The "Circle" fit the "Blob"-scenario better, and a few variations for "Circle" was simulated as well. Of the variations "Circle, repeated" fared best, using measurements from five circles instead of one. The problem with "Circle, repeated" is that the improvement in mean deviation is paid in computation time.

Simulating $P_{NI}(r)$ for time to collision

The simulations above were performed for time to collision as well. The scenarios can be seen in figure 1.5. When evaluating the simulations the unordered scenario will be of most importance since it is the only scenario for which we have a desired $g(\tau)$.

Figure 1.5: Illustrations of the different scenarios tested for time to collision. Left: points uniformly distributed over the area of two overlaping circles. The overlap is so that $^1/_3$ och the total area is unique to each circle, and $^1/_3$ is shared. The headings are uniformly distributed. Since the points are placed uniformly and randomly, the desired $g(\tau) = 1$. Right: Blob but with all headings towards the mass center. No expected $g(\tau)$ have been constructed.

The $P_{NI}(\tau)$ presented below all are based on the same principles as the $P_{NI}(r)$ above. All $P_{NI}(\tau)$ uses the euclidian distance on a non-looping, finite plane, all positions being uniform random inside a shape. For calculating time to collision headings and velocities are needed, for this purpose headings are assign uniformly, and all points have the same velocity.

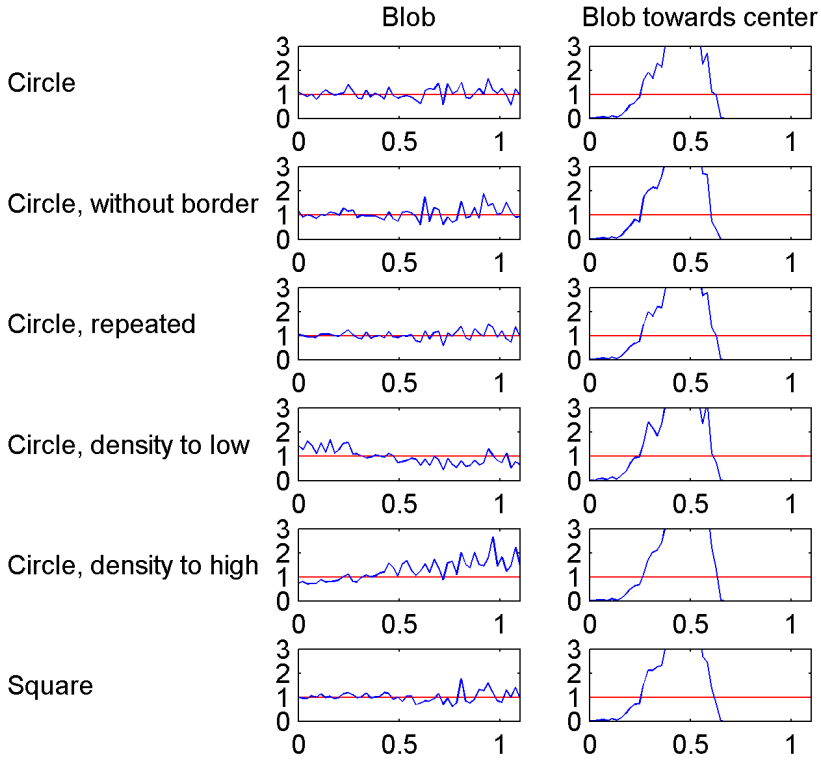


Figure 1.6: Pair distribution function $g(\tau)$ for different structures using different $P_{NI}(\tau)$. The default area are 1 area unit and 900 points where used, giving more than $4 \cdot 10^5$ pairs. The data were binned into 50 bins. Each row has a single $P_{NI}(\tau)$ and each column has a single $P(\tau)$. Also see table 1.2.

	Mean	Median	σ
Circle	0.1318 (0.1471)	0.1295 (0.1460)	0.0173 (0.0182)
Circle, without border	0.1348 (0.1535)	0.1343 (0.1517)	0.0164 (0.0191)
Circle, repeated	0.0974 (0.1052)	0.0972 (0.1054)	0.0076 (0.0079)
Circle, density to low	0.2635 (0.2717)	0.2630 (0.2704)	0.0218 (0.0224)
Circle, density to high	0.3017 (0.3341)	0.3035 (0.3339)	0.0296 (0.0323)
Square	0.1323 (0.1459)	0.1310 (0.1422)	0.0174 (0.0176)

Table 1.2: Deviation of 100 runs from expected values for $g(\tau)$ in figure 1.4, for the unordered blob scenario. Values exclude large τ to avoid a possible exploding tail. Values in parenthesis includes all values of τ, even the tail.

In figure 1.6 we see that even though the amount of points is the same as for the tests of $g(r)$, the corresponding $g(\tau)$ is noticeable noisier. The noise is a product of that not all pairs collide, resulting in less data.

The best in terms of deviation is once again "Circle, repeated", but it is resource demanding. The $P_{NI}(x)$ later referred to as "position scrambling" is "Circle" to save computation time.

Position scrambling vs time scrambling

When the $P_{NI}(x)$ above is applied to data, velocity is preserved and position and heading is randomised, henceforth referred to as "position scrambling". The article by Karamouzas et al. [Karamouzas *et al.* , 2014a] used an alternate method. Velocity, heading and position are preserved, but the time for which the individs are present are scrambled, henceforth referred to as "time scrambling". Individuals were considered non-interacting since they are not present at the same time in the scrambled data set even though they are present at the same time in the unscrambled data. The time scrambling will give rise to random scrambling shapes constrained by the experiment area, on average a square the size of the experiment area.

From the simulations above we know that density is important to preserve from the data to the reference. For position scrambling a alpha shape area for the swarms positions are calculated, and used to define the area of the used "Circle". The alpha shape is a shape that cover all points (of interest) and depending on a variable α excess space is removed so that (given the parameters) the resulting shape is the smallest possible, while still covering all points, see figure 1.7 for visual aid. The conservation of area, positions and velocities, ensures that position scrambling use the correct density.

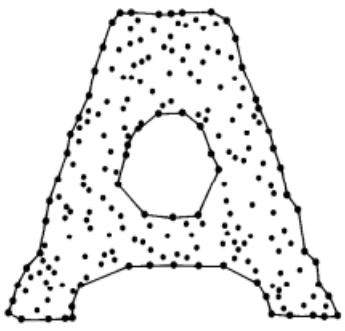


Figure 1.7: From source: "'α-shapes', which seem to capture the intuitive notions of 'fine shaoe' and 'crude shape' of point sets."[Edelsbrunner *et al.* , 2006]

Time scrambling is said to use the correct density as well. From the supplementary material to the article by Karamouzas et al. [Karamouzas *et al.* , 2014a]: "The resulting 'time-scrambled' dataset maintains the same spatially-averaged density at any given instant and the same time-averaged flow rate of pedestrians across any location in the scene as in the original dataset." It is not clear if they assume this to be a direct result of the scrambling method or if they made additional choices in how the time-scrambling was performed to ensure this. When we later make efforts to duplicate the method from the article, we always make sure that the number of present individuals is the same for data and reference at any point in time. This ensures the right density for the experiment area in total.

Are r or τ sufficient? - Using pair distribution function

The next step used by Karamouzas et al. [Karamouzas *et al.* , 2014a] is to determine if it is probably that r or τ is sufficient by analysing the behaviour of $g(r,x)$ and $g(\tau,x)$. If one of the variables are sufficient then $g(r,x) = g(r)$ or $g(\tau,x) = g(\tau)$. The secondary variables used are rate of approach and relative heading.

The methods by Karamouzas et al. [Karamouzas *et al.* , 2014a] where replicated for fish school data, with a few exceptions. It is assumed that time scrambling preserves density by preserving the number of individuals present, it is not clear in the article how this was handled in detail. The fish data was cleaned from frames where fewer than 80% of the school where found by the tracking, no such cleaning is mentioned in the article. The article does not specify what statistical test was used to test if different $g(r,x)$ and $g(\tau,x)$ is significantly different, why it is not clear if the same tests were used.

Results from Karamouzas et al.

The conclusion by Karamouzas et al. [Karamouzas *et al.* , 2014a] for the human crowd data was that $g(r)$ showed qualitative differences for different binned data, $g(r,x) \neq g(r)$, but $g(\tau)$ did not,($g(\tau,x) = g(\tau)$) thus indicating τ is a sufficient characteristic but r is not, see figure 1.8.

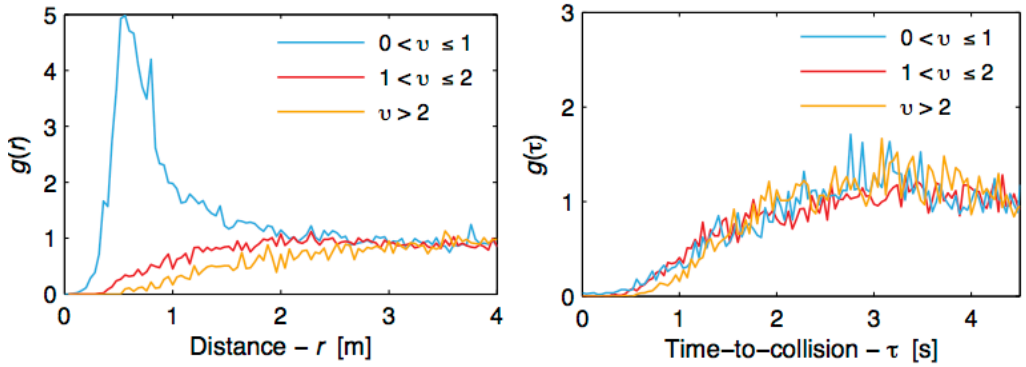
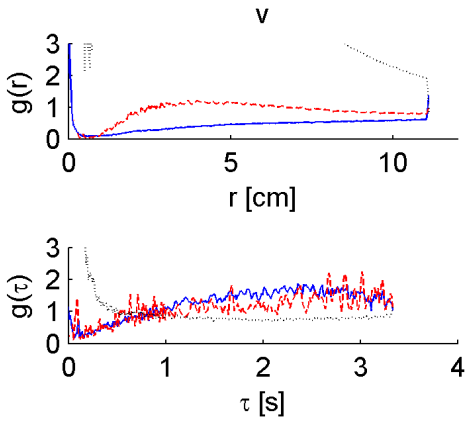


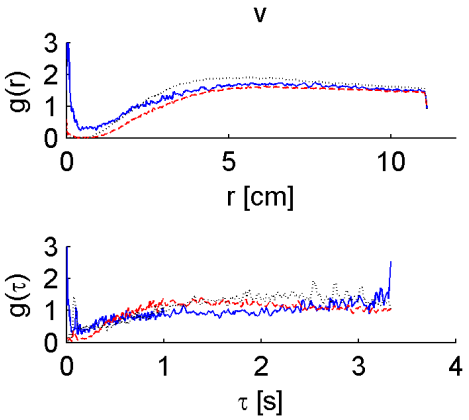
Figure 1.8: The pair distribution function g: figures from [Karamouzas *et al.* , 2014a]. Left: "The pair distribution function g as a function of inter-pedestrian separation r shows very different behavior when plotted for pedestrian pairs with different rate of approach v = dr/dt. Units of v are m/s." Right: " In contrast, when g is computed as a function of time-to-collision, , curves corresponding to different v collapse onto each other."

Fish schools

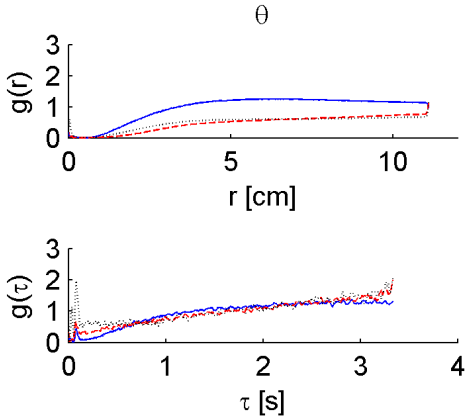
The results from the fish school data show that $g(x)$ behaves significantly different for the different scrambling methods, see figure 1.9. Some, such as 1.9c and 1.9d have the same general shape but differs in magnitude. Others, such as 1.9a and 1.9b differs both in shape and magnitude.



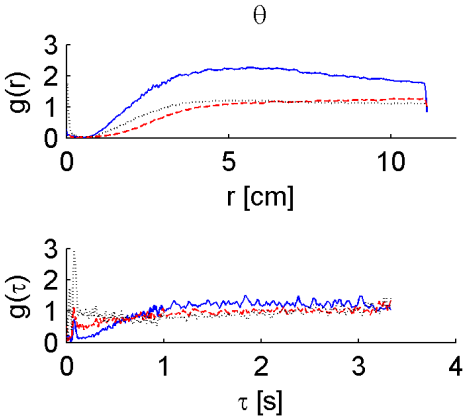
(a) Position scrambling, binned on v. Blue: $v > 10$, red: $10 \geq v > 3.3$, black: $3.3 \geq v$.



(b) Time scrambling, binned on v. Blue: $v > 10$, red: $10 \geq v > 3.3$, black: $3.3 \geq v$.



(c) Position normalisation, binned on θ. Blue: $0 \geq \theta > \pi/3$, red: $\pi/3 \geq \theta > 2\pi/3$, black: $2\pi/3 \geq \theta > \pi$.



(d) Time scrambling, binned on θ. Blue: $0 \geq \theta > \pi/3$, red: $\pi/3 \geq \theta > 2\pi/3$, black: $2\pi/3 \geq \theta > \pi$.

Figure 1.9: The pair distribution function as a function of the separation between the two fishes in the pair compared to the pair distribution function as a function of the time to collision. The pairs are binned according to the relative heading of the fishes in each pair or the rate at wich they approach each other. v is measured in cm/s, θ in radians. Data from a school of 70 fishes.

In addition to the visual information, statistical analysis were used to determine if $g(r,x) = g(r)$ or $g(\tau,x) = g(\tau)$. Wilcoxon rank sum tests were used for this, since it does not make any assumptions regarding the test-populations distribution, it test the null hypothesis that two populations

are the same, and it is already implemented in MATLAB as the function `ranksum`.

For the data set of 30 fish, using time scrambling, the findings of the original article were repeated. $g(r)$ shows a dependency of rate of approach and on relative heading, while $g(\tau)$ does not. This does not hold true for the data sets of 70 and 150 fish however, where many $g(\tau)$ are significantly different. For the $g(x)$ obtained using position scrambling, all $g(r)$ are significantly different while some $g(\tau)$ are the same. For P-values, see table 1.

