Alma Mater Studiorum · Università di Bologna

Scuola di Scienze
Dipartimento di Fisica e Astronomia
Corso di Laurea in Fisica

Classification of Clausocalanus Furcatus Mobility

Relatore:
Prof. Armando Bazzani

Correlatore:
Dott. Nico Curti

Presentata da:
Diego Cardinali

Anno Accademico 2019/2020

A mio nonno Marino
Sit tibi terra levis

06/06/1942
10/10/2020

Abstract

The purpose of this work is the classification of the main properties of the mobility of the species *Clausucalanus furcatus*. This classification is then used to propose a model for the motion of the specimens divided in a *fast* regime a *slow* regime, with an ulterior distinction between environments in which food is present or absent. For the slow regime the specimens appear to move mostly upwards or downwards for short time periods, whereas the motion in the fast regime is compatible with the Lévy flight foraging hypothesis. In environments where food is scarce *C. Furcatus* also appears to adopt some energy conservation strategies, increasing the time between more expensive activities such as long jumps.

Sommario

Scopo di questo lavoro è la classificazione delle principali proprietà di mobilità della specie *Clausocalanus Furcatus*. Questa classificazione è poi utilizzata per proporre un modello per il moto degli esemplari diviso in un regime *veloce* e un regime *lento*, con ulteriore distinzione sulla presenza o meno di cibo nell'ambiente. Nel regime lento gli individui tendono a muoversi di moto ascendente o discendente per brevi periodi di tempo, mentre il regime veloce appare compatibile con l'ipotesi di foraggiamento tramite processi di Lévy. In ambienti privi di cibo emerge anche l'adozione da parte di *C. Furcatus* di strategie di risparmio energetico, dilatando i tempi tra attività più dispendiose quali spostamenti particolarmente lunghi.

Contents

Introduction

The purpose of this work is the classification of the main properties of the mobility of *Clausocalanus furcatus*, a species of copepods that can be found in various marine locations including the gulf of Naples (IT), where the samples analysed in this work have been collected.
The main goal of this thesis is to propose a simplified model for the motion of the specimens of *C. furcatus* based mainly on two factors: velocity of the individual and presence (or absence) of food in the environment. Regarding the velocity, the motion of the creatures has been divided in *fast* and *slow* regime based on specific thresholds that emerged during the analysis and that seem to depend on whether or not food is present. The second factor, other than dictating the velocity threshold, also seems to be related to the behaviour of the creatures concerning spatial exploration.
This work is divided into chapters, what follows now is a brief overview of the content of each one.
The first chapter contains a brief introduction to some concepts of probability, mostly focused on power-law probability density functions, and stochastic processes. Afterwards Markov processes are introduced and then, more specifically, diffusion processes. These are determinant to introduce the main stochastic process used in the proposed model, the Lévy flight. After presenting these objects, the Lévy flight foraging hypothesis is quickly introduced. The chapter concludes with the concept of hazard function from survival analysis, where a physical interpretation is given for the hazard function of a power-law probability density function.
The second chapter contains an overview of the concept of Frenet-Serret frame for a 3-dimensional curve, then the calculations for the curvature and torsion of a helix, that are used to test the accuracy and behaviour of a portion of the code subsequently used in the analysis.
The third chapter follows the steps of the analysis performed, first by examining the velocities and determining threshold values to distinguish between motion regimes, then by exploring various properties of the motion. After the division, a test is conducted to check whether or not the mean behaviour of the individuals matches that of the whole set of data, followed by a classification of the durations and lengths of the trajectories. The final part of the chapter contains a more detailed overview of the space exploration properties for the trajectories by examining the curvature and torsion probability density functions.
The fourth and final chapter contains a summarisation of all the previous results, which are combined to formulate the proposed model for the motion of *C. Furcatus*.
To conclude, a small appendix contains an overview of the Kolmogorov-Smirnov test, which is used throughout this thesis to check the confidence level of all curve fits on the

probability density functions.
All the code used for the analysis in this thesis is free and open source software, and it is available at [Car20a], [Car20b] under the terms of *The Unlicense*.

1. Elements of Probability and Statistics

1.1 Probability

Let $\mathscr{P}(x)$ be a probability density function (PDF) for a random variable ξ, then its associated cumulative distribution (CDF) is defined as

$$P(x) \equiv P(\xi \leq x) \stackrel{\text{def}}{=} \int_{-\infty}^{x} \mathscr{P}(y)\,\mathrm{d}y \ . \tag{1.1}$$

By definition of PDF it follows immediately that

$$P(+\infty) = \lim_{x\to+\infty} \int_{-\infty}^{x} \mathscr{P}(y)\,\mathrm{d}y = 1 \ , \tag{1.2}$$

so another useful definition is that of the complementary CDF [1]

$$\overline{P}(x) \stackrel{\text{def}}{=} 1 - P(x) = \int_{x}^{+\infty} \mathscr{P}(y)\,\mathrm{d}y \ . \tag{1.3}$$

In the *Data Analysis* chapter a commonly used PDF will be the power law

$$\mathscr{P}(x) = C\,(x - x_0)^{\alpha} \ , \tag{1.4}$$

with $C, \alpha, x_0 \in \mathbb{R}$ such that the normalisation condition is satisfied over the dominion. Whenever these distributions will be applied they will represent the actual PDF only for values of x such that $x - x_0 > 0$, so the fact that the base is positive also implies that for all the cases in this thesis it will always be assumed $\alpha < 0$, otherwise the PDF would diverge. Given these considerations, it is useful to rewrite (1.4) as

$$\mathscr{P}(x) = \begin{cases} C\,(x - x_0)^{\alpha} & x \geq x_0 \\ 0 & x < x_0 \end{cases} \ . \tag{1.5}$$

The CDF for a PDF in the form (1.5) is found by solving the integral (1.1)

$$\begin{aligned} P(x) &= \int_{-\infty}^{x} C\,(y - x_0)^{\alpha}\,\mathrm{d}y = \int_{x_0}^{x} C\,(y - x_0)^{\alpha}\,\mathrm{d}y = \\ &\underset{\alpha\neq-1}{=} \left[\frac{C}{\alpha+1}\,(y - x_0)^{\alpha+1}\right]_{x_0}^{x} = \frac{C}{\alpha+1}\,(x - x_0)^{\alpha+1} \ . \end{aligned} \tag{1.6}$$

[1] Often also referred to as survival function and denoted by S.

1.2 Stochastic Processes and Markov Processes

A stochastic system is a system that evolves probabilistically in time or, more precisely, any system in which there exists a time-dependent random variable $\xi(t)$, of which various values x_i can be measured at times t_i. In any stochastic system one can always assume the existence of a set of PDFs called joint probability densities

$$\mathscr{P}(x_1, t_1; x_2, t_2; ...) \ . \tag{1.7}$$

Given $\mathscr{P}$, it is also possible to introduce a conditional joint probability of a series of events $(x_1, t_1; x_2, t_2; ...)$ provided that other events $(y_1, \tau_1; y_2, \tau_2; ...)$ have already happened. This conditional PDF is indicated as

$$\mathscr{P}(x_1, t_1; x_2, t_2; ...|y_1, \tau_1; y_2, \tau_2; ...) \stackrel{\text{def}}{=} \frac{\mathscr{P}(x_1, t_1; x_2, t_2; ...; y_1, \tau_1; y_2, \tau_2; ...)}{\mathscr{P}(y_1, \tau_1; y_2, \tau_2; ...)} \ , \tag{1.8}$$

where times are usually considered ordered decreasingly (even though these probability densities are independent of the order of times), giving

$$t_1 \geq t_2 \geq ... \geq \tau_1 \geq \tau_2 \geq ... \ . \tag{1.9}$$

To define a stochastic process it is necessary to know at least all joint probabilities (1.7) [Gar09], and the process will be called separable if this is enough to define it. There are four kinds of such processes: completely independent, Bernoulli trials, martingales and Markov processes. For this thesis only the last ones are relevant, and so will be expanded upon.

In a Markov Process the knowledge of the present determines the future via a relation called Markov assumption. If times are ordered as in (1.9), the Markov assumption requires the conditional probability to be entirely determined by the knowledge of the most recent condition, i.e.

$$\mathscr{P}(x_1, t_1; x_2, t_2; ...|y_1, \tau_1; y_2, \tau_2; ...) = \mathscr{P}(x_1, t_1; x_2, t_2; ...|y_1, \tau_1) \ . \tag{1.10}$$

This assumption means that everything can be defined in terms of the simple conditional probabilities $\mathscr{P}(x_1, t_1|y_1, \tau_1)$. As an example, given

$$\mathscr{P}(x_1, t_1; x_2, t_2|y_1, \tau_1) = \mathscr{P}(x_1, t_1|x_2, t_2; y_1, \tau_1)\, \mathscr{P}(x_2, t_2|y_1, \tau_1) \ ,$$

then from the Markov assumption it follows

$$\mathscr{P}(x_1, t_1; x_2, t_2; y_1, \tau_1) = \mathscr{P}(x_1, t_1|x_2, t_2)\, \mathscr{P}(x_2, t_2|y_1, \tau_1) \ ,$$

where each $\mathscr{P}$ is now fully characterised based only on the immediately preceding event. For the sake of completeness, this example can be generalised for an arbitrary joint probability as [Gar09]

$$\mathscr{P}(x_1, t_1; ...; x_n, t_n) = \mathscr{P}(x_n, t_n) \prod_{i=1}^{n-1} \mathscr{P}(x_i, t_i|x_{i+1}, t_{i+1}) \ . \tag{1.11}$$

1.3 Diffusion Processes

A diffusion process is a particular kind of Markov process widely used to model the movement of objects of any kind, from particles to small animals. Any diffusion process can be divided into two phases: in the first one the objects are in a non-equilibrium environment, and thus there is a concentration gradient that drives the process; when the concentration of the particles drifting is uniform the gradient becomes zero and thus the particles do not drift anymore as an ensemble (but they still move randomly in all directions). Diffusion processes are usually modelled via partial differential equations, as these tend to be easy to solve numerically with a good degree of precision. Let ξ be a random variable associated to a continuous-state, continuous-time Markov process $\Xi(t)$, that is, a process in which both ξ and t are continuous. $\Xi(t)$ is also called the sample path of ξ, i. e. the time progression of the samples drawn from ξ. The conditional CDF for the process is

$$P(x,t|y,\tau) = P(\Xi(t) \leq x|\Xi(\tau) = y) \ . \tag{1.12}$$

if the derivative of P with respect to x exists, it is called the transition density function of the process and its expression is

$$\mathscr{P}(x,t|y,\tau) = \partial_x P(x,t|y,\tau) \ . \tag{1.13}$$

In order for Ξ to be considered a diffusion process the required conditions are:

- continuity of $\mathscr{P}$, which implies continuity of Ξ [Ibe13];
- the mean $\langle\cdot\rangle$ of Ξ is related to $\mathscr{P}$ in the form

$$\lim_{\Delta t\to 0} \frac{\langle \Xi(t+\Delta t) - \Xi(t) \,|\Xi(t) = y\rangle}{\Delta t} = \frac{1}{\Delta t}\int_{\mathbb{R}} (x-y)\,\partial_x \mathscr{P}(x,t+\Delta t|y,t)\,\mathrm{d}x \stackrel{\text{def}}{=} a(y,t) \ , \tag{1.14}$$

  where a is known as the instantaneous drift of Ξ;

- the variance of Ξ is related to $\mathscr{P}$ in the form

$$\lim_{\Delta t\to 0} \frac{\left\langle [\Xi(t+\Delta t) - \Xi(t)]^2 |\Xi(t) = y\right\rangle}{\Delta t} = \frac{1}{\Delta t}\int_{\mathbb{R}^n} (x-y)^2\,\partial_x \mathscr{P}(x,t+\Delta t|y,t)\,\mathrm{d}x = b(y,t) \ , \tag{1.15}$$

  where b is known as the instantaneous variance of Ξ.

If Δt is discrete, let p be the probability that Ξ moves of Δx after a time increase of Δt and $q = 1 - p$ the probability that it moves of $-\Delta x$. If $p - q$ tends to 0 as Δt and Δx tend to 0 it is possible to introduce the coefficients

$$\mu \stackrel{\text{def}}{=} \lim_{\Delta t, \Delta x \to 0} \frac{(p-q)\,\Delta x}{\Delta t} \ , \tag{1.16}$$

$$\mathcal{D} \stackrel{\text{def}}{=} \lim_{\Delta t, \Delta x \to 0} \frac{(\Delta x)^2}{2\Delta t} \ . \tag{1.17}$$

If μ and $\mathcal{D}$ are constants the diffusion process is described by the Fokker-Planck equation

$$\partial_t \mathscr{P}\left(x, t | x_0, t_0\right) = -\mu \partial_x \mathscr{P}\left(x, t | x_0, t_0\right) + \mathcal{D} \partial_x^2 \mathscr{P}\left(x, t | x_0, t_0\right) \ . \tag{1.18}$$

A characteristic quantity for diffusion processes is the mean square displacement (MSD). Let there be a system with n particles identified by the index i, each with starting position x_{0i} and a position $x_i\left(t\right)$ at time t, then the MSD is

$$X^2\left(t\right) \stackrel{\text{def}}{=} \frac{1}{n} \sum_{i=1}^{n} \left\langle \left[x_i\left(t\right) - x_{0i}\right]^2 \right\rangle \ . \tag{1.19}$$

This quantity is closely related to the diffusion coefficient $\mathcal{D}$ by the power law [BH00]

$$X^2 \propto D t^{\gamma} \tag{1.20}$$

with $\gamma \in \mathbb{R}$. The value of γ is determinant for the type of diffusion. If $\gamma = 1$ the MSD is proportional to the diffusion coefficient the process is called standard (or Fickian) diffusion, and it obeys Fick's law, otherwise it is called anomalous diffusion, and in particular sub-diffusion if $\gamma < 1$, super-diffusion if $\gamma > 1$ [Kla+90].
The case for $\gamma = 1$ is the standard Brownian Motion, as it is proven in [Zan], meaning that the MSD is directly proportional to the diffusion coefficient, so space exploration grows linearly in time.
In the case $\gamma < 1$ the underlying probability density for a particle to be at a position r at time t is non-Gaussian. This regime is usually obtained by adding physical or temporal constraints to the system, limiting the freedom of movement for the particles. The regime $\gamma > 1$ on the other hand is also often called enhanced transportation, and it will be expanded upon in the next section of this chapter.

1.4 Lévy Flights and the Lévy Flight Foraging Hypothesis

Lévy Flights are processes characterised by a PDF for the length of the step in the form

$$\mathscr{P}(x) \propto x^{\alpha} \qquad\qquad -3 \leq \alpha < -1 \ , \tag{1.21}$$

so that their sample path is described by clusters of random points interjected by infrequent longer jumps, the flights that give the phenomenon its name. The mean and the variance of $\mathscr{P}$ diverge along with the MSD [Kla+90]. This last quantity in particular satisfies Eq. (1.20) for $\gamma > 1$, meaning that Lévy flights are topical examples of super-diffusive processes.
Because of the aforementioned non-convergence problems, Lévy Flights are often truncated when applied to real-life scenarios, using a PDF in the form

$$\begin{cases} \mathscr{P}(x) \propto x^{\alpha} & x_0 < x \leq \overline{x} \ , \\ 0 & \text{otherwise} \ . \end{cases} \tag{1.22}$$

In this case $\overline{x}$ takes the name of cutoff parameter and ensures that the previously divergent parameters now converge to a finite value. In this thesis all Lévy flights will be automatically truncated by programmatically choosing as cutoff parameter a value after which the noise on the data becomes too significant. The value of x_0 is usually 0, but in this work it is necessary to introduce it as not all distributions immediately present a Lévy flight-like fall, so this parameter is used to start fitting when the power-law regime begins.
While Lévy flights have many applications, the one that will be explored in this work is that of the Lévy Flight foraging hypothesis; formulated first in 1999 by Viswanathan et al. in [Vis+99], it can be formally stated as follows:

> *"Since Lévy flights (actually, truncated Lévy flights) optimize random searches, biological organisms must have therefore evolved to exploit Lévy flights."* [Rap+09]

The original work from 1999 stated that, under the assumptions of randomly distributed revisitable foraging sites, if Markov assumption (i. e. absence of memory) also holds, the optimal strategy for visiting as many sites as possible is that of an inverse square power-law.
This hypothesis has been heavily debated over the years, finding various confirmations, but also having many results overturned. A recent analysis [Hum+12] found strong evidence for a truncated Lévy flight driven exploration in wandering albatross (*Diomedea exulans*) and black-browed albatross (*Thalassarche melanophrys*) when information about the environment is not complete, i. e. the animal can not rely on visual, auditory or olfactory clues to identify the prey from afar or is not familiar with the environment.
Since the purpose of this work is to investigate the behaviour of specimens of *Clausocalanus Furcatus* both in conditions where food is present and where it is not, a main driving idea of this investigation will be that the individuals tend to explore the space in a Lévy flight-like distribution when going over a certain threshold velocity, as to verify whether or not this hypothesis could apply to the motion of this species.