Interaction energy $E(\tau)$

The article by Karamouzas et al. [Karamouzas *et al.* , 2014a] formulate an interaction energy $E(\tau)$, that is assumed to have a Boltzmann-like relation $g(\tau) \propto \exp(-E(\tau)/E_0)$. For modelling of $E(\tau)$ below the following is used [Karamouzas *et al.* , 2014a]:

$$E(\tau) = \ln(g(\tau)^{-1}) \tag{1.3}$$

This Boltzmann-approach assumes that the system is in statistical equilibrium, and that the relevant properties of the swarm are time independent.

Results from Karamouzas et al.

In the article by Karamouzas et al. [Karamouzas *et al.* , 2014a] it is found that

$$E(\tau) \propto \tau^{-2} e^{-\tau/\tau_0} \tag{1.4}$$

where τ_0 is a cut off for small $E(\tau)$.

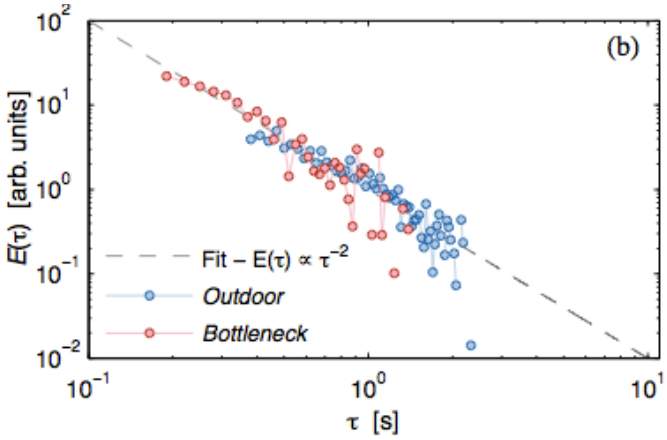
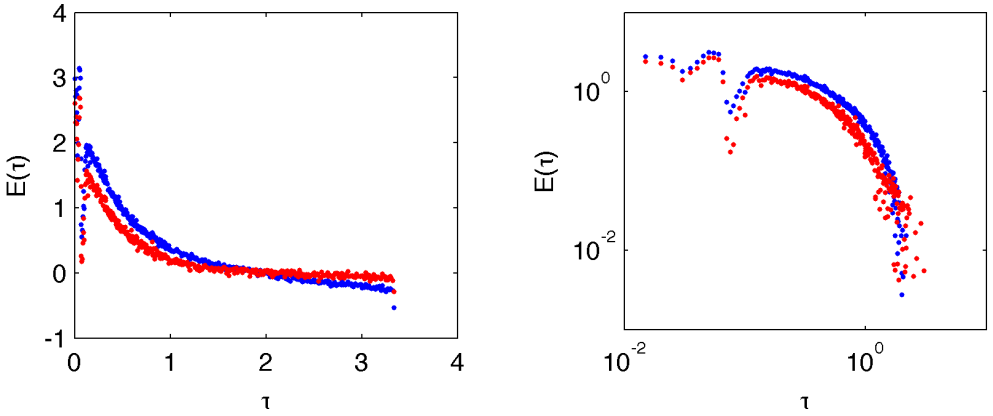


Figure 1.10: The interaction energy figure from [Karamouzas *et al.* , 2014a]. Note that this figure only serves to show the fit to the exponent 2, and is not to be interpet as the two datasets can be combined. From the article: "The overall constant k is normalized so that E(1) = 1." and "The interaction energy in both datasets is well described by a power law with exponent 2.". From the supplemental material: "[...]the interaction energy in both datasets can be well modeled with an exponent of 2 [t(174) = 0.809, P = 0.42 for the Outdoor and t(106) = 0.171, P = 0.865 for the Bottleneck]. We note that, for visual clarity, the data [...] are down-sampled".

Fish schools

The data for individual $g(\tau)$ above show different behaviour, depending on school size, scrambling method etc. After combining the data from different school sizes into one data set it becomes clear that the schools interaction energy follow the same distribution regardless of school size. $g(\tau)$ for the position scrambling and the time scrambling have similar behaviour but differ by a scale factor.

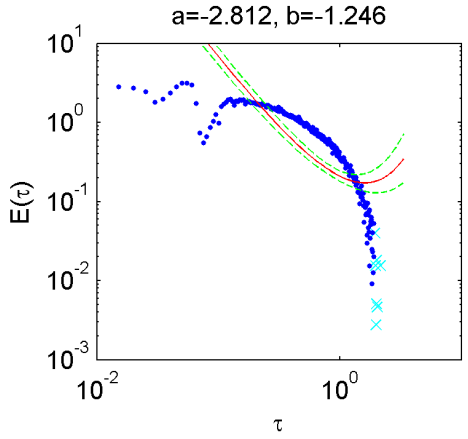


(a) Linear plot. (b) Loglog plot.

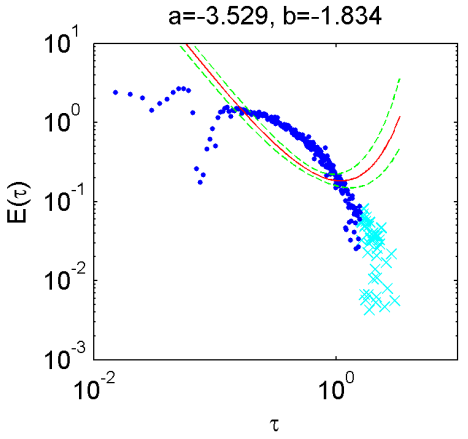
Figure 1.11: Data from time series of 30, 70 and 150 fishs, combined into one data set. Compairision of the different scrambling methods. Blue: position scrambling, red: time scrambling.

Look at figure . $E(\tau)$ becomes small for large τ, possible a result of that the influence diminishes when time to collision is large. Data points from the first τ where $E(\tau) \leq 0$ and forward where excluded from the fit. This corresponds to removing all $g(\tau) \geq 1$. This approximately corresponds the approach in the original article where all data points after a cut-off τ_0 where discarded. Furthermore the 5% most deviating data points where excluded in some fits in order to prevent outliers, such as points close to $\tau = 0$, to be overrepresented. The fit itself was performed using the MATLAB function `nlinfit`.

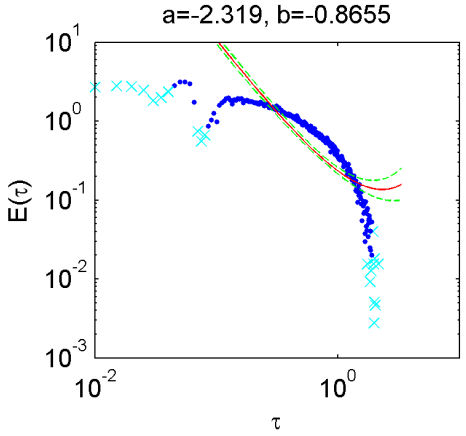
As previously mentioned, the article [Karamouzas *et al.* , 2014a] proposed the model (1.4). Using the model above for the data sets used in this thesis have proven not promising, as can seen in figure 1.12, for either scrambling method.



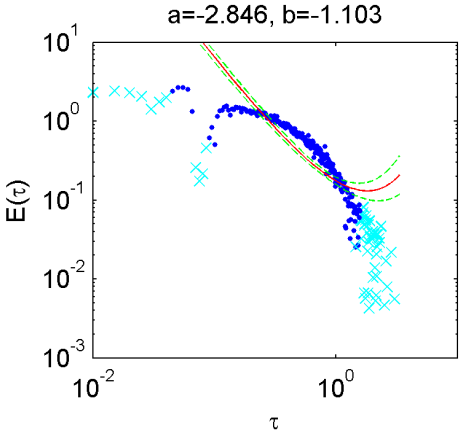
(a) Position scrambling. Values after the cut off are excluded.



(b) Position scrambling. Values after the cut off are excluded.



(c) Time scrambling. Values after the cut off are excluded as are the 5% most deviating points.



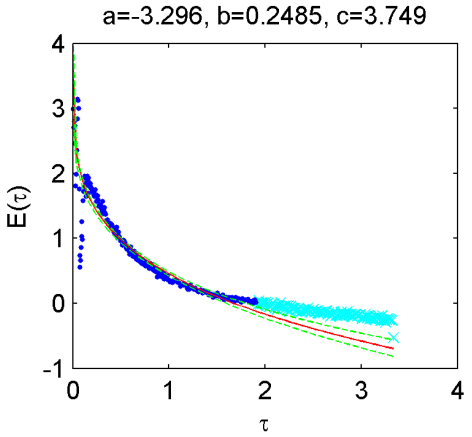
(d) Time scrambling. Values after the cut off are excluded as are the 5% most deviating points.

Figure 1.12: Data from time series of 30, 70 and 150 fishs. Fitted to (1.4). Blue: experiment data, cyan: removed data points, red: fitted model, green: confidence 95%.

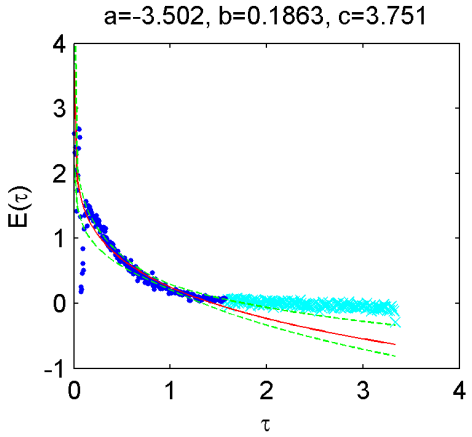
Since the model from the article proved not promising, a few other models were tried. The first:

$$E(\tau) = a\tau^b + c \tag{1.5}$$

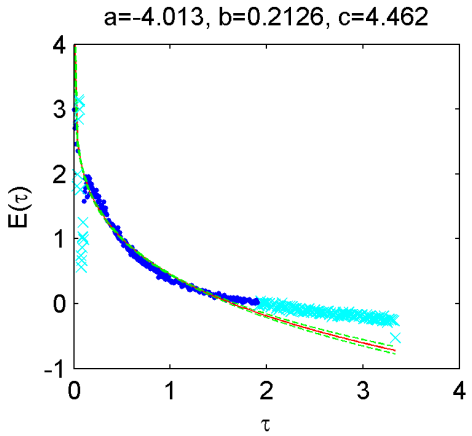
which is based of (1.4), but ignores the $e^{-\tau/\tau_0}$ term and leaves it open wether the $E(\tau)$-falloff is quadratic or not. Model (1.5) fares better than (1.4), but it is not a perfect fit.



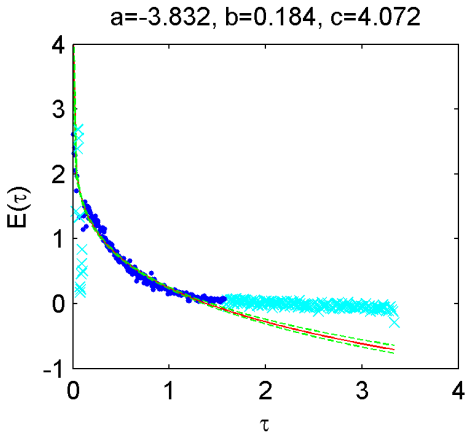
(a) Position scrambling. Values after the cut off are excluded.



(b) Position scrambling. Values after the cut off are excluded.



(c) Time scrambling. Values after the cut off are excluded as are the 5% most deviating points.



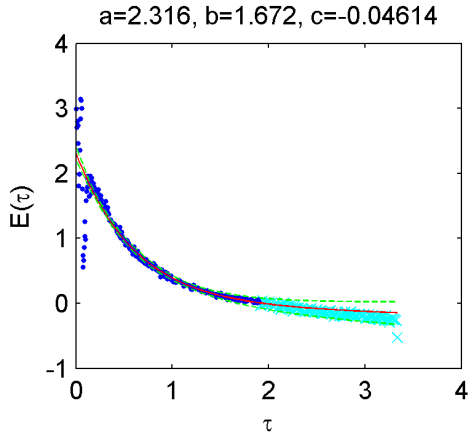
(d) Time scrambling. Values after the cut off are excluded as are the 5% most deviating points.

Figure 1.13: Data from time series of 30, 70 and 150 fishs. Fitted to (1.5). Blue: experiment data, cyan: removed data points, red: fitted model, green: confidence 95%.

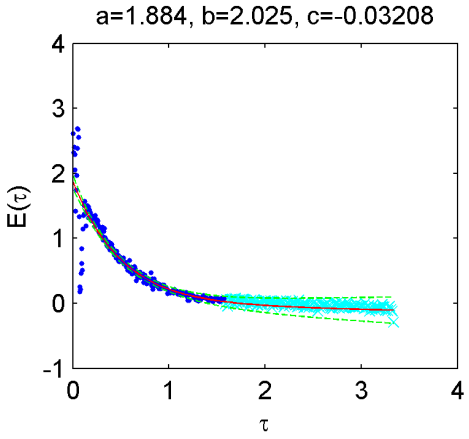
The last model presented here is defined as:

$$E(\tau) = ae^{-b\tau} + c\tau \tag{1.6}$$

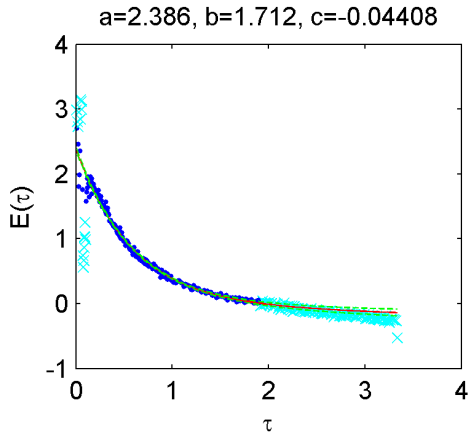
This model contain the $e^{-\tau/\tau_0}$ term but assumes a linear term as well. Model (1.6) is the best fit for the fish data, even removed data points for large values of τ match the model well. Note that there is no conclusive evidence that (1.6) is the underlaying model. The results should be interpreted as (1.6) works as a descriptive model.



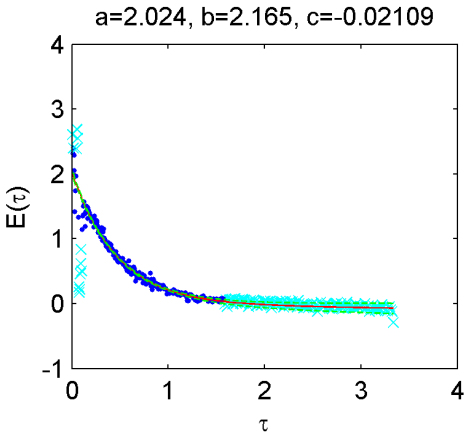
(a) Position scrambling. Values after the cut off are excluded.



(b) Position scrambling. Values after the cut off are excluded.



(c) Time scrambling. Values after the cut off are excluded as are the 5% most deviating points.



(d) Time scrambling. Values after the cut off are excluded as are the 5% most deviating points.

Figure 1.14: Data from time series of 30, 70 and 150 fishs. Fitted to (1.6). Blue: experiment data, cyan: removed data points, red: fitted model, green: confidence 95%.

Additional figures and tables for fitting to the different school sizes are included in appendix 3. Some school sizes results in a better fit than others.

Human crowds

We return to the work of Karamouzas et al. [Karamouzas *et al.* , 2014a]. The article itself does not provide us with the exact fit found but with help of the selected data [Karamouzas *et al.* , 2014b] the same model fitting is

performed. First observation is that the two datasets can not be combined into one, see figure 1.15. Furthermore the model (1.6) is a better descriptor for the Outdoor data set than (1.4).

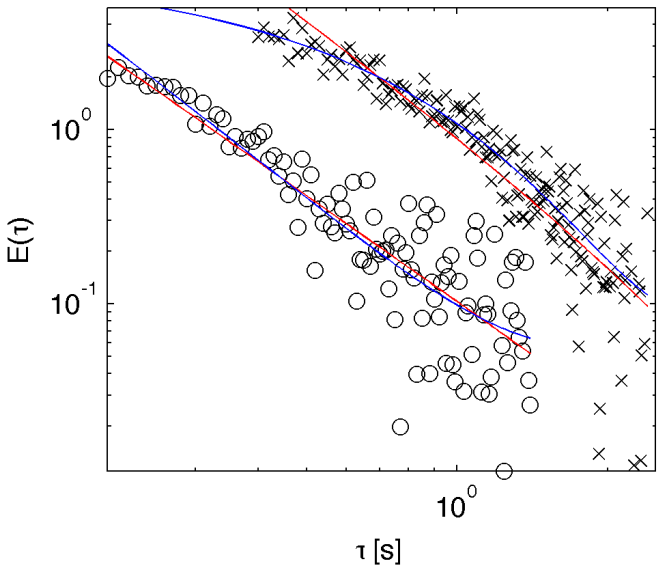
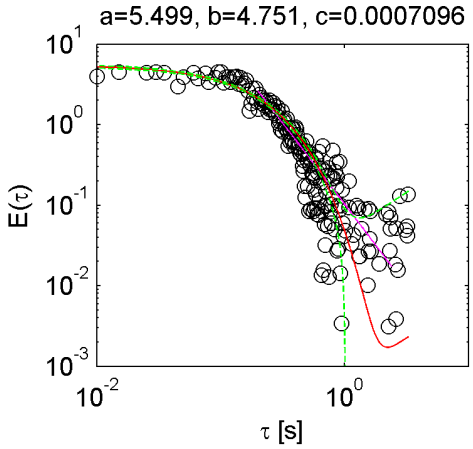


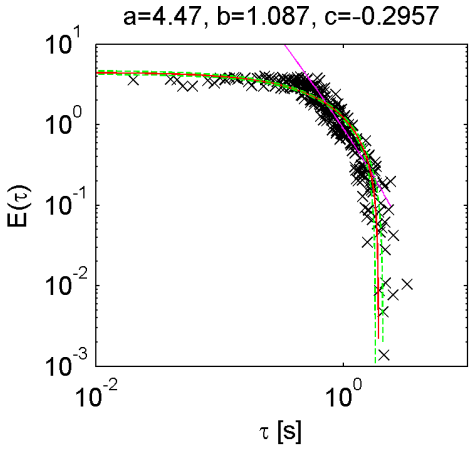
Figure 1.15: The interaction energy for the two datasets used in Karamouzas et al. [Karamouzas *et al.* , 2014a]. For the data provided [Karamouzas *et al.* , 2014b] data points have been removed for large time to collision (since tha interaction energy are too low), and for small time to collision. Circles are from the bottleneck data set, crosses are from the outdoor data set. Red is the fitted model (1.4), blue is the fitted model (1.6), both fitted using MATLABs function `nlinfit`. For the Outdoor dataset the model (1.6) fits better than the proposed model. For the Bottleneck dataset the fitted (1.6) curves upwards instead of downwards.

The data removal performed by Karamouzas et al. [Karamouzas *et al.* , 2014a] is not identical to the data removal used for the fish data. Using the values for time to collision - from article [Karamouzas *et al.* , 2014b] - we perform the same removal, used on the fish data, on the human crowd data. In figure 1.16 the interaction energy for the human crowds have been binned in the same way that the data for fish is in the previous subsection. Note that for small time to collision (1.4) is not a good fit, (1.6) is better. One also notices that the removal of points for large time to collision used on the fish data is more aggressive than the one used by Karamouzas et al. For the

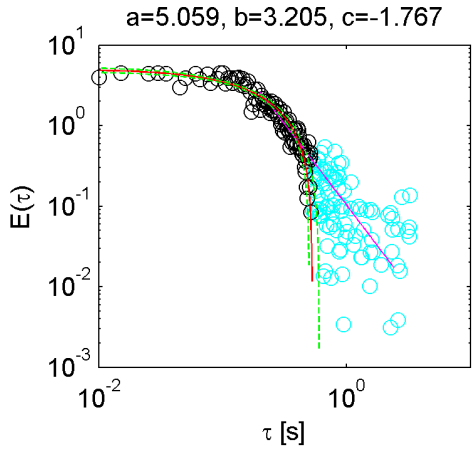
fish data the removal of the tail data produced fits that still captures the tail, see figure 1.14, this is less true for the human data. This is probably a result of the the noisier tail and the more aggressive cut of point.



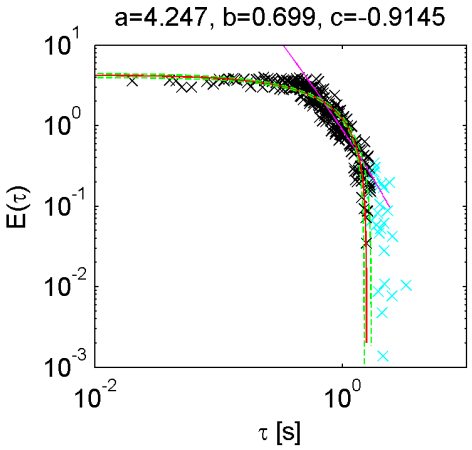
(a) Bottleneck dataset.



(b) Outdoor data set.



(c) Bottleneck dataset. Values after the cut off are excluded.



(d) Outdoor data set. Values after the cut off are excluded.

Figure 1.16: Time to collision data from [Karamouzas *et al.* , 2014b] fitted to (1.6). Both using a more allowing cut of point and the more agressive cut off used on the fish data. Also included are the fitted model (1.4), see figure 1.15. Black: experiment data, cyan: removed data points, pink: fit to model (1.4), red: fit to model (1.6), green: confidence 95%. For small time to collision (1.6) provides a much better fit. Note that in 1.16a the fitted function turns upwards, this is probably in part because the noisy nature of the data.

Chapter 2

Agent based modelling approach

This chapter uses a agent based modelling approach and compares models using different varaibles to determine behaviour, focus on distance and time to collision. The models used are all from previous work, or variations thereof.

The chapter begin with descriptions of the models used, some using mainly distance, some mainly using time to collision, to determine behaviour. The chapter continues with derived metrics used to evaluate the models, followed by results from simulations.

Models

The models below have been divided into two categories: "Distance dependent models" and "Time to collision dependent models". The two variables: r and τ is used interwoven in several models below, and for one model ("Potential gradient") we could just as easily argued it should be categorised in the other category. All results will be presented with this in mind.

Distance dependent models

Models where it is assumed that different behaviour occurs at different distances from other individuals and obstacles [Couzin *et al.* , 2002] [Grégoire *et al.* , 2003] [Mogilner *et al.* , 2003]. Put differently: the neighbouring individuals have different priority, weighted by the distance to the focal individual. The velocity of a individual at time $t + 1$ is determined by:

$$\mathbf{v}_i(t+1) = \frac{1}{n_{align}} \sum_{j=n_{align}} w_j \mathbf{v}_j(t) + \frac{1}{n_{force}} \sum_{j=n_{force}} f_{ij} + \eta_i(t) \qquad (2.1)$$

where $n_{interaction}$ is all neighbouring fishes in the zone of the interaction in question, w_j is weights for alignment, f_{ij} is the force with which fish j acts upon fish i (repulsion/attraction depending on distance), η_i is noise.

This group of models is meant to ensure that individuals do not collide (using repulsion on close distance), moves in an organised fashion (using the alignment weights), and stay together as a school (attraction on larger distances).

Behavorial zones

Couzin et al [Couzin *et al.* , 2002], uses discrete behavioural zones: a zone of repulsion, a zone of alignment, and a zone of attraction. The zones are defined by inter-fish distance, with the possible exception of a blind zone. A average for all fish in each zone is used to determine the new heading. Beyond the radius for the zone of attraction, no interaction between individuals takes place.

Note that $|\mathbf{v}_i|$ stays the same over time, this model only determines heading.

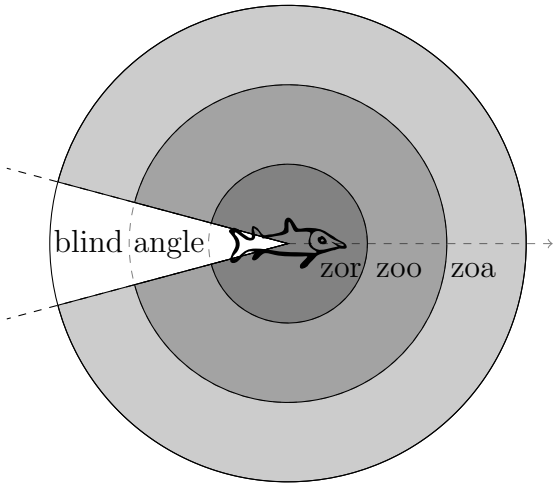


Figure 2.1: A single fish in the behavorial zones model. zor = zone of repulsion, zoo = zone of orientation, zoa = zone of attraction. Other fish present in the other zones will affect the behaviour of the focal fish depending on which zone they are in. Fish in the zone of repulsion repels the focal fish, fish in the zone of orinetaion affects the focal fish orientation, and fish in the zone of attraction attract the focal fish. The wedge behind the fish is the possible blind angle.

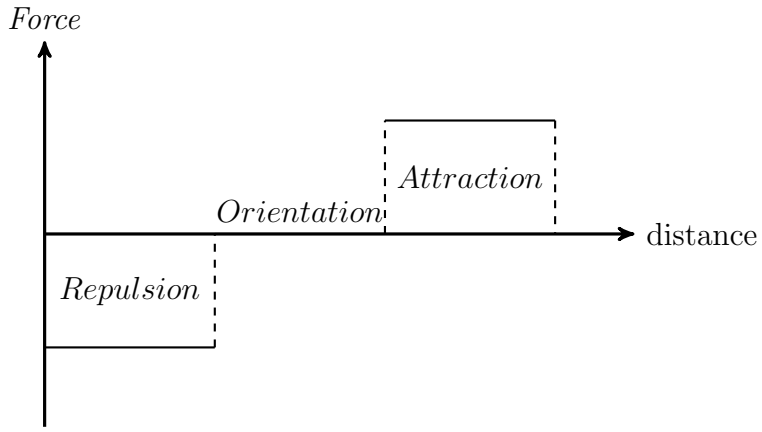


Figure 2.2: The force acting upon a fish from another fish as a function of the distance between them. When the fish are at a close distance they are in each other's zone of repulsion (and act upon each other with a repulsive force), at distances close to the mean nearest neighbour distance the fishes are in the zone of orientation (and only care about the orientation), at distances a bit beyond that the fishes are in the zone of attraction (and act upon each other with an attractive force). Beyond the zone of attration the fish ignores one another.

Behavorial force

"Behavorial force" utilises a continuous force that is a function of the inter-fish distance. In contrast to the previous model all fishes are part of n_{align} and n_{force}. At close distances fishes exert a repulsive force towards one another, at the mean nearest neighbour distance the force is zero, and at intermediate distances it exerts a attractive force that diminishes at greater distances.

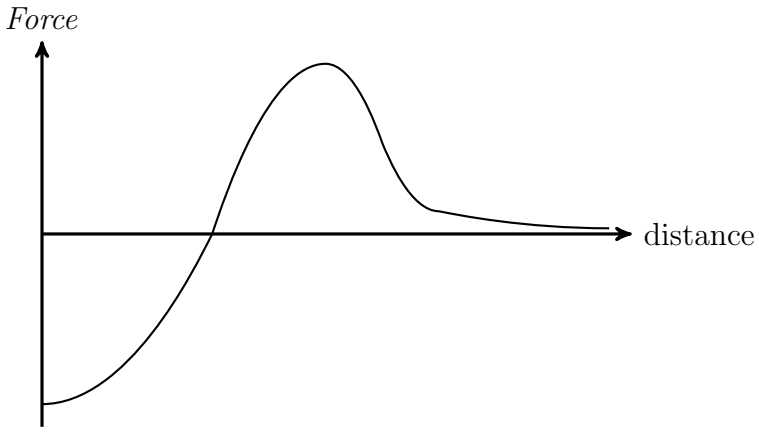


Figure 2.3: The force acting upon a fish from another fish as a function of the distance between them. When the fish is at a close distance they act upon each other with an repulsive force, at the mean nearest neighbour distance the force is at an equilibrium and the mutal force is zero, at distances beyond that the fish act upon each other with an attractive force. The force decays to zero at large distances.

Alignment factors in as a weighted mean:

$$c_{align} \cdot \sum \frac{\mathbf{v}_j}{|\mathbf{v}_j| \cdot r_{ij}} \tag{2.2}$$

where $r_{ij} = |\mathbf{x}_i - \mathbf{x}_j|$. This alignment factor prioritise close neighbours but stil uses data from far away neighbours.

In the later implementation the force function was constructed using the tracked data. Distances to the nearest neighbour where used to find a fitting function, to be used to construct the force:

$$f_{ij} = -\frac{d}{d\mathbf{x}} f_{fit}(r_{ij}) \tag{2.3}$$

Time to collision dependent models

The models below is derived from a repurposed model for human crowd movement[Moussaïd *et al.* , 2011]. All individuals are assumed to have a personal destination point which they are heading towards, avoiding collisions on the way. The way chosen by each individual is the shortest route

$$r_{min} = \min_{\theta}(r_{max}^2 + r_c(\theta)^2 - 2r_{max}r_c(\theta)\cos(\theta_0 - \theta)) \tag{2.4}$$

where r_{max} is the detection horizon, beyond this distance individuals do not consider collision avoidance relevant. θ_0 is the heading directly towards the destination point. $r_c(\theta)$ is the distance to the closest collision in the direction of θ.