1.5 Hazard Functions

Later in this work a key part of the behavioural analysis performed will be focused on characterising the duration of various tracks, so it is relevant to introduce an important concept of survival analysis: the hazard function. Under the hypothesis of the Markov assumption for the time evolution of the system, the complementary CDF satisfies [GBR15]

$$\overline{P}(t+\Delta t) = [1-\lambda(t)\,\Delta t]\,\overline{P}(t) + o(\Delta t)\ , \tag{1.23}$$

where λ is the hazard function, i. e. the event rate at t conditional on a duration of at least t. In the continuous case, for the limit of Δt that goes to 0, from this relation it is possible to find an expression for $\lambda(t)$ by simple algebraic manipulation of (1.23):

$$\begin{aligned}
&\frac{\overline{P}(t+\Delta t) - o(\Delta t)}{\overline{P}(t)} = 1-\lambda(t)\,\Delta t \\
\Longrightarrow &1 - \frac{\overline{P}(t+\Delta t) - o(\Delta t)}{\overline{P}(t)} = \lambda(t)\,\Delta t \\
\Longrightarrow &\frac{1}{\overline{P}(t)}\frac{\overline{P}(t) - \overline{P}(t+\Delta t) - o(\Delta t)}{\Delta t} = \lambda(t)\ ,
\end{aligned}$$

and by taking the limit it follows that

$$\begin{aligned}
\lambda(t) &= \lim_{\Delta t\to 0}\frac{1}{\overline{P}(t)}\frac{\overline{P}(t) - \overline{P}(t+\Delta t) - o(\Delta t)}{\Delta t} \\
\Longrightarrow \lambda(t) &= -\frac{\mathrm{d}_t\overline{P}(t)}{\overline{P}(t)}\ .
\end{aligned}$$

An immediate consequence of 1.3 is

$$\mathrm{d}_t\overline{P}(t) = \mathrm{d}_t\,(1-P(t)) =$$

so the final expression for the hazard function is

$$\lambda(t) = \frac{\mathscr{P}(t)}{\overline{P}(t)} = \frac{\mathscr{P}(t)}{1-\int_{-\infty}^{t}\mathscr{P}(\tau)\,\mathrm{d}\tau}\ . \tag{1.24}$$

Via direct substitution of the Eqs. (1.5, 1.6) in the newfound expression, it is possible to find the theoretical expression for the hazard function of the power law probability density function

$$\lambda(t) = \frac{C\,(t-t_0)^{\alpha}}{1-\frac{C}{\alpha+1}\,(t-t_0)^{\alpha+1}}\ . \tag{1.25}$$

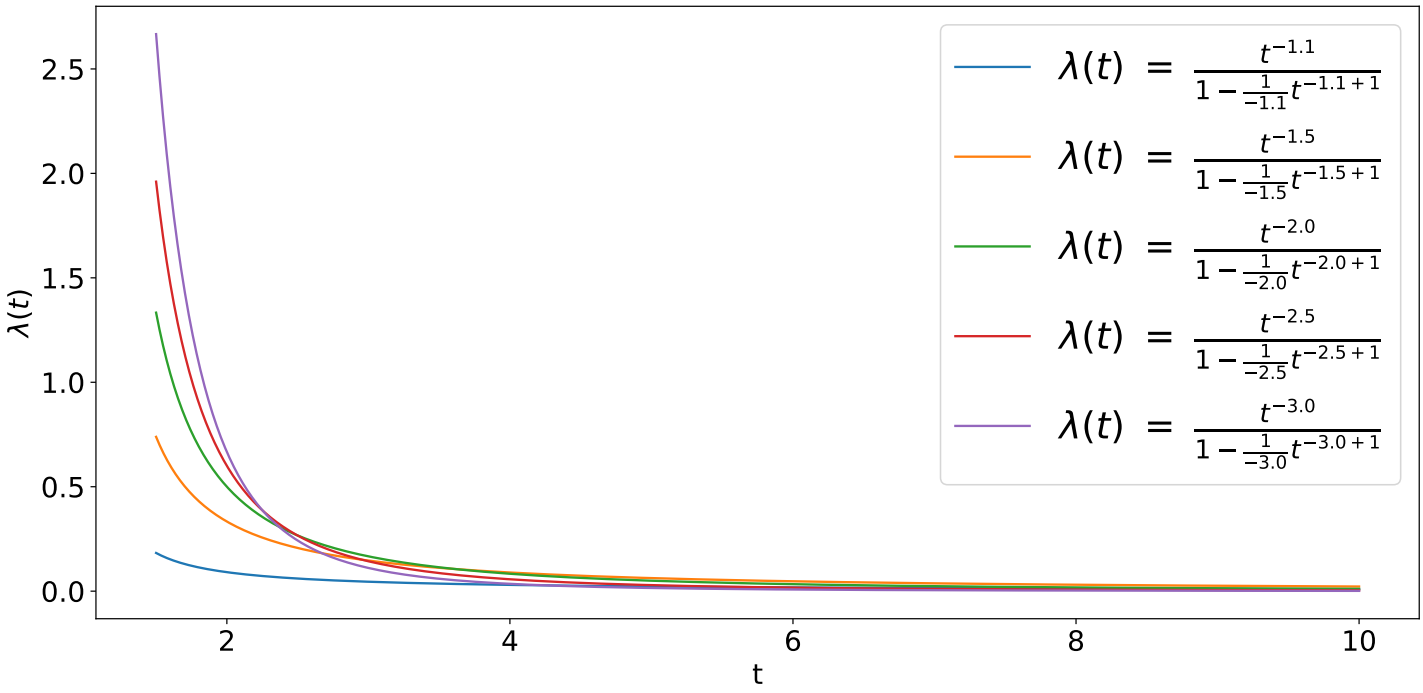


Figure 1.1: Various hazard functions for PDFs in the form (1.4) with $C = 1$, $t_0 = 1.5$ and a cutoff parameter $\bar{t} = 10$. These values for α are all in the interval $-3 \leq \alpha < -1$ to remain within the boundaries of the Lévy flight phenomena. The minimum value chosen for the exponent was $\alpha = -1.1$ as $\alpha = -1$ would cancel the denominator.

In the case of a Lévy flight distribution ($-3 \leq \alpha < -1$) some notable examples for the hazard function can be seen in Fig. 1.1 (case $C = 1$, $t_0 = 1.5$, $\bar{t} = 10$). These allow to give a reasonable physical interpretation to the hazard function of a power-law probability density function. If the PDF is interpreted as the distribution of durations of a trip for a particle, the hazard function represents the conditional probability density function that a trip will end at a certain time $t + \Delta t$ provided that it lasted t.
As it can be seen in the figure, the function is strictly decreasing, meaning that for short trips it is much more likely for the particle to stop. As time passes and the trip becomes longer however, the probability to stop quickly decreases, until the cutoff parameter $\bar{t}$ is reached. The purpose of the cutoff parameter is to make it impossible for the particle to travel for an infinite amount of time, since that would clearly not be physical.
This incremental incentive to make the trip longer suggest a sort of intentionality, i. e. when a trip lasts longer it means that the traveller has a specific intent of reaching a further point, so in the case of an animal it may be interpreted as a deliberate choice to search for some points of interest (food, a mate, a shelter...).