The speed updates to ensure a time to collision of at least τ_{min}:

$$|\mathbf{v}_i(t+1)| = |\mathbf{v}_i(t)| + \frac{\min(v_{max}, r_c(\theta)/\tau_{min}) - |\mathbf{v}_i(t)|}{\tau_{min}} \tag{2.5}$$

where v_{max} is the maximum speed, θ is the heading for r_{min}.

Figure 2.4: Decision making for a single fish (fish$_i$). For all possible directions the fish evaluates (2.4) and chooses the angle corresponding to the lowest value. The angle straight towards the set goal provides the shortest route but causes a potential collision in an close future. An angle that does not lead to a collision soon is prefered.

In crowded cases, it is assumed that the individuals unintentionally move as a consequence of collisions.

$$f_{ij} = \max(0, r_{expected} - r_{ij}) \cdot \mathbf{n}_{ij} \tag{2.6}$$

where $r_{expected}$ is the expected minimum distance between fish i and fish j or the wall. $\mathbf{n}_{ij}$ is the unit vector from fish j to fish i.

One problem is that this requires that the individuals are heading towards a destination point, desiring the heading θ_0. In the original article [Moussaïd *et al.* , 2011] this was not a issue since the humans in the crowd where heading towards door openings, down hallways or similar. For fish schools in a featureless tank there is no such obvious destinations. Variations of this models concerns how to choose such a destination point.

Potential gradient

Using the forces defined above in 2 to construct a potential field. The heading at which the gradient of the potential field is downwards steepest is the destination heading. This circumvents the need of (2.4).

Potential field

Using the forces defined above in 2 to construct a potential field. All individuals within d_{max} are considered when calculating this. The point that has the lowest potential is set as the destination point.

Mean destination

The heading of the nearest neighbours will be used to guess where their destination point is. The mean of these neighbour's destination points is the new destination point.

Derived metrics

Metrics are needed to evaluate the models in section 2. Positions of individuals have already been mentioned in section , velocities are obtained by a two-point differentiation of the positions, accelerations are obtained similarly from the velocities. Other, less self-explanatory, metrics will be discussed below.

Basic static metrics

Mean speed $\bar{v}$ Mean speed for the swarm:

$$\bar{v} = \frac{1}{N} \sum_{i=1}^{N} |\mathbf{v}_i| \tag{2.7}$$

where $\mathbf{v}_i$ is the velocity of fish i.

Border distance $r_{CMborder}$ Distance from center of mass to closest point on the borders of the testing area.

Mean nearest neighbour distance The mean distance to the closest fish, as a measurement of the fish density

$$r_{nn} = \frac{1}{N} \sum_{i=1}^{N} |\mathbf{x}_i - \mathbf{x}_{n_i}| \tag{2.8}$$

where $\mathbf{x}_i$ is the position for fish i, and $\mathbf{x}_{n_i}$ is the position of the fish currently closest to fish i.

School states

Two metrics, O_p and O_r, can be used to describe the collective state of the school[Tunstrøm *et al.* , 2013] [Couzin *et al.* , 2002]. This is potentially useful for determining behaviour both of the tracked fish school and model simulations.

State	Description	O_p	O_r
Swarm	Individuals aggregate, are locally and globally disordered	Low	Low
Milling	Locally high degree of alignment, globally forming a milling formation	Low	High
Polar	High degree of alignment both locally and globally, the collective moves forward.	High	Low

Table 2.1: The different collective states. [Tunstrøm *et al.* , 2013]

O_p Polarisation O_p describes the degree of alignment in the swarm. Definition as in [Tunstrøm *et al.* , 2013][Couzin *et al.* , 2002]: the absolute value of the mean individual heading.

$$O_p = \frac{1}{N}\left|\sum_{i=1}^{N}\mathbf{u}_i\right|, \{0 \leq O_p \leq 1\} \tag{2.9}$$

where $\mathbf{u}_i$ is the heading (unit direction) of fish i. N is the number of fishes in the school. Higher values indicate more fish are aligned, up to 1 where all fish are aligned.

O_r Rotation O_r describes the swarms degree of rotation around its center of mass. Definition as in [Tunstrøm *et al.* , 2013][Couzin *et al.* , 2002]: the mean angular momentum.

$$O_r = \frac{1}{N}\left|\sum_{i=1}^{N}\mathbf{u}_i \times \frac{\mathbf{r}_{CMi}}{|\mathbf{r}_{CMi}|}\right|, \{0 \leq O_r \leq 1\} \tag{2.10}$$

where $\mathbf{r}_{CMi}$ is the vector pointing from the shool's center of mass towards fish i. Higher values indicate stronger rotation.

Diffusion

Diffusion tries to catch some of the dynamical properties of a school in motion. The diffusion measurements below can be fitted to a power-law dependence [Cavagna *et al.* , 2013]:

$$Dt^a \tag{2.11}$$

where a is the diffusion exponent that governs long time diffusion. $a = 1$ corresponds to brownian motion/normal diffusion, higher values indicates that individuals diffuse faster. D is the diffusion coefficient, and matters most in short time spans. Higher values indicate more mobile individuals.

Individual displacement Average mean-square displacement as a function of time. Definition as in [Cavagna *et al.* , 2013]

$$\delta_r^2(t) = \frac{1}{T-t}\frac{1}{N}\sum_{t_0=0}^{T-t-1}\sum_{i=1}^{N}(\mathbf{x}_i(t_0+t) + \mathbf{x}_i(t_0))^2 \tag{2.12}$$

where $\mathbf{x}_i(t)$ is the position for fish i at time t. Covers all time spans of duration t in the interval [0,T].

Center of mass displacement Average mean-square displacement from center of mass for the school as a function of time. Definition as in [Cavagna *et al.* , 2013]

$$\delta_{CM}^2(t) = \frac{1}{T-t}\frac{1}{N}\sum_{t_0=0}^{T-t-1}\sum_{i=1}^{N}(|\mathbf{x}_i(t_0+t) - \mathbf{x}_{CM}(t_0+t)| + |\mathbf{x}_i(t_0) - \mathbf{x}_{CM}(t_0)|)^2 \tag{2.13}$$

where $\mathbf{x}_{CM}(t)$ is the position for the schools center of mass at time t.

Nearest neighbour displacement Average mean-square displacement to the nearest neighbour as a function of time. Definition as in [Cavagna *et al.* , 2013]

$$\delta_{nn}^2(t) = \frac{1}{T-t}\frac{1}{N}\sum_{t_0=0}^{T-t-1}\sum_{i=1}^{N}(|\mathbf{x}_i(t_0+t) - \mathbf{x}_{n(t_0,i)}(t_0+t)| + |\mathbf{x}_i(t_0) - \mathbf{x}_{n(t_0,i)}(t_0)|)^2 \tag{2.14}$$

where $\mathbf{x}_{n(t_0,i)}(t)$ is the position at time t for fish is nearest neighbour from time t_0.

Neighbour reshuffling The article cited above [Cavagna *et al.* , 2013] also defines a metric for neighbour reshuffling, but it will not be used in this thesis. It is defined as mean neighbour overlap as a function of time:

$$Q_M(t) = \frac{1}{T-t}\frac{1}{N}\sum_{t_0=0}^{T-t-1}\sum_{i=1}^{N}\frac{M_i(t_0,t_0+t)}{M} \tag{2.15}$$

here M is the number of nearest neighbours considered. $M_i(t_1,t_2)$ is the number of individuals that are among the M nearest neighbours at both time t_1 and t_2, for fish i. Since it is dependent on being able to track and identify a large number of fishs over longer time spans it is not suitable for the data available.

Simulations

The models in section 2 can be adjusted by using different values for various parameters. Some of these can be derived from the raw data.

The derived values have been used for simulations starting from a point in time in the raw data then continued using the models discussed.

As mentioned in section the raw data looses track of the identity of individuals in the school, which makes a one-to-one comparison difficult for most of the tracked data. For these purposes sequences where found where all the identified fish at the start of the sequence are still present at the end of the sequence. Individuals that where not detected by the tracking algorithm at the start point will not be simulated, even though they show up in the raw data before the end. Since there was no long timespan for which this was true, the timespan of these sequences are only 4 seconds. For the same reason only tracking data for 30 fish are used.

Simulations were performed where different ratios of the school are simulated or direct from the raw data respectively. For these simulations noise was omitted.

In the tables and plots in this section metrics from section 2 will be used. Mean, median, standard deviation is given for the last timestep if applicable. For the different diffusion metrics, the a and D values are obtained by fitting using MATLABs function `fit`. Deviations from the original fish school data is then measured.

In some simulations a small number of fish in a leading position have been preserved from the experimental data. Leading position is determined as the fish who have largest number of school members positioned behind them in the first frame. This is to partly to help with the case where the school

happen upon a wall/border and avoids it in a direction, partly to test the influence of flock leaders in different models.

Distance: Behavorial zones

The size of the zone of repulsion can roughly be determined from the mean distance to nearest neighbour. The zone of attraction is estimated using the distribution of nearest neighbour distances. Also, the max turning speed from the raw data are used. For determining max turning speed, the highest values (5%) where discarded as potential tracking error. Since the model does not change the speed of individuals, the mean speed of the original data was used initally.

This still leaves the size of the zone of alignment to be estimated, likewise the field of view.

This model's lack of speed regulation is a noticeable drawback when comparing to tracked data. Tracked individuals regulate their speed to adapt to their environment which the model does not replicate. In the absence of noise and using a unit speed the school forms an organised lattice unit. In order to evaluate the model further, the speed was taken directly from the original data.

If the field of view is 360° (no blind angle) simulated individuals seek to be the center of the school, leading to a swarming behaviour. When the blind angle is increased the school becomes more and more capable of polarisation.

The model sometimes behaves noisily, with individuals getting temporarily locked in sequences of sudden turns back and forth. This is probably a result of the zone sizing, making it possible for an individual to place itself just right to oscillate between two points.

The diffusion measurements indicate the model diffuse less long term but more short term. All diffusions are sub-diffusion ($a < 1$). For more details se tables 2.2, 6.

	Mean	Median	σ	Dev. mean	Dev. median	Dev. σ
O_p	1	1	0	0.052	0.054	0.039
O_r	0.15	0.1	0.093	-0.065	-0.039	0.11
$r_{CMborder}$	2.5e+02	2.6e+02	58	-58	-44	74
r_{nn}	53	53	7.6	0.26	0.43	5
δ_r^2: a	0.76	0.78	0.12	-0.13	-0.12	0.079
δ_r^2: D	16	15	7.4	2.3	2.7	2.4
δ_{CM}^2: a	0.48	0.47	0.16	-0.013	0.011	0.26
δ_{CM}^2: D	19	20	7.3	8.5	8.8	4.3
δ_{nn}^2: a	0.5	0.5	0.21	-0.06	-0.019	0.34
δ_{nn}^2: D	13	13	8.7	9.5	7.8	9.3
Deviation	1.9e+02	1.8e+02	92			
Cum. dev.	1.1e+04	1.2e+04	4.3e+03			

Table 2.2: Data for 6 simulations of the behavorial zones model. All short term diffusions have noticible higher values (faster diffusion) than the experiment data, in contrast the long term diffusion is slower.

Distance: Behavorial force

Nearest neighbour distances from the raw data where fitted to curves using MATLABs `fit` function. The data was first split into 1000 bins and normalised. Goodness of fit where compared using root mean square error. Only lower order functions where tested in order to avoid overfitting. The best fit was a loggaussian:

$$f_{fit} = c_1 \exp(-\frac{(\ln d_{nn} - c_2)^2}{c_3^2}) \tag{2.16}$$

which gives the gradient:

$$\frac{d}{d\Delta x}(f_{fit}) = -\frac{2c_1(\ln d_{ij} - c_2)\exp(-\frac{(\ln d_{ij}-c_2)^2}{c_3^2})}{c_3^2 d_{ij}^2}\Delta x_{ij} \tag{2.17}$$

where c_{1-3} are coefficients decided by the fitting. Δx_{ij} is the distance in x for fish i to fish j.

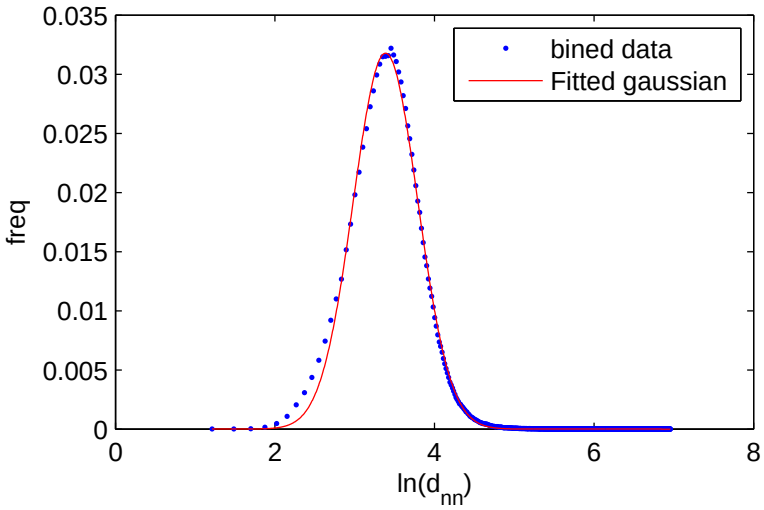


Figure 2.5: The fitted function (2.16) ploted together with the tracked data. The data was first split into 1000 bins and normalised, then fitted using MATLABs `fit` function.

Left to be estimated are field of view and the alignment constant c_{align}.

As this model also doesn't regulate the speed of the individuals, speed from the original data was used.

Compared to the other distance dependent model, this model behaves less noisily, meaning less sequences of sudden turns back and forth. This is a direct result of the smoother force distribution. In general it behaves in a way that visually similar to the raw data but deviates when choosing specific paths. A few trials with simulating only one fish and keeping the rest as is from the experimental data were performed, and the initial results for this model showed the most promise out of the models tried.