2. Frenet-Serret Curve Theory

2.1 Frenet-Serret Apparatus

To properly introduce a Frenet-Serret apparatus for a point it is first necessary to characterise its motion. Motion can be defined as a continuous succession of events described by a differentiable application

$$\begin{aligned} \vec{r} : \mathbb{R} &\longrightarrow \mathbb{R}^3 \\ t &\longmapsto \vec{r}\,(t) \end{aligned} \ , \tag{2.1}$$

where t is usually identified with the time and $\vec{r}$ is often referred to as motion law. The plot of $\vec{r}$ in $\mathbb{R}^3$ is called trajectory or orbit. Given two points $\vec{r}_1 = (x, y, z)$ and $\vec{r}_2 = (x + \mathrm{d}x, y + \mathrm{d}y, z + \mathrm{d}z)$, the length of $\vec{r}_2 - \vec{r}_1$ constitutes the arc length of an infinitesimal portion of the curve

$$\mathrm{d}s^2 = \mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2 \ , \tag{2.2}$$

and by introducing the velocity $\vec{v}$ as the derivative with respect to t of the trajectory, the arc length can be expressed as

$$\mathrm{d}s = \sqrt{\mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2} = \sqrt{(\mathrm{d}_t x)^2 + (\mathrm{d}_t y)^2 + (\mathrm{d}_t z)^2}\mathrm{d}t = \|\vec{v}\,(t)\|\,\mathrm{d}t \ . \tag{2.3}$$

The kinematic component of the motion can be separated from the purely geometrical one by rewriting (2.1) as

$$\vec{r} = \vec{r}\,(s) \ , \qquad\qquad s = s\,(t) \ . \tag{2.4}$$

The first equation now describes only the trajectory, so it can be used to introduce three unitary orthogonal vectors which constitute the so called Frenet-Serret frame [Tur99]

$$\vec{T}\,(s) \stackrel{\text{def}}{=} \mathrm{d}_s \vec{r}\,(s) \ , \tag{2.5}$$

$$\vec{N}\,(s) \stackrel{\text{def}}{=} \frac{1}{\left\| d_s \vec{T} \right\|} \mathrm{d}_s \vec{T}\,(s) = \frac{1}{\kappa} \mathrm{d}_s \vec{T}\,(s) \ , \tag{2.6}$$

$$\vec{B}\,(s) \stackrel{\text{def}}{=} \vec{T}\,(s) \times \vec{N}\,(s) \ . \tag{2.7}$$

The quantity κ is called curvature and it is the inverse of the curvature radius, i. e. the radius of the circle arc that best approximates the curve at that particular point. It follows that if κ tends to 0, the curve tends to a straight line. The three vectors of the

frame are called respectively tangent, normal and binormal. By using these vectors it is possible to build what is called a Frenet-Serret apparatus via a set of equations that completely characterise the curve at any point in space. These equations are

$$\mathrm{d}_s\vec{T} = \kappa\vec{N} \ , \tag{2.8}$$

$$\mathrm{d}_s\vec{N} = -\kappa\vec{T} + \tau\vec{B} \ , \tag{2.9}$$

$$\mathrm{d}_s\vec{B} = -\tau\vec{N} \ , \tag{2.10}$$

where the quantity τ is called torsion and is defied via the relation (2.10). The torsion describes the characteristic velocity at which the curve is leaving its osculating plane, that is defined as the plane orthogonal to the binormal vector. The preceding equations are called Frenet-Serret equations and the set of $\vec{T}$, $\vec{N}$, $\vec{B}$, κ and τ is called the Frenet-Serret apparatus.

2.2 Implementation and Testing of the Frenet-Serret Apparatus

The analysis of the 3D trajectories presented in the *Data Analysis* chapter was performed via a code that can be found free of license at [Car20b]. The computational part was written in C++, while the graphical part was implemented in Python. In order to test the accuracy of the numerical methods used, a test was conducted on a helix of equations

$$\begin{cases} x = R\cos(t) \\ y = R\sin(t) \\ z = At \end{cases} \ , \tag{2.11}$$

with a radius $R = 3.0$ and a pitch $2\pi A = 2\pi \cdot 0.3$. To check if the behaviour of the code is correct the equation must be firstly parametrised with s such that the velocity becomes unitary. This can be achieved by using

$$s(t) = \sqrt{R^2 + A^2}\ t \ , \tag{2.12}$$

so that

$$\vec{r}(s) = \left(R\cos\left(\frac{s}{\sqrt{R^2+A^2}}\right), R\sin\left(\frac{s}{\sqrt{R^2+A^2}}\right), \frac{As}{\sqrt{R^2+A^2}}\right) \ . \tag{2.13}$$

The tangent can be obtained by direct derivation of the motion law as

$$\vec{T} = \mathrm{d}_s\vec{r} = \frac{1}{\sqrt{R^2+A^2}}\left(-R\sin\left(\frac{s}{\sqrt{R^2+A^2}}\right), R\cos\left(\frac{s}{\sqrt{R^2+A^2}}\right), \frac{A}{\sqrt{R^2+A^2}}\right) \ ,$$

and its norm is coherently

$$\|\mathrm{d}_s\vec{r}\| = \frac{1}{\sqrt{R^2+A^2}}\sqrt{R^2+A^2} = 1 \ .$$

The curvature and the normal vector immediately follow as

$$\mathrm{d}_s\vec{T} = \kappa\vec{N} = \frac{1}{R^2+A^2}\left(-R\cos\left(\frac{s}{\sqrt{R^2+A^2}}\right), -R\sin\left(\frac{s}{\sqrt{R^2+A^2}}\right), 0\right) \ , \tag{2.14}$$

$$\kappa = \left\|\mathrm{d}_s\vec{T}\right\| = \frac{R}{R^2+A^2} \ , \tag{2.15}$$

proving that for a helix the curvature is constant. The binormal vector is obtained from the cross product of $\vec{T}$ and $\vec{N}$ as

$$\vec{B} = \frac{1}{\sqrt{R^2+A^2}}\left(A\sin\left(\frac{s}{\sqrt{R^2+A^2}}\right), -A\cos\left(\frac{s}{\sqrt{R^2+A^2}}\right), R\right) \ , \tag{2.16}$$

from which the torsion

$$\mathrm{d}_s\vec{B} = -\tau\vec{N} = \frac{A}{R^2+A^2}\left(\cos\left(\frac{s}{\sqrt{R^2+A^2}}\right), \sin\left(\frac{s}{\sqrt{R^2+A^2}}\right), 0)\right)$$

$$\implies \tau = \frac{A}{R^2+A^2} \ . \tag{2.17}$$

The last relation found shows that the torsion is also constant for a helix. Using the parameters for the test one finds

$$\kappa = 0.\overline{3300} \ , \qquad\qquad \tau = 0.\overline{0330} \ .$$

Two numeric tests were conducted with different parametrisations. The first aimed at studying an almost continuous case, where the parameter t ranged from 0 to 6π with an increment of 0.0002π every step, for a total of 30000 steps. In this case the values obtained (after being divided by the increment to correctly rescale t) are

$$\kappa = 0.33003300 \ , \qquad\qquad \tau = 0.03300330 \ ,$$

which are perfectly coherent with the theoretical one up to the eighth decimal place. For the second test the parameters used were more resembling of common values found in the actual data, with a chosen velocity of 10[1], and the time ranging from 0 to 10 with an increment of 0.0666667, for a total of 150 steps. This time the values are

$$\kappa = 0.32997114 \ , \qquad\qquad \tau = 0.03300304 \ ,$$

and as expected, these are less precise, but still resemble the theoretical data with a good degree of precision. The helices obtained with the program using these parameters are reported in Fig. 2.1. The results are as expected, with the correct positioning of all unitary vectors on the curve.

[1]Since in this case the scale is arbitrary no unit is reported.

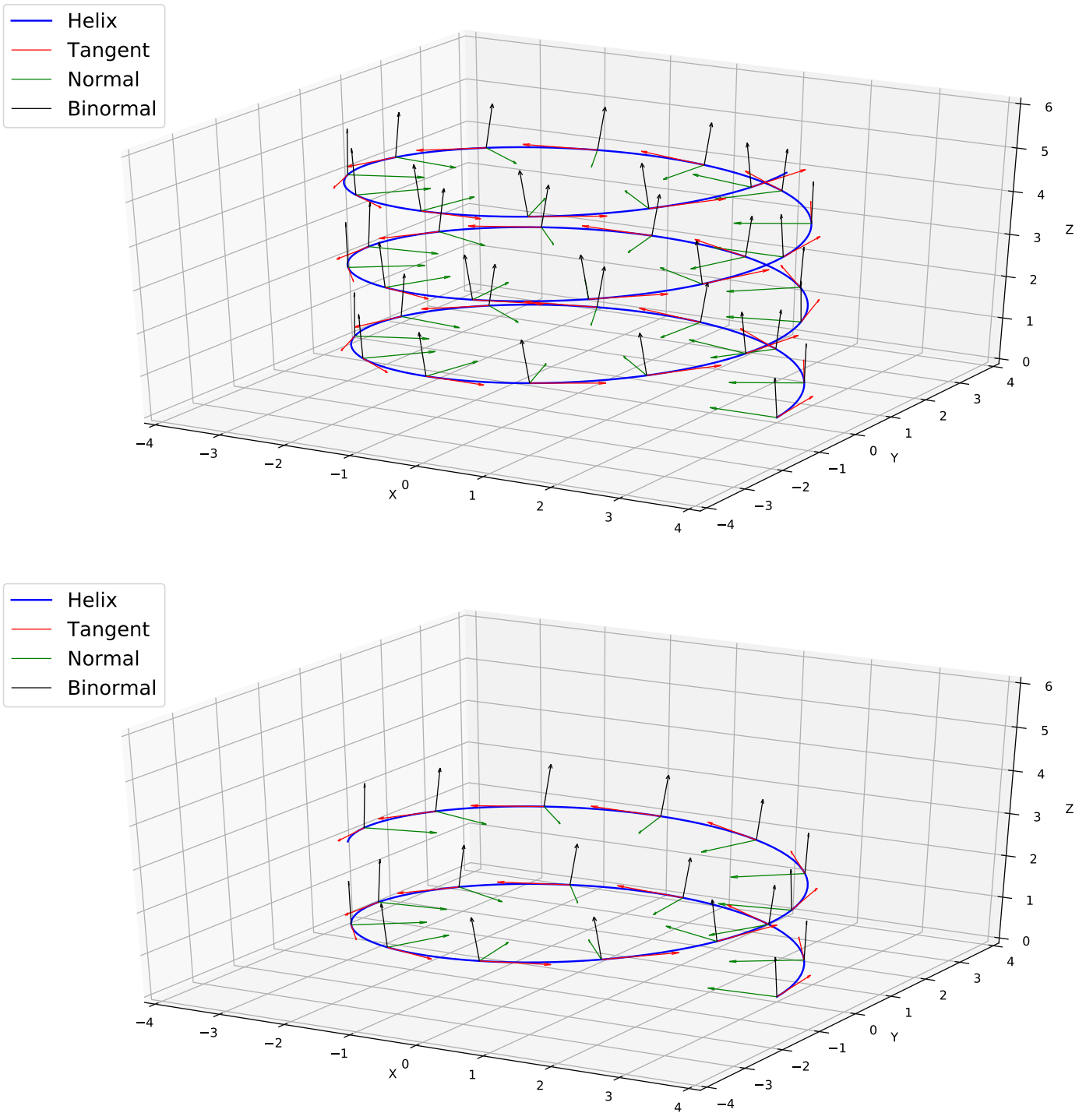


Figure 2.1: Test helices obtained via the code at [Car20b]. Both helices have a radius $R = 3$ and a pitch $2\pi A = 2\pi \cdot 0.3$ so the curvature is $\kappa = 0.\overline{3300}$ and the torsion is $\tau = 0.\overline{0330}$. The helices have different parametrisations. The top one has the parameter $t \in [0, 6\pi]$ with an increment of 0.0002π, for a total of 30000 steps, the bottom one has $t \in [0, 10]$ with an increment of 0.0666667, for a total of 150 steps. In order to maintain legibility only one set of unitary vectors every k points is drawn, for the top plot $k = 800$, for the bottom one $k = 8$. The red unitary vector is the tangent, the green one is the normal and the black one is the binormal.

3. Data Analysis

3.1 Structure of Data

The data analysed in this work consists of four experiments [Pas+18], which will be referred to as E1, E2, E3, E4. Numerous zooplankton samples were collected from the Gulf of Naples in 2008 (for E1, E2) and 2009 (for E3, E4). Healthy adult females of the species *Clausocalanus furcatus* were selected among these and isolated in two groups of 30 and 37 individuals, that were consequently recorded in a 1ℓ aquarium for one hour either in the presence of food or without it. E1 and E3 were recorded in the presence of small food particles, with a density of $\rho_{f1} = 5 \cdot 10^5$ units/ℓ and $\rho_{f3} = 5 \cdot 10^6$ units/ℓ respectively, whereas E2 and E4 were conducted in the absence of food using filtered sea water.
A series of coordinates in the 2-dimensional space were acquired via an automated system of cameras with a spatial resolution of 78 μm at 15 fps. The system can be seen in Fig. 3.1. The merging of the trajectories in 3-D was performed via a C++ code used to combine the simultaneous 2-D tracks by comparing the common z values and discard ghost tracks (non moving objects). All trajectories shorter than 5 s were discarded. Information about the final 3-D trajectories are given in Tab. 3.1.
An example of a track can be seen in Fig. 3.2. This particular track is a segment of an even longer track from $E3$, in which the specimen is moving at a high velocity ($v > 6mm/s$) in a convoluted pattern.

Experiment	Tracks	Average duration (s)	Minimum duration (s)	Maximum duration (s)
E1 (food)	255	23.9	5	241
E2 (no food)	384	28.9	5	342
E3 (food)	625	24.4	5	278
E4 (no food)	425	35.2	5	277

Table 3.1: Number of tracks, average duration, minimum duration and maximum duration of the four experiments E1, E2, E3, E4. E3 and E4 have significantly more data than E2 and, especially, E1, as not only do the other experiments have more tracks, they also have a higher average duration, meaning more points overall.

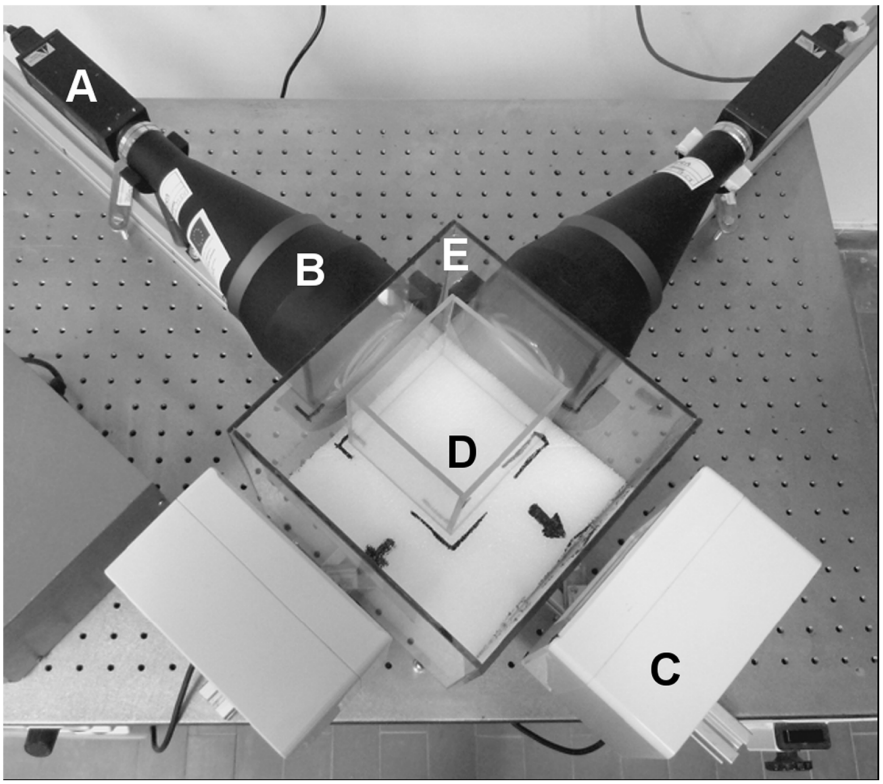


Figure 3.1: An image of the experimental setup taken from [Bia+13]. The labelled components are respectively: a digital camera (A); a telecentric lens (B); an infrared light source (C); a 1 ℓ aquarium containing the samples (D); an empty 8 ℓ aquarium used for protection from external disturbances (E). The sampling rate chosen for the experiments was 15 fps, the spatial resolution of the cameras was 78 μm.