	Mean	Median	σ	Dev. mean	Dev. median	Dev. σ
O_p	1	1	0	0.052	0.054	0.039
O_r	0.068	0.047	0.058	-0.15	-0.14	0.09
$r_{CMborder}$	2.2e+02	2e+02	71	-95	-92	75
r_{nn}	57	57	9.6	3.7	4.4	6.2
δ_r^2: a	0.91	0.91	0.04	0.026	0.025	0.027
δ_r^2: D	12	12	7.3	-1.6	-1.8	1.3
δ_{CM}^2: a	0.39	0.37	0.13	-0.1	-0.094	0.18
δ_{CM}^2: D	11	11	5.6	0.64	-0.2	5.1
δ_{nn}^2: a	0.36	0.29	0.26	-0.2	-0.28	0.27
δ_{nn}^2: D	11	8	11	7.9	4.4	10
Deviation	1.7e+02	1.6e+02	98			
Cum. dev.	7.9e+03	7.7e+03	4.3e+03			

Table 2.3: Data for 6 simulations of the behavorial force model. This model have the lowest cummulative deviation from the experiment data.

Time to collision: Potential gradient

In a non-crowded situation the individuals on the border (and in front) of the school has nothing stopping them from swimming at max speed. Soon there is nothing stopping the fish in the centre of the school to swim at max speed. Unless the school comes too close to one of the borders the school will behave as a grid structure moving at said max speed, on other circumstances a fish can find a local potential minima and get stuck.

The short term individual displacement is high is this model, and the long term nearest neighbour displacement super-diffus.

	Mean	Median	σ	Dev. mean	Dev. median	Dev. σ
O_p	1	1	0	0.052	0.054	0.039
O_r	0.21	0.18	0.12	-0.011	-0.021	0.18
$r_{CMborder}$	95	91	40	-2.2e+02	-2e+02	1.1e+02
r_{nn}	38	43	14	-15	-13	12
$\bar{v}$	3.8	4.1	1.2	1	1.3	1
δ_r^2: a	0.78	0.76	0.093	-0.1	-0.13	0.062
δ_r^2: D	71	76	23	57	56	20
δ_{CM}^2: a	0.47	0.46	0.16	-0.017	0.0084	0.21
δ_{CM}^2: D	32	31	7.2	22	18	12
δ_{nn}^2: a	0.91	0.87	0.67	0.35	0.35	0.8
δ_{nn}^2: D	6.5	4.7	6.8	3	0.76	8.3
Deviation	3.7e+02	4e+02	96			
Cum. dev.	2.4e+04	2.4e+04	5.1e+03			

Table 2.4: Data for 6 simulations of the potential gradient model. The distance to the border is noticible smaller than the experiment data. This is potenitally because the border was represented with the same forces as fellow fish.

Time to collision: Potential field

For this model the max speed and the minimum time to collision τ can be taken from the raw data, again omitting extreme values as potential tracking error (5%). As before the maximum turning speed from the experimental data is used.

Left to be estimated is the collision avoidance horizon r_{max}, and the field of view.

This model makes individuals try to intermix more than in the original data, and if a blind zone is used the school easily separates. The intermixing is possibly a result of the destination point being placed a far bit away in the school, while the true destination point might just be whatever point makes the fish roughly keep its place relative to the school. As with the previous model: a fish can find a local potential minima and get stuck.

	Mean	Median	σ	Dev. mean	Dev. median	Dev. σ
O_p	1	1	0	0.054	0.07	0.043
O_r	0.1	0.057	0.099	-0.11	-0.11	0.16
$r_{CMborder}$	2.5e+02	2.5e+02	76	-64	-64	66
r_{nn}	37	37	12	-16	-16	11
$\bar{v}$	3.1	3.2	0.51	0.29	0.41	1.1
δ_r^2: a	0.34	0.36	0.24	-0.54	-0.49	0.25
δ_r^2: D	9.1e+02	48	2.1e+03	9e+02	29	2.1e+03
δ_{CM}^2: a	0.32	0.33	0.24	-0.17	-0.23	0.3
δ_{CM}^2: D	9.1e+02	48	2.1e+03	9e+02	38	2.1e+03
δ_{nn}^2: a	0.44	0.45	0.31	-0.12	-0.13	0.45
δ_{nn}^2: D	8.1e+02	23	1.9e+03	8.1e+02	22	1.9e+03
Deviation	3.7e+02	3.9e+02	1.8e+02			
Cum. dev.	2.1e+04	2.2e+04	7.1e+03			

Table 2.5: Data for 6 simulations of the potential field model. All short term diffusions have noticible higher values (faster diffusion) than the experiment data, in contrast the long term diffusion is slower.

Time to collision: Mean destination

For this model the following variables have to be estimated: number of neighbours to consider, and how many time steps ahead the destination point is estimated.

If the field of view is less than 360° (has a blind angle) the school's frontmost fish have a probability to loose track of the school and swim away on their own. Other works have deduced that information transfer in groups of fish flow from front to back and not from back to front [Katz *et al.* , 2011]. This discredits this model that need the information from individuals behind them to stay on course.

	Mean	Median	σ	Dev. mean	Dev. median	Dev. σ
O_p	1	1	0	0.052	0.054	0.039
O_r	0.19	0.23	0.13	-0.032	0.022	0.16
$r_{CMborder}$	3.4e+02	3.3e+02	22	25	23	1.2e+02
r_{nn}	56	57	5	3.3	0.56	6.2
$\bar{v}$	7.5	7.6	0.5	4.7	4.4	1.3
δ_r^2: a	0.71	0.71	0.15	-0.18	-0.17	0.14
δ_r^2: D	1.2e+02	1e+02	58	1e+02	90	62
δ_{CM}^2: a	0.33	0.32	0.083	-0.16	-0.23	0.18
δ_{CM}^2: D	83	87	24	73	75	26
δ_{nn}^2: a	0.65	0.55	0.37	0.092	0.098	0.23
δ_{nn}^2: D	21	14	27	18	8.6	27
Deviation	3.9e+02	3.6e+02	1.1e+02			
Cum. dev.	2.3e+04	2.3e+04	3.1e+03			

Table 2.6: Data for 6 simulations of the mean destination model. This is the only model were fish are farther from the border than in the experiment data.

Distance vs. time to collision

These simulations are not conclusive in deciding which of the models that is superior to the rest, and all show weaknesses when compared to gathered experimental data.

It is apparent that the lack of speed regulation in the two distance dependent models are a disadvantage, and the fix of keeping the original speeds might be an unfair advantage in the comparisons. The use of time to collision to regulate speed is a noticeable improvement, but the lack of a tested way to decide a destination point lowers the performance of the time to collision models.

The interwoven nature of time to collision and distance in the models compared makes it hard to draw any conclusion regarding preferred variables from the simulation data.

Chapter 3

Discussion

Even though this report did not contain the definitive answer to what is the most prominent factor for decision making on individual level of an swarm it does lean towards time to collision as the better one, compared to distance. Below are some thoughts on areas in this report that needs further investigation and what could have been done differently.

General

All models and the analysis of the proposed interaction energy have here approximated the shape of a fish to be a circle. This was done partly to cut computation times, but also because there was no tested way to determine how much space the fish actually occupy. A considered approximation was to construct an oblong oval or a rounded rectangle, with the elongated axel in the direction of the fish velocity. This caused a bit of problem with the headings being noisily, and the oval approach was dropped before a solution was found. A longer study could look into using ovals instead.

Statistical mechanical approach

The approach adapted from Karamouzas et. al. [Karamouzas *et al.* , 2014a], yielded interesting results but leaves several questions unanswered. The choice of normalisation method affect the resulting $g(x)$ noticeably, and the best method for $g(\tau)$ might not even be among those tested. The used $g(r)$ show the expected artefact with peaking at higher values. They also show peaking at low values of τ, for several fish sizes, all $g(x)$ and normalisation methods. It is unclear wether it is an artefact of the method used, the experimental setup or a real feature.

The area normalisation method that faired best was not the one used. This was solely because of time constraints and the method mentioned as "Circle, repeated" would, in hindsight, have been preferable.

It was tested separately on a small scale to add a weight to the data of distances and times to collision, being effectively the number of fish in the pair able to see the other fish in the pair. The weight modifications did not initially yield interesting differences to not using weights, it only added noise by removing data, which is why the weighted approach was not pursued and is omitted from this report. If more time was given this could have been looked into further.

The Boltsmann-like relation $g(\tau) \propto \exp(-E(\tau)/E_0)$ and the formulation of an interaction energy does makes some assumptions not touched upon in this report. It assumes that the system is in (or near) statistical equilibrium, so that the relation falls out as a result of entropy maximisation. In the original article this was motivated by that relevant characteristics (density and mean speed) were essentially time independent [Karamouzas *et al.* , 2014a]. No further attempts were made to justify this as part of this thesis.

The model (1.6) fits both the fish data and the human data better than the proposed (1.4), especially if one does not remove data points. This opens up for further investigation of what the appropriate range and model is.

Agent based modelling approach

A draw-back to the model-based approach is that many different microscopic rules can give the same macroscopic behaviour, and there is no guarantee that a model that matches some observed characteristics well will match other characteristics or the underlying rules.

The comparison of distance versus time to collision using a model-based approach has several problems. The distance dependent models used lack speed regulation, making them inferior. When using the speeds from the original data they instead get an noticeable advantage. All time to collision dependent models used do use distances in some way to determine a destination point. Time to collision is a product of distance, speed, and heading for the individual fish and all neighbours and obstacles, making it costly to compute. An advantage of time to collision is that it makes it possible to prioritise information and social cues of neighbouring individuals based on wether collision is a possibility.

If possible, a model using only time to collision, also when deciding a destination point, or as a substitute for distance in a behavioural force model, would be an interesting next step.

A few tests where only a one of the individuals in a school was simulated were performed. Although interesting it is difficult to draw big conclusions from such trials due to the risk of overfitting etc. Other research groups have used a neural networks approach to find a model according to which (many) single fish behave [Herbert-Read *et al.* , 2011], and it gives an interesting angle to the modelling problem.

Any model that makes itself too complicated to compute on an individual level makes itself unlikely to be executed in real time by the individuals in a swarm, thus discrediting itself. Time to collision is mentioned above as computationally harder than distance. In the field of game and movie graphics there have been advances in crowd simulations, producing collision avoidance heuristics and algorithms that are faster [Guy *et al.* , 2009]. It is possible that different good-enough heuristics are closer to the underlying rules than a hard-to-compute but exact algorithm.

The addition of leader fish in the simulations were interesting and there was some comparisons made between keeping a small number of front placed fish and keeping fish otherwise placed in the school. This was cut from the report due to time but could be interesting to pursue on itself.

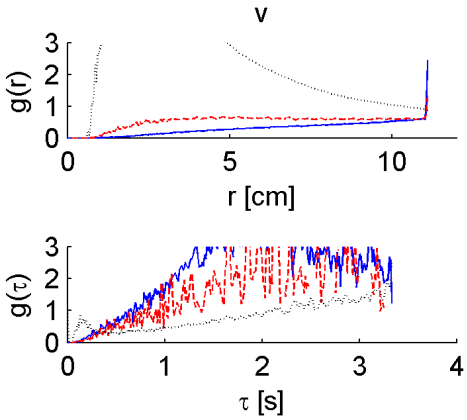
Additional tables and figures

Statistical mechanical approach

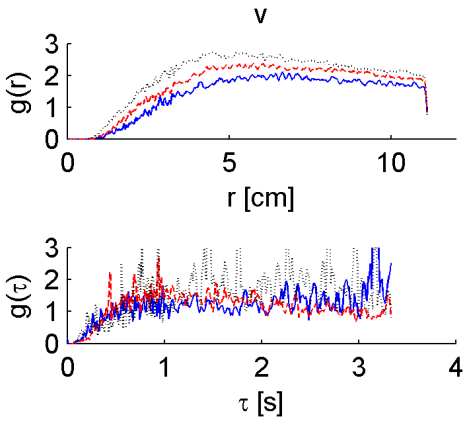
For further details and context for the figures and tables below, see section 1.

g(x) bin 1 *vs.* **g(x) bin 2**	**30 fish**	**70 fish**	**150 fish**
P-values for area normalising for $g(r)$			
$0 >= \theta > \pi/3$ *vs.* $\pi/3 >= \theta > 2\pi/3$	<0.001*	<0.001*	<0.001*
$0 >= \theta > \pi/3$ *vs.* $2\pi/3 >= \theta > \pi$	<0.001*	<0.001*	<0.001*
$\pi/3 >= \theta > 2\pi/3$ *vs.* $2\pi/3 >= \theta > \pi$	<0.001*	0.032*	<0.001*
$v > 10$ *vs.* $10 >= v > 3.3$	<0.001*	<0.001*	<0.001*
$v > 10$ *vs.* $3.3 >= v$	<0.001*	<0.001*	<0.001*
$10 >= v > 3.3$ *vs.* $3.3 >= v$	<0.001*	<0.001*	<0.001*
P-values for area normalising for $g(\tau)$			
$0 >= \theta > \pi/3$ *vs.* $\pi/3 >= \theta > 2\pi/3$	<0.001*	0.605	<0.001*
$0 >= \theta > \pi/3$ *vs.* $2\pi/3 >= \theta > \pi$	<0.001*	0.020*	0.642
$\pi/3 >= \theta > 2\pi/3$ *vs.* $2\pi/3 >= \theta > \pi$	0.058	0.004*	<0.001*
$v > 10$ *vs.* $10 >= v > 3.3$	<0.001*	0.006*	0.004*
$v > 10$ *vs.* $3.3 >= v$	<0.001*	0.071	<0.001*
$10 >= v > 3.3$ *vs.* $3.3 >= v$	<0.001*	0.759	0.002*
P-values for time scrambling normalising for $g(r)$			
$0 >= \theta > \pi/3$ *vs.* $\pi/3 >= \theta > 2\pi/3$	<0.001*	<0.001*	<0.001*
$0 >= \theta > \pi/3$ *vs.* $2\pi/3 >= \theta > \pi$	<0.001*	<0.001*	<0.001*
$\pi/3 >= \theta > 2\pi/3$ *vs.* $2\pi/3 >= \theta > \pi$	<0.001*	0.212	<0.001*
$v > 10$ *vs.* $10 >= v > 3.3$	<0.001*	<0.001*	<0.001*
$v > 10$ *vs.* $3.3 >= v$	<0.001*	<0.001*	0.937
$10 >= v > 3.3$ *vs.* $3.3 >= v$	<0.001*	<0.001*	<0.001*
P-values for time scrambling normalising for $g(\tau)$			
$0 >= \theta > \pi/3$ *vs.* $\pi/3 >= \theta > 2\pi/3$	0.061	<0.001*	<0.001*
$0 >= \theta > \pi/3$ *vs.* $2\pi/3 >= \theta > \pi$	0.899	<0.001*	0.024*
$\pi/3 >= \theta > 2\pi/3$ *vs.* $2\pi/3 >= \theta > \pi$	0.179	0.575	<0.001*
$v > 10$ *vs.* $10 >= v > 3.3$	0.176	<0.001*	<0.001*
$v > 10$ *vs.* $3.3 >= v$	0.011*	<0.001*	0.002*
$10 >= v > 3.3$ *vs.* $3.3 >= v$	0.077	<0.001*	<0.001*

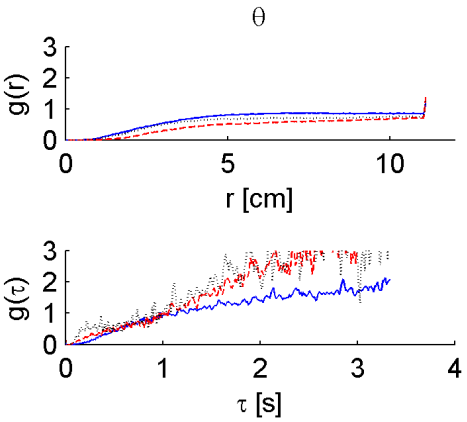
Table 1: P-values for Wilcoxon rank sum test on the $g(x)$ in figure 1, 1.9 and 2. Asterixs marks values indicate that $g(x)$ bin 1 being significant different than $g(x)$ bin 2.



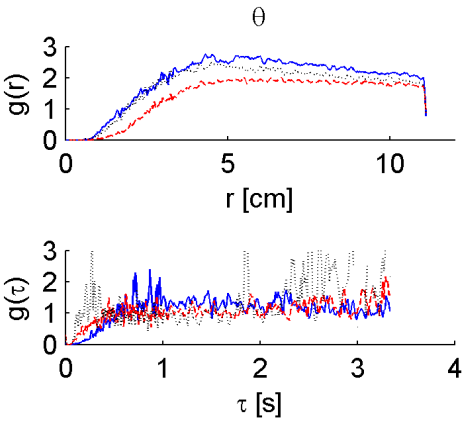
(a) Normalised using area normalisation, binned on v. Blue: $v > 10$, red: $10 \geq v > 3.3$, black: $3.3 \geq v$.



(b) Normalised using time scrambling normalisation, binned on v. Blue: $v > 10$, red: $10 \geq v > 3.3$, black: $3.3 \geq v$.

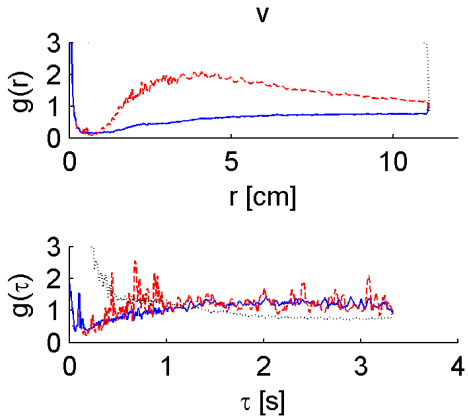


(c) Normalised using area normalisation, binned on θ. Blue: $0 \geq \theta > \pi/3$, red: $\pi/3 \geq \theta > 2\pi/3$, black: $2\pi/3 \geq \theta > \pi$.

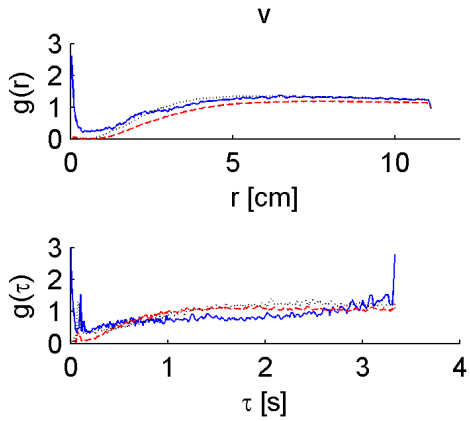


(d) Normalised using time scrambling normalisation, binned on θ. Blue: $0 \geq \theta > \pi/3$, red: $\pi/3 \geq \theta > 2\pi/3$, black: $2\pi/3 \geq \theta > \pi$.

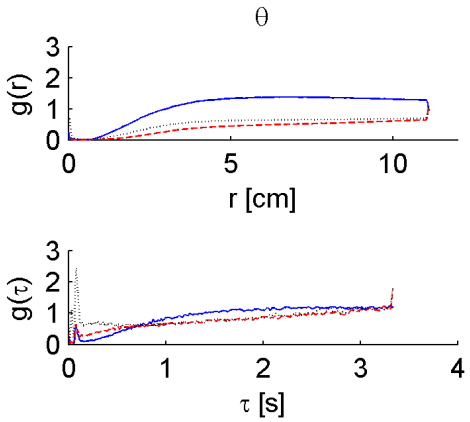
Figure 1: The pair distribution function as a function of the seperation between the two fish in the pair compared to the pair distribution function as a function of the time to collision. The pairs are binned according to the relative heading of the fish in the pair or the rate at wich they approach each other. v is measured in cm/s, θ in radians. Data from a school of 30 fish.



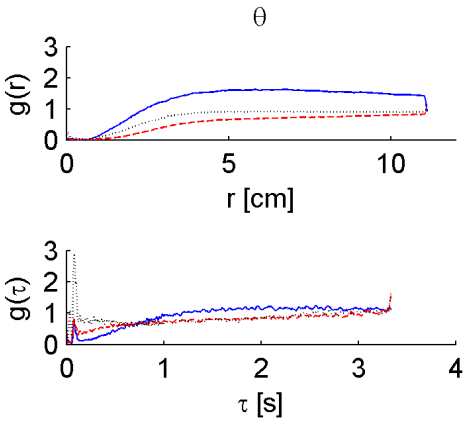
(a) Normalised using area normalisation, binned on v. Blue: $v > 10$, red: $10 \geq v > 3.3$, black: $3.3 \geq v$.



(b) Normalised using time scrambling normalisation, binned on v. Blue: $v > 10$, red: $10 \geq v > 3.3$, black: $3.3 \geq v$.



(c) Normalised using area normalisation, binned on θ. Blue: $0 \geq \theta > \pi/3$, red: $\pi/3 \geq \theta > 2\pi/3$, black: $2\pi/3 \geq \theta > \pi$.



(d) Normalised using time scrambling normalisation, binned on θ. Blue: $0 \geq \theta > \pi/3$, red: $\pi/3 \geq \theta > 2\pi/3$, black: $2\pi/3 \geq \theta > \pi$.

Figure 2: The pair distribution function as a function of the seperation between the two fish in the pair compared to the pair distribution function as a function of the time to collision. The pairs are binned according to the relative heading of the fish in the pair or the rate at wich they approach each other. v is measured in cm/s, θ in radians. Data from a school of 150 fish.

Number of fish	30	70	150
Number of timepoints	8.3e+04	9e+04	8.8e+04
Number of r measurements	2.7e+06	1e+07	2.6e+07
Number of τ measurements	3.1e+04	7.4e+04	1.3e+05

Table 2: Amount of useble data from timeseries used to make the figures in section 1. This is after frames with too few tracked fish have been removed (at least 80% must be detected by the tracking).

Model: $a \cdot \tau^b + c$				
	Area normalising		*Time scrambling normalising*	
	All data points	*With data cleaning*	*All data points*	*With data cleaning*
a	-3.3	-4.01	-3.5	-3.83
b	0.248	0.213	0.186	0.184
c	3.75	4.46	3.75	4.07
Model: $a \cdot e^{-b\tau} + c\tau$				
	Area normalising		*Time scrambling normalising*	
	All data points	*With data cleaning*	*All data points*	*With data cleaning*
a	2.32	2.39	1.88	2.02
b	1.67	1.71	2.02	2.16
c	-0.0461	-0.0441	-0.0321	-0.0211
Model: $a \cdot \tau^{-2} \exp(b\tau)$				
	Area normalising		*Time scrambling normalising*	
	All data points	*With data cleaning*	*All data points*	*With data cleaning*
a	-2.81	-2.32	-3.53	-2.85
b	-1.25	-0.865	-1.83	-1.1

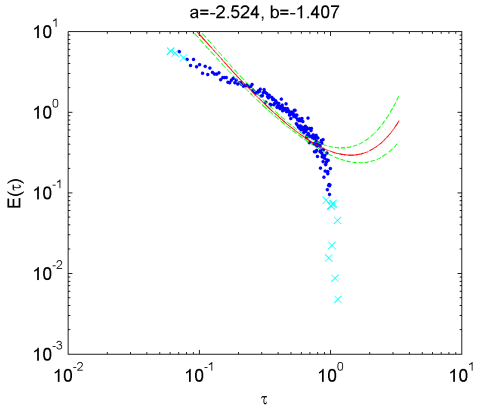
Table 3: Constants from the model fitting in section 1. The data points excluded were so for having too low values of $E(\tau)$ or being too far away from the fit (the 5% furthest from the estimated model) .

Model: $a \cdot \tau^b + c$						
	Area normalising			*Time scrambling normalising*		
	30	*70*	*150*	*30*	*70*	*150*
a	1.99	-2.94	-3.81	-1.7e+03	-2.25	-5.29
b	-0.461	0.331	0.186	0.000901	0.4	0.109
c	-1.82	3.32	4.34	1.7e+03	2.37	5.7
Model: $a \cdot e^{-b\tau} + c\tau$						
	Area normalising			*Time scrambling normalising*		
	30	*70*	*150*	*30*	*70*	*150*
a	4.98	2.35	2.14	6.47	1.79	1.93
b	3.53	1.67	1.51	7.51	1.83	1.73
c	0.0798	-0.0925	-0.00983	0.601	-0.133	0.00172
Model: $a \cdot \tau^{-2} \exp(b\tau)$						
	Area normalising			*Time scrambling normalising*		
	30	*70*	*150*	*30*	*70*	*150*
a	-2.52	-2.59	-2.32	-3.61	-3.77	-2.36
b	-1.41	-1.27	-1.01	-3.12	-2.85	-0.754

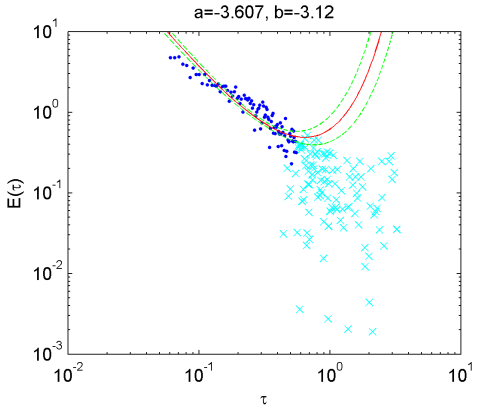
Table 4: Constants from the model fitting in section **??**. Points have been excluded from the data for having too low values of $E(\tau)$ or too far away from the fit (the 5% furthest from the estimated model) .

Model: $a \cdot \tau^b + c$						
	Area normalising			*Time scrambling normalising*		
	30	*70*	*150*	*30*	*70*	*150*
a	1.88	-2.51	-2.97	-372	-2.02	-4.19
b	-0.473	0.375	0.228	0.00459	0.399	0.13
c	-1.69	2.9	3.5	372	2.17	4.61
Model: $a \cdot e^{-b\tau} + c\tau$						
	Area normalising			*Time scrambling normalising*		
	30	*70*	*150*	*30*	*70*	*150*
a	5.83	2.21	2.08	6.86	1.65	1.85
b	4.24	1.54	1.48	8.03	1.61	1.68
c	0.22	-0.117	-0.0112	0.705	-0.182	0.000453
Model: $a \cdot \tau^{-2} \exp(b\tau)$						
	Area normalising			*Time scrambling normalising*		
	30	*70*	*150*	*30*	*70*	*150*
a	-2.46	-3.18	-2.79	-3.48	-4.43	-2.89
b	-1.13	-1.86	-1.31	-2.43	-3.81	-1.08

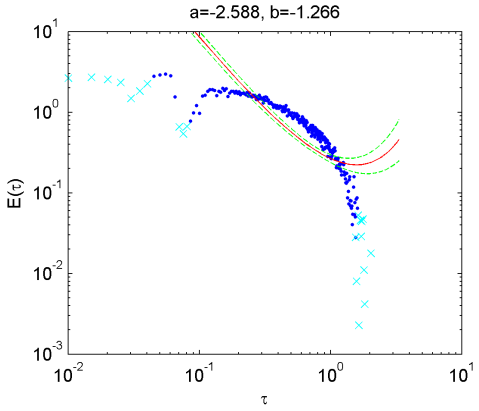
Table 5: Constants from the model fitting in section **??**.



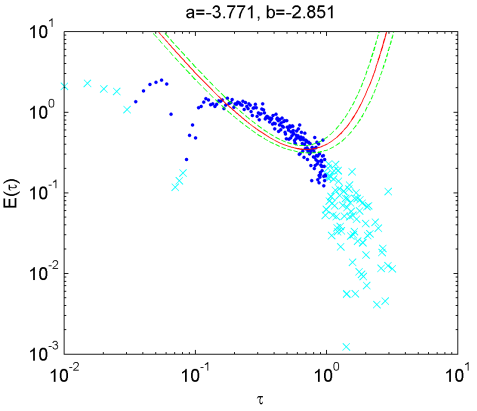
(a) Normalised using area normalisation. 30 fish.



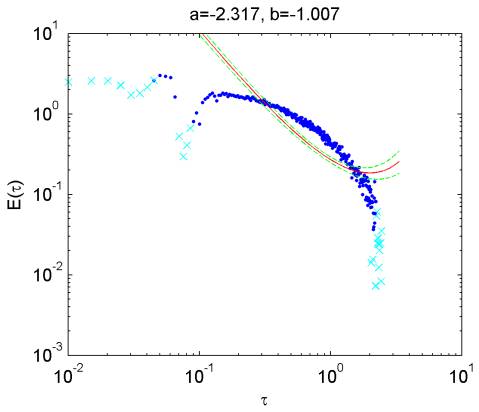
(b) Normalised using time scrambling normalisation. 30 fish.



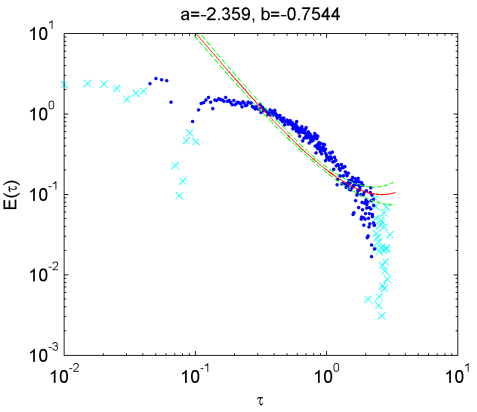
(c) Normalised using area normalisation. 70 fish.