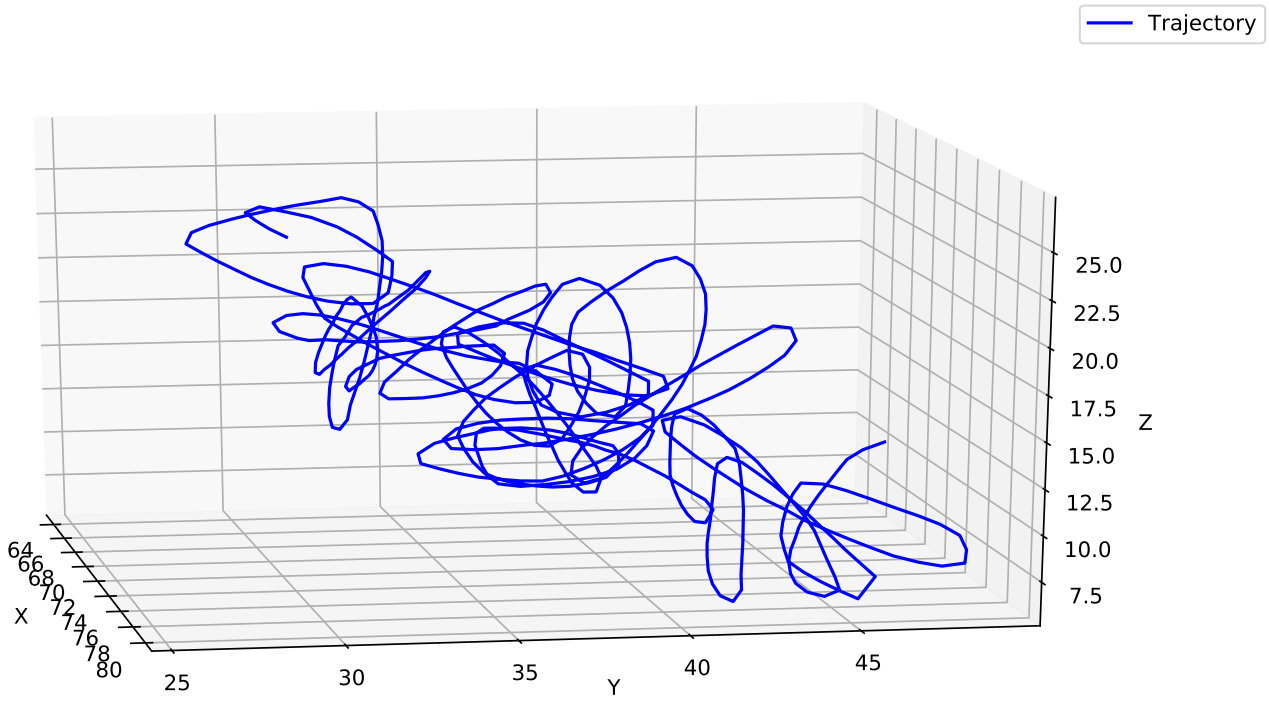


Figure 3.2: A long trajectory taken from E3, the loop structures tend to be quite common along with spiralling ones and quick dives both upwards and downwards.

3.2 Velocities: Fast and Slow Regimes

The first step of the analysis involved the study of the velocities of *C. Furcatus* specimens. The time-step between two points is $\bar{t} = 1/15\ s = 0.0\overline{6}\ s$, so given two sets coordinates $\vec{x}_1, \vec{x}_2 \in \mathbb{R}^3$, the velocity is simply

$$\vec{v}_1 = \frac{\vec{x}_2 - \vec{x}_1}{\bar{t}}\ . \tag{3.1}$$

For each experiment the PDF of the velocity modules can be seen in Figs. 3.3 for the two experiments with food and 3.4 for the ones without. For E2 and E4 the distributions have similar structures, both with the largest peak between 1 mm/s and 2 mm/s and a smaller peak between 2 mm/s and 3 mm/s. Both also have another peak between 5 mm/s and 7 mm/s, however the E2 PDF lacks a fourth peak, which is instead present in the E4 PDF between 9 mm/s and 10 mm/s. An important part of the analysis is to find whether there is a difference in behaviour as velocity increases, therefore a threshold velocity has been defined to distinguish between *fast* and *slow* regime. For E2 and E4 this velocity was chosen at the end of the first peak at $v_{2,4}^{\text{cut}} = 2.1\ mm/s$, and all of the subsequent analysis has been conducted separately between the two regimes to highlight behavioural differences.
The distributions for E1 and E3 showcase more differences, as the E1 PDF has two distinct peaks at slow velocities, one between 1 mm/s and 2 mm/s, the other between 2 mm/s and 3 mm/s, whereas the E3 PDF only manifests the first one. The E1 PDF also has a smaller peak between 5 mm/s and 6 mm/s, and a final and taller peak at 10 mm/s, while the E3 distribution lacks the smaller peak and instead manifests a way larger peak at high velocities between 10 mm/s and 13 mm/s. This last one is notably lower than the peak at slow velocities, but it also has a longer tail than the one of E1, as the PDF for a velocity of 12 mm/s is already below 10^{-3} for E1, while it is still peaking for E3. These anomalies were further investigated by trying to compute the E1 velocities in two alternative ways to see if this would have helped to reproduce a PDF more similar to that of E3.
The first attempt made was to compute a double step velocity, i. e. given three consecutive points $\vec{x}_1, \vec{x}_2, \vec{x}_3 \in \mathbb{R}^3$, the velocity was computed skipping $\vec{x}_2$ using

$$\vec{v}_1 = \frac{\vec{x}_3 - \vec{x}_1}{2\bar{t}}\ , \tag{3.2}$$

however, as it can be seen in the first graph of Fig. 3.5, this only increased the spacing between the two low velocity peaks with little effect on the width of the fastest peak. The second attempt was made using a running average procedure to smooth the peaks, i. e. given $2N + 1$ consecutive velocities $\vec{v}_{-N}, ..., \vec{v}_0, ..., \vec{v}_N \in \mathbb{R}^3$ the averaged velocity

was computed as

$$\|\vec{v}\|_0^{avg} = \frac{1}{2N+1} \sum_{i=-N}^{N} \|\vec{v}_i\| \ , \tag{3.3}$$

where $\|\vec{v}\|$ is the 3-dimensional euclidean norm of the vector. The second graph in Fig. 3.5 shows the averaged PDF for $N = 2$, and as expected this smoothed the peaks, however without any significant effect on making the overall shape of the distribution look more similar to the E3 one. Several other values for N were used, up to $N = 7$, without any significant improvement, therefore this route was abandoned.
Considering the tests made, the threshold velocity for E1 and E3 was chosen at $v_{1,3}^{\text{cut}} = 6\ mm/s$, the minimum value between the third and fourth peaks of E1 and far enough from the two peaks of E3. This value was chosen to clearly divide the final peak from the previous ones, and although it is not the minimum for $E3f$ it is close enough to it with a low value for the PDF.
After identifying the threshold velocities, every track in all the experiments was further divided based on the two motion regimes found. Each track was divided in sub-tracks based on the regime change, with the switch from slow to fast (or vice versa) registered if the animal changed its behaviour for at least 4 frames, i. e. $0.27s$. Information about the new trajectories are reported in Tab. 3.2. The fast regime will be referred to with an "f" appended to the experiment name, the slow regime with an "s".