(d) Normalised using time scrambling normalisation. 70 fish.

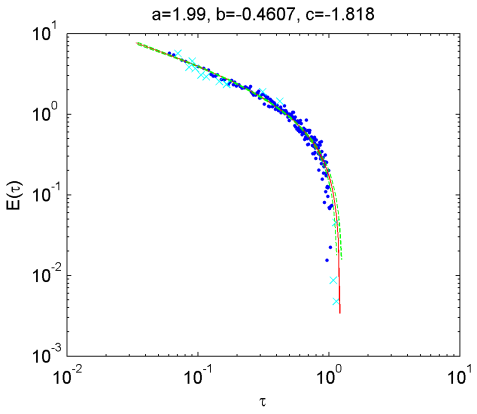


(e) Normalised using area normalisation. 150 fish.

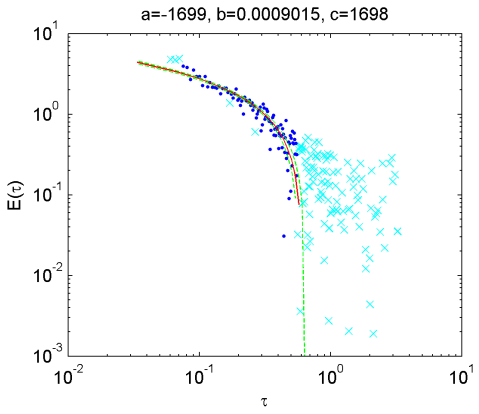


(f) Normalised using time scrambling normalisation. 150 fish.

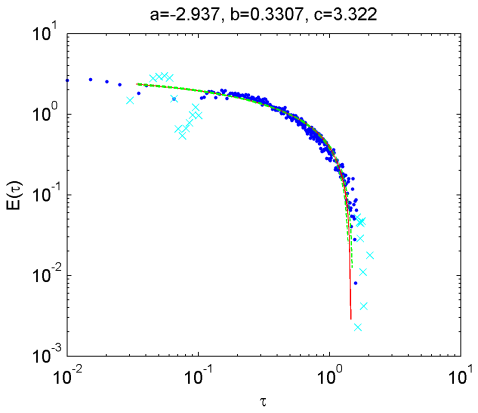
Figure 3: Data from different school sizes, fitted to (1.4). Blue: experiment data, cyan: removed data points, red: fitted model, green: confidence 95%.



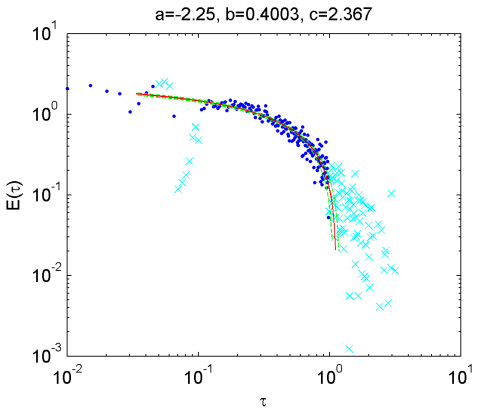
(a) Normalised using area normalisation. 30 fish.



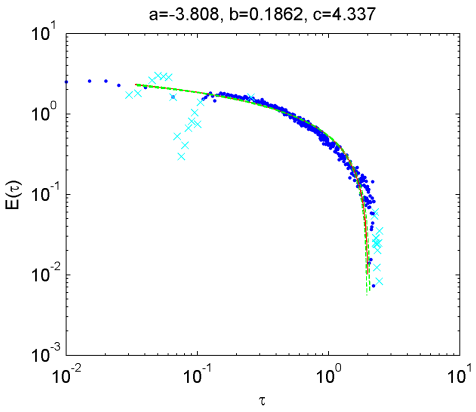
(b) Normalised using time scrambling normalisation. 30 fish.



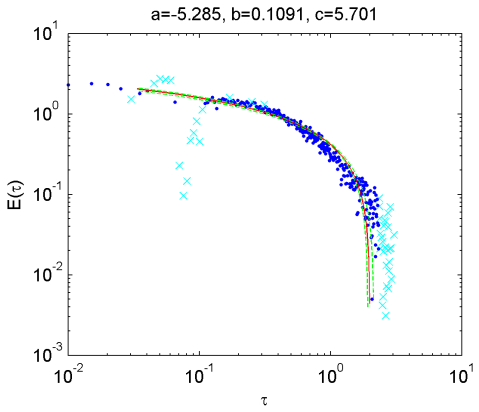
(c) Normalised using area normalisation. 70 fish.



(d) Normalised using time scrambling normalisation. 70 fish.

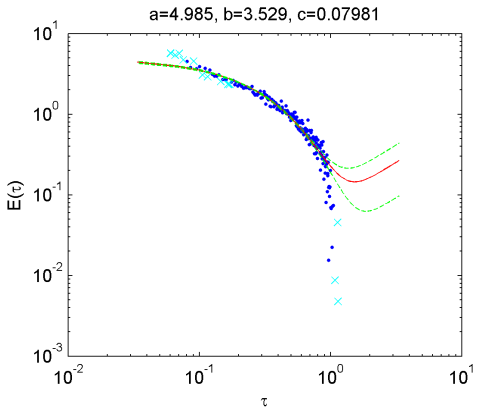


(e) Normalised using area normalisation. 150 fish.

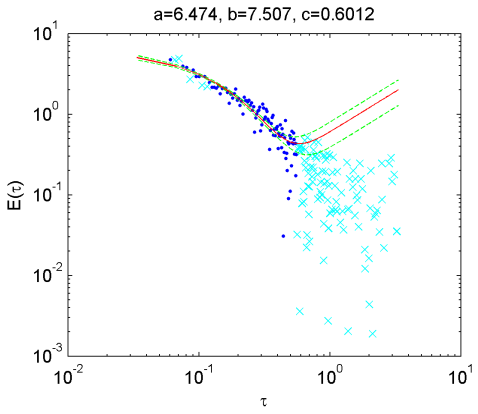


(f) Normalised using time scrambling normalisation. 150 fish.

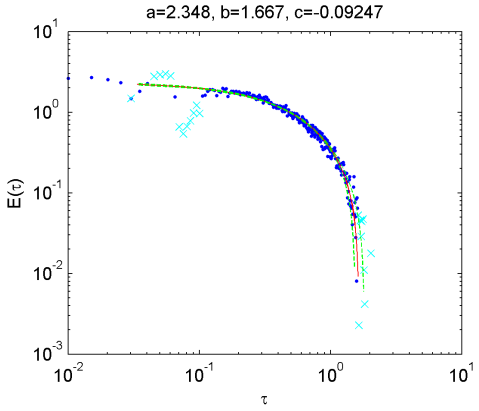
Figure 4: Data from different school sizes, fitted to (1.5). Blue: experiment data, cyan: removed data points, red: fitted model, green: confidence 95%.



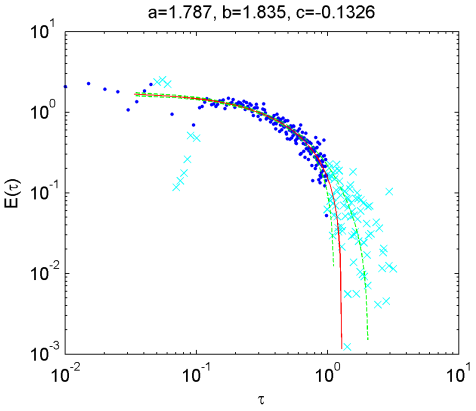
(a) Normalised using area normalisation. 30 fish.



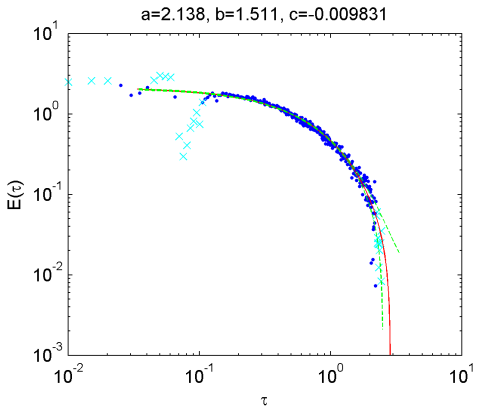
(b) Normalised using time scrambling normalisation. 30 fish.



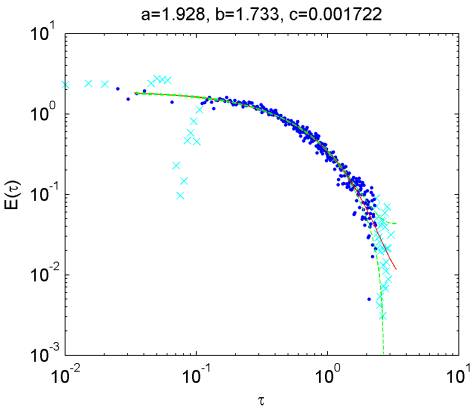
(c) Normalised using area normalisation. 70 fish.



(d) Normalised using time scrambling normalisation. 70 fish.

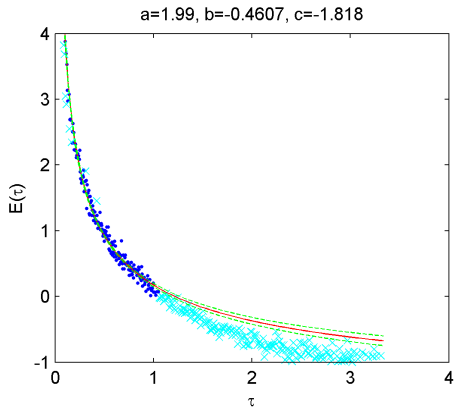


(e) Normalised using area normalisation. 150 fish.

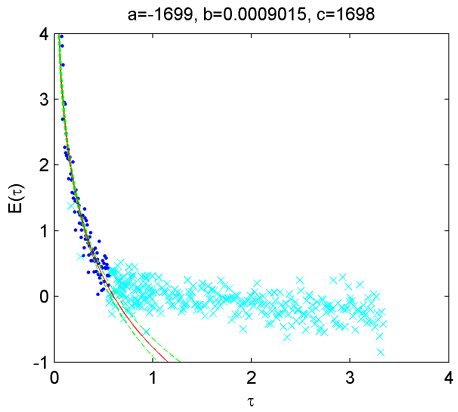


(f) Normalised using time scrambling normalisation. 150 fish.

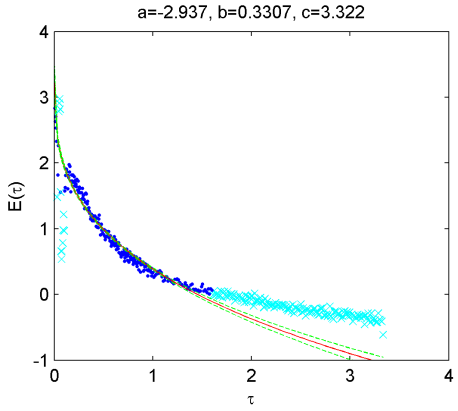
Figure 5: Data from different school sizes, fitted to (1.6). Blue: experiment data, cyan: removed data points, red: fitted model, green: confidence 95%.



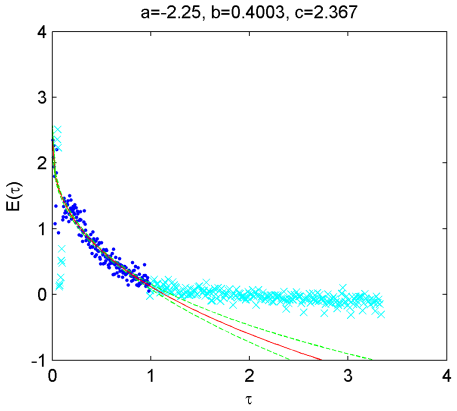
(a) Normalised using area normalisation. 30 fish.



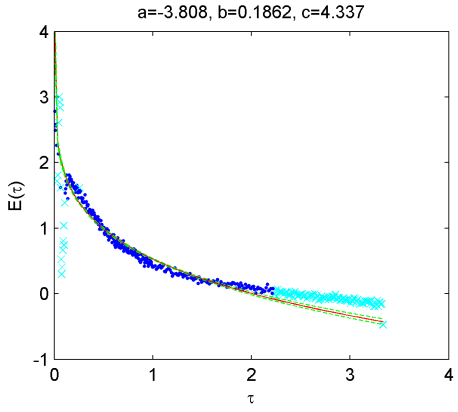
(b) Normalised using time scrambling normalisation. 30 fish.



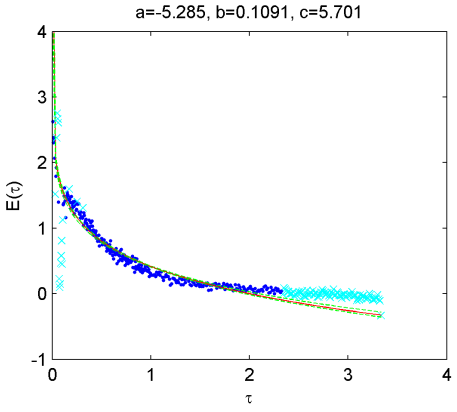
(c) Normalised using area normalisation. 70 fish.



(d) Normalised using time scrambling normalisation. 70 fish.

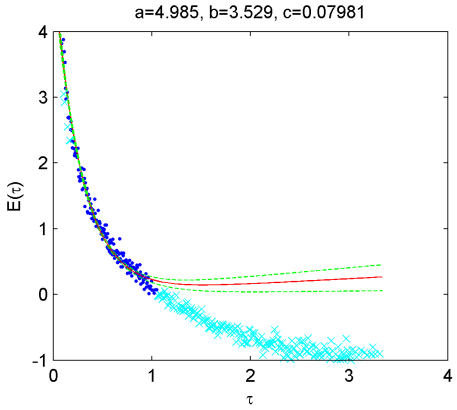


(e) Normalised using area normalisation. 150 fish.

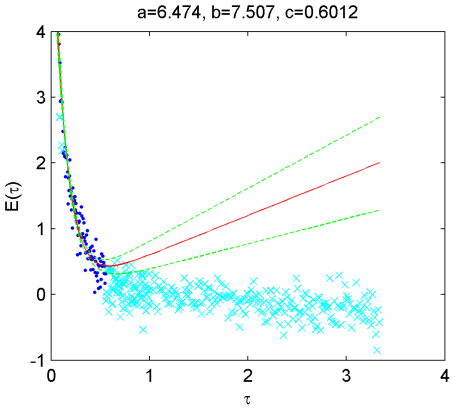


(f) Normalised using time scrambling normalisation. 150 fish.

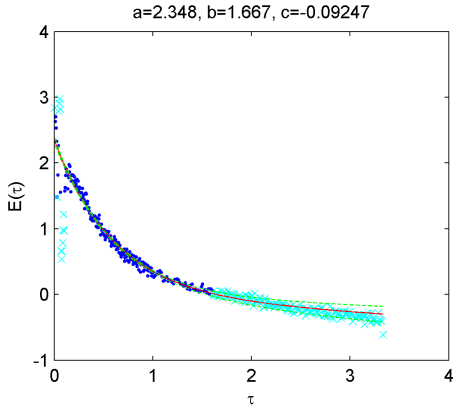
Figure 6: Data from different school sizes, fitted to (1.5). Blue: experiment data, cyan: removed data points, red: fitted model, green: confidence 95%.



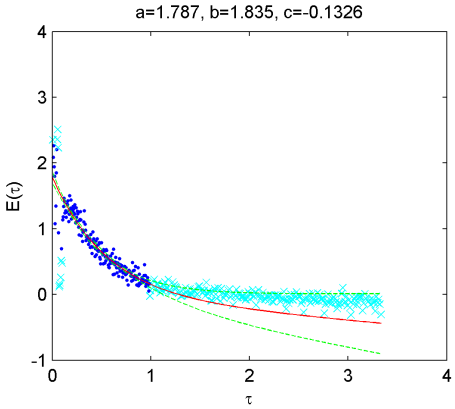
(a) Normalised using area normalisation. 30 fish.



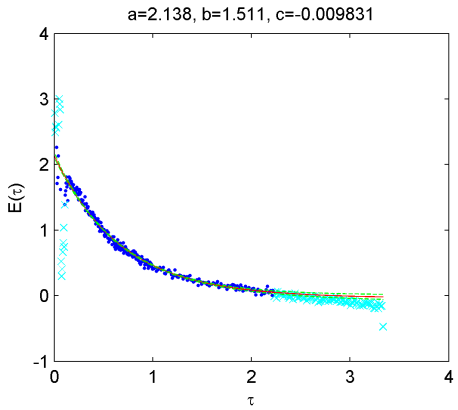
(b) Normalised using time scrambling normalisation. 30 fish.



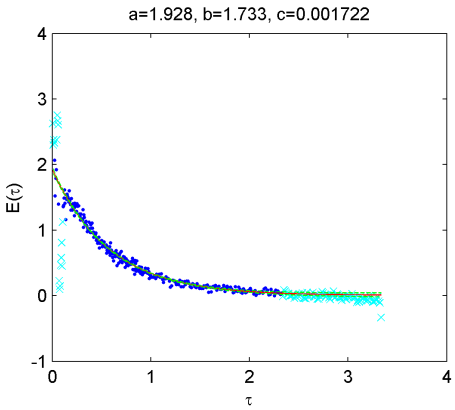
(c) Normalised using area normalisation. 70 fish.



(d) Normalised using time scrambling normalisation. 70 fish.



(e) Normalised using area normalisation. 150 fish.



(f) Normalised using time scrambling normalisation. 150 fish.

Figure 7: Data from different school sizes, fitted to (1.6). Blue: experiment data, cyan: removed data points, red: fitted model, green: confidence 95%.

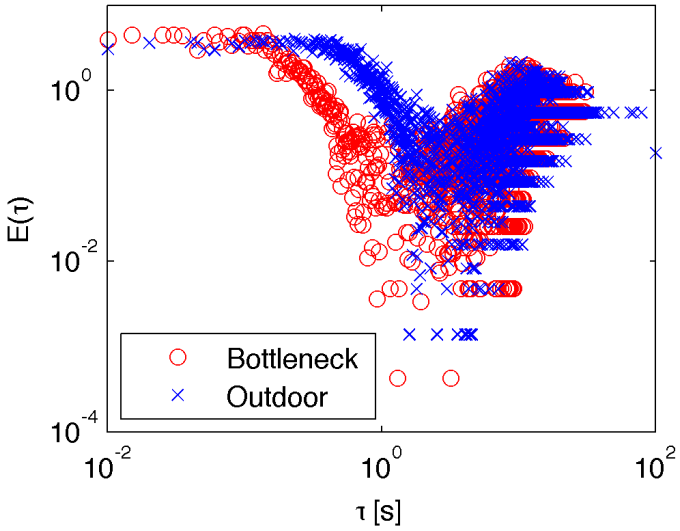


Figure 8: The interaction energy for the two datasets used in Karamouzas et al. [Karamouzas *et al.* , 2014a], all data points are included. For large times to collision the number of data points used to construct $g(\tau)$ is few making the resulting $E(\tau)$ fractions of relativly small numbers wich can be seen as lines in the plot. This justifies the need of an cut off point.

Model simulation approach

For further details and context for the figures and tables below, see sections 2 and 2.

	Mean	Median	σ	Dev. mean	Dev. median	Dev. σ
O_p	1	0.98	0.066	0.1	0.11	0.048
O_r	0.2	0.21	0.1	0.028	0.093	0.13
$r_{CMborder}$	2.6e+02	2.7e+02	68	-53	-37	76
r_{nn}	52	51	11	-0.83	-2.3	9.3
δ_r^2: a	0.78	0.76	0.075	-0.1	-0.11	0.054
δ_r^2: D	17	16	7.9	2.3	3.3	2.4
δ_{CM}^2: a	0.48	0.47	0.16	0.05	0.071	0.23
δ_{CM}^2: D	19	21	7.5	9.6	9.9	5.1
δ_{nn}^2: a	0.51	0.5	0.28	-0.073	-0.048	0.2
δ_{nn}^2: D	13	15	7.4	9.7	12	6
Deviation	1.8e+02	1.7e+02	91			
Cum. dev.	1.1e+04	1.1e+04	4.2e+03			

Table 6: Data for 6 simulations of the behavorial zones model. 3 fishes have been preserved from the experimental data.

	Mean	Median	σ	Dev. mean	Dev. median	Dev. σ
O_p	1	1	0.043	0.11	0.1	0.037
O_r	0.15	0.13	0.11	-0.023	-0.024	0.14
$r_{CMborder}$	2.3e+02	2.1e+02	75	-86	-87	82
r_{nn}	55	57	11	2.3	3.5	8.2
δ_r^2: a	0.89	0.88	0.037	0.0072	0.015	0.024
δ_r^2: D	13	14	7.2	-1	-0.76	1.5
δ_{CM}^2: a	0.42	0.41	0.13	-0.0049	0.039	0.18
δ_{CM}^2: D	11	12	6	1.5	-0.58	4.9
δ_{nn}^2: a	0.69	0.6	0.42	0.1	0.18	0.27
δ_{nn}^2: D	6.3	6.1	5.4	3	1.2	4.4
Deviation	1.5e+02	1.4e+02	98			
Cum. dev.	7e+03	6.9e+03	4.1e+03			

Table 7: Data for 6 simulations of the behavorial force model. 3 fishes have been preserved from the experimental data.

	Mean	Median	σ	Dev. mean	Dev. median	Dev. σ
O_p	0.9	0.92	0.039	0.0019	0.013	0.057
O_r	0.29	0.29	0.15	0.11	0.11	0.15
$r_{CMborder}$	1.4e+02	86	1.1e+02	- 1.7e+02	- 1.4e+02	1.4e+02
r_{nn}	46	47	11	-7.1	-7.3	11
$\bar{v}$	4.1	4.4	1.1	1.3	1.5	1.1
δ_r^2: a	0.78	0.78	0.11	-0.1	-0.097	0.075
δ_r^2: D	64	65	18	50	51	15
δ_{CM}^2: a	0.48	0.52	0.2	0.05	0.072	0.24
δ_{CM}^2: D	31	28	13	22	17	18
δ_{nn}^2: a	0.86	0.78	0.57	0.27	0.3	0.71
δ_{nn}^2: D	8.6	8.9	5.8	5.3	5	7.5
Deviation	3.4e+02	3.5e+02	98			
Cum. dev.	2.2e+04	2.1e+04	6.1e+03			

Table 8: Data for 6 simulations of the potential gradient model. 3 fishes have been preserved from the experimental data.

	Mean	Median	σ	Dev. mean	Dev. median	Dev. σ
O_p	1.1	1.2	0.46	0.21	0.26	0.43
O_r	0.18	0.21	0.061	-0.016	0.0089	0.083
$r_{CMborder}$	2.6e+02	2.6e+02	80	-49	-54	40
r_{nn}	46	42	14	-6.5	-10	12
$\bar{v}$	3.2	3.1	0.65	0.4	0.6	0.97
δ_r^2: a	0.49	0.47	0.1	-0.39	-0.41	0.13
δ_r^2: D	54	35	54	40	22	54
δ_{CM}^2: a	0.36	0.35	0.081	-0.069	-0.095	0.16
δ_{CM}^2: D	63	42	62	53	30	62
δ_{nn}^2: a	0.65	0.42	0.54	0.069	-0.14	0.62
δ_{nn}^2: D	37	15	56	33	14	55
Deviation	2.9e+02	3.3e+02	1.1e+02			
Cum. dev.	1.8e+04	2e+04	5.6e+03			

Table 9: Data for 6 simulations of the potential field model. 3 fishes have been preserved from the experimental data.

	Mean	Median	σ	Dev. mean	Dev. median	Dev. σ
O_p	0.91	0.88	0.081	0.0065	-0.012	0.091
O_r	0.17	0.14	0.12	-0.0017	0.004	0.094
$r_{CMborder}$	3.4e+02	3.5e+02	35	25	28	1.3e+02
r_{nn}	53	54	12	0.02	4.2	14
$\bar{v}$	6.6	6.5	0.65	3.8	3.3	1
δ_r^2: a	0.7	0.76	0.17	-0.19	-0.14	0.15
δ_r^2: D	89	91	17	74	78	21
δ_{CM}^2: a	0.33	0.33	0.15	-0.099	-0.14	0.21
δ_{CM}^2: D	82	94	32	72	91	37
δ_{nn}^2: a	0.7	0.62	0.32	0.11	0.044	0.42
δ_{nn}^2: D	14	16	8.5	11	13	9.2
Deviation	3.4e+02	3.4e+02	59			
Cum. dev.	2.1e+04	2.2e+04	4.3e+03			

Table 10: Data for 6 simulations of the mean destination model. 3 fishes have been preserved from the experimental data.

Bibliography

[Cri, n.d.] *Rdf schematic.* `http://commons.wikimedia.org/wiki/File:Rdf_schematic.jpg`. Accessed: 2015-04-25.

[Cavagna *et al.* , 2013] Cavagna, A., Queirós, S. M. Duarte, Giardina, I., Stefanini, F., & Viale, M. 2013. Diffusion of individual birds in starling flocks. *Proceedings of the Royal Society of London B: Biological Sciences*, **280**(1756).

[Couzin *et al.* , 2002] Couzin, Iain D., Krause, Jens, James, Richard, Ruxton, Graeme D., & Franks, Nigel R. 2002. Collective Memory and Spatial Sorting in Animal Groups. *Journal of Theoretical Biology*, **218**(1), 1 – 11.

[Edelsbrunner *et al.* , 2006] Edelsbrunner, H., Kirkpatrick, D., & Seidel, R. 2006. On the Shape of a Set of Points in the Plane. *IEEE Trans. Inf. Theor.*, **29**(4), 551–559.

[Grégoire *et al.* , 2003] Grégoire, Guillaume, Chaté, Hugues, & Tu, Yuhai. 2003. Moving and staying together without a leader. *Physica D: Nonlinear Phenomena*, **181**(34), 157 – 170.

[Guy *et al.* , 2009] Guy, Stephen J., Chhugani, Jatin, Kim, Changkyu, Satish, Nadathur, Lin, Ming C., Manocha, Dinesh, & Dubey, Pradeep. 2009. ClearPath: Highly Parallel Collision Avoidance for Multi-Agent Simulation. *Pages 177–187 of: ACM SIGGRAPH/EUROGRAPHICS SYMPOSIUM ON COMPUTER ANIMATION*. ACM.

[Herbert-Read *et al.* , 2011] Herbert-Read, James E., Perna, Andrea, Mann, Richard P., Schaerf, Timothy M., Sumpter, David J. T., & Ward, Ashley J. W. 2011. Inferring the rules of interaction of shoaling fish. *Proceedings of the National Academy of Sciences*, **108**(46), 18726–18731.

[Karamouzas *et al.* , 2014a] Karamouzas, Ioannis, Skinner, Brian, & Guy, Stephen J. 2014a. Universal Power Law Governing Pedestrian Interactions. *Phys. Rev. Lett.*, **113**(Dec), 238701.

[Karamouzas *et al.* , 2014b] Karamouzas, Ioannis, Skinner, Brian, & Guy, Stephen J. 2014b. *Universal Power Law Governing Pedestrian Interactions.* `http://motion.cs.umn.edu/PowerLaw/`. [Online; accessed 2015-02-21].

[Katz *et al.* , 2011] Katz, Yael, Tunstrøm, Kolbjørn, Ioannou, Christos C., Huepe, Cristián, & Couzin, Iain D. 2011. Inferring the structure and dynamics of interactions in schooling fish. *Proceedings of the National Academy of Sciences*, **108**(46), 18720–18725.

[Mogilner *et al.* , 2003] Mogilner, A., Edelstein-Keshet, L., Bent, L., & Spiros, A. 2003. Mutual interactions, potentials, and individual distance in a social aggregation. *Journal of Mathematical Biology*, **47**(4), 353–389.

[Moussaïd *et al.* , 2011] Moussaïd, Mehdi, Helbing, Dirk, & Theraulaz, Guy. 2011. How simple rules determine pedestrian behavior and crowd disasters. *Proceedings of the National Academy of Sciences*, **108**(17), 6884–6888.

[Sarnitskiy, n.d.] Sarnitskiy, Grigory. *Lennard-Jones Radial Distribution Function.* `http://commons.wikimedia.org/wiki/File:Lennard-Jones_Radial_Distribution_Function.svg`. Accessed: 2015-04-25.

[Sbalzarini & Koumoutsakos, 2005] Sbalzarini, I.F., & Koumoutsakos, P. 2005. Feature point tracking and trajectory analysis for video imaging in cell biology. *Journal of Structural Biology*, **151**(2), 182 – 195.

[Tunstrøm *et al.* , 2013] Tunstrøm, Kolbjørn, Katz, Yael, Ioannou, Christos C., Huepe, Cristián, Lutz, Matthew J., & Couzin, Iain D. 2013. Collective States, Multistability and Transitional Behavior in Schooling Fish. *PLoS Computational Biology*, **9**(2), e1002915.