Experiment	Tracks	Average duration (s)	Minimum duration (s)	Maximum duration (s)
E1 (slow, food)	371	8.4	0.27	200.9
E1 (fast, food)	377	7.8	0.27	240.7
E2 (slow, no food)	1716	4.4	0.27	18.9
E2 (fast, no food)	1757	2.0	0.27	24.5
E3 (slow, food)	1664	3.6	0.27	47.5
E3 (fast, food)	1978	4.6	0.27	176.1
E4 (slow, no food)	2750	3.4	0.27	14.0
E4 (fast, no food)	2849	2.0	0.27	120.7

Table 3.2: Number of tracks, average duration, minimum duration and maximum duration of the four experiments after dividing the tracks into the fast and slow regimes. It should be noted that all experiments had at least some tracks which started or ended with 1 to 3 points of a particular regime, hence the actual minimum duration was of only 1 frame despite the division of the track on 4 frames minimum. These cases were however not significant and excluded from the analysis, as they were always less than 3.5% of the total number of tracks.

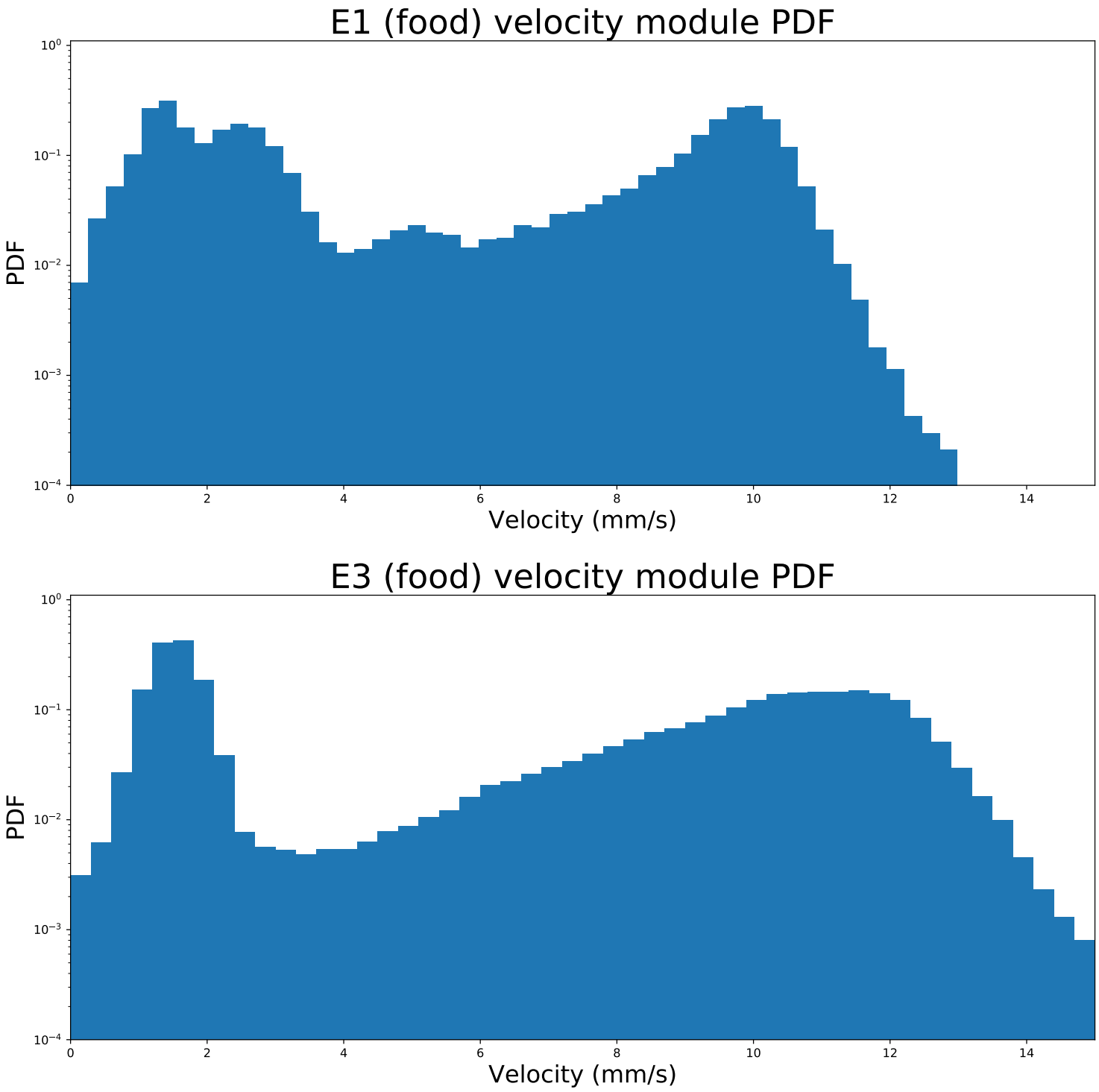


Figure 3.3: Velocity module logarithmic scale PDFs for E1 (top) and E3 (bottom), conducted in the presence of food. As the overall shapes of the distributions appear different, the subsequent analysis was conducted dividing the velocities in slow regime and fast regime with the threshold chosen at $v_{1,3}^{\text{cut}} = 6mm/s$ to isolate the final peak from the rest of the data and to clearly distinguish between faster tracks and slower tracks.

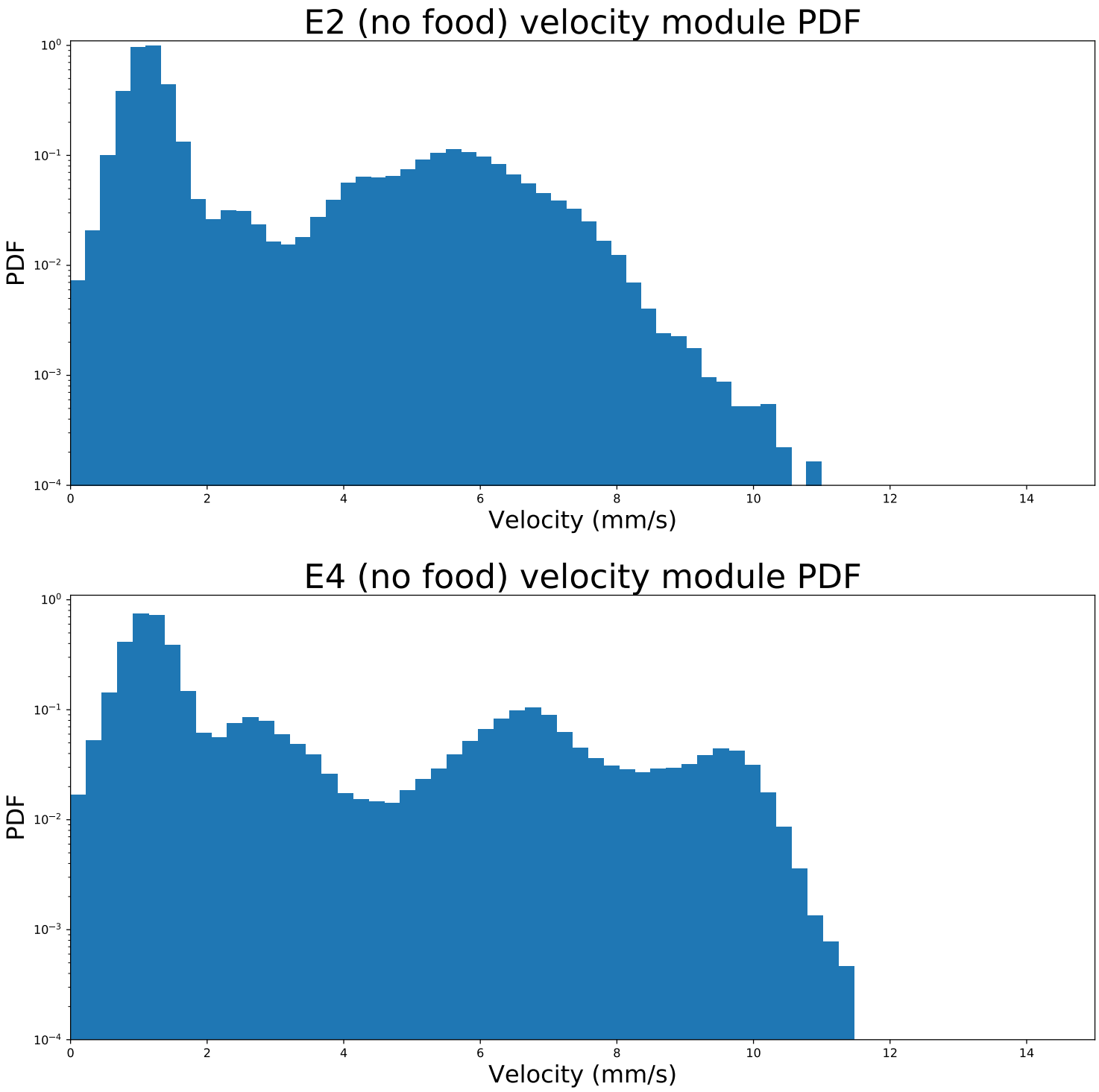


Figure 3.4: Velocity module logarithmic scale PDFs for E2 (top) and E4 (bottom), conducted in the absence of food. The overall shapes of the distributions appear similar for the first part, while the E4 PDF shows a final peak that the one from E2 lacks. The threshold velocity to divide in fast and slow regime was chosen at $v^{\text{cut}_{2,4}} = 2.1mm/s$ as it is the minimum after the first (and tallest) peak and the only minimum the two distributions have in common.

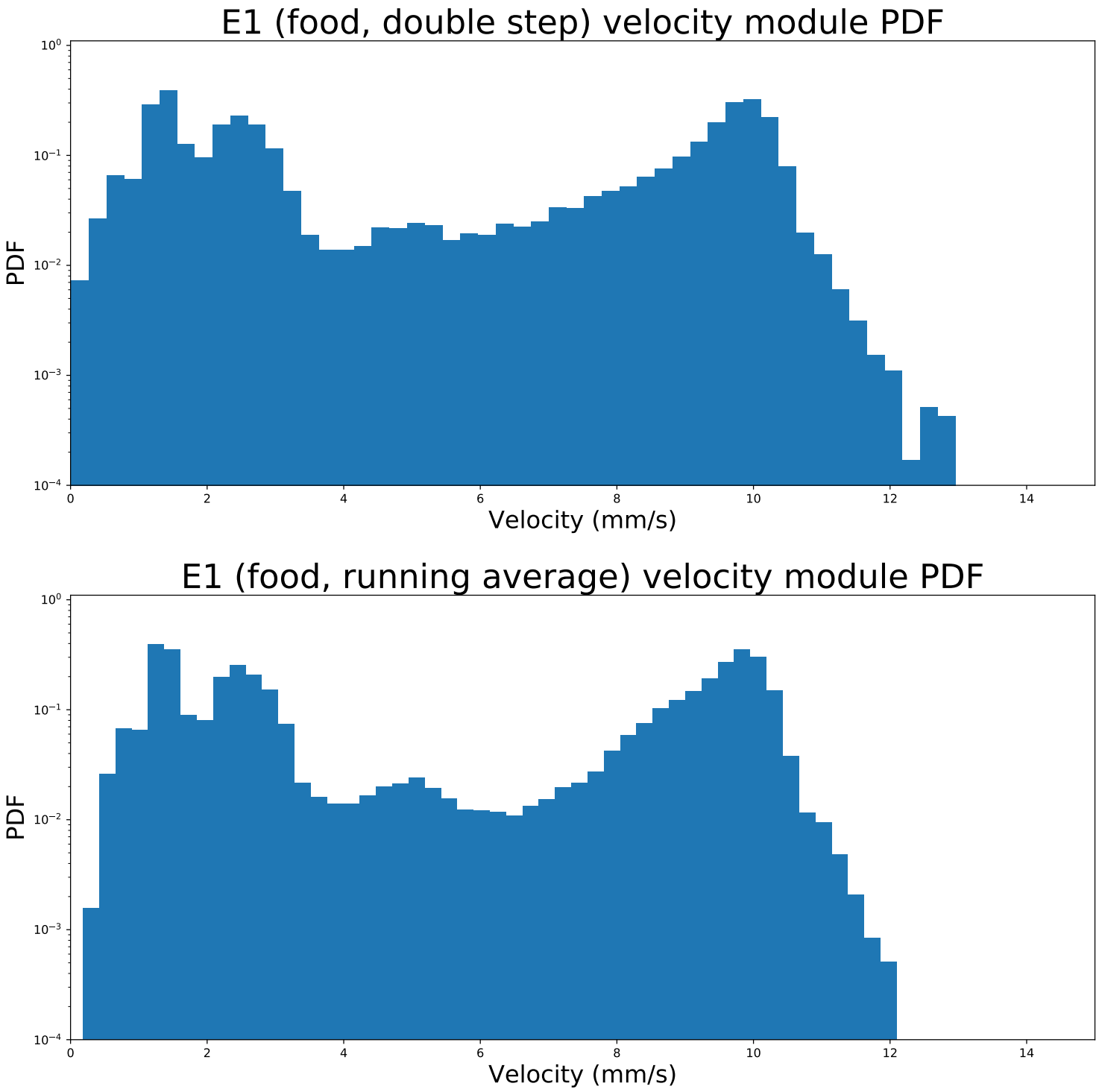


Figure 3.5: Velocity module logarithmic scale PDFs for E1 using alternative ways to compute velocities: a double step velocity (top) and a 2-step running average (bottom). These tests were aimed at reproducing from E1 a PDF more resembling to that of E3 to explain the differences in distributions, but they were not successful, so ultimately the analysis was conducted on the unaltered PDF in Fig. 3.3.

3.3 Individual Mean Velocities

After dividing the data, the first step in the analysis was aimed at understanding if the behaviour of the individual could be used to describe that of the population. In order to do so, the mean velocity of every specimen was computed over each track before finding the PDF. Fig. 3.6 shows the PDFs found for the mean velocity of each individual, which have been compared to the previous PDFs in Figs. 3.3, 3.4.
For the slow regime E1s and E2s show some consistent behaviour with the previous ones as the peaks of the distributions fall in the same intervals, but the height of the peaks have been completely altered. The E1s main peak has increased in height from 0.4 to 1, while the one in E2s rose from 1 to 3. E3s shows a shape similar to that of the previous PDF, but the peak has also increased from 0.4 to 1. In E4s the peak has also risen from 0.9 to 3 and there is a non-neglectable portion of specimens that appear to be almost still, while a small portion of velocities localised between 0.6 mm/s and 0.8 mm/s appears to be missing. This behaviour is totally lacking in Fig. 3.4. In this regime the hypothesis of using the individuals to identify the sample does not appear to be consistent.
The fast regime shows some notable changes in the structure of the PDFs. E1f in this case shows two smaller yet distinct peaks centred at 8 mm/s and 9.5 mm/s, while it manifested a single peak centred at 10 mm/s on the original PDF. The first peak is also higher than the original peak, from 0.3 to 0.5. E2f retains most of its features but it has its peak increased from 0.1 to 0.6. The large peak from 10 mm/s to 12 mm/s in E3f has disappeared and has been replaced by a large peak ranging from 8 mm/s to 10 mm/s, depriving the PDF of one of its most notable features. E4f loses its smaller peaks around 3 mm/s and 10 mm/s, while the one centred between 6 mm/s and 7 mm/s is retained and has increased in height from about 0.1 to about 0.25. These distributions are completely different to those in Figs. 3.3, 3.4.
These considerations show that the analysis of both regimes can not be carried out on the individuals, but instead requires a large population to be effective and fully characterise the behaviour of the specimens.
The scatter plots for the mean velocity of each individual as a function of the track length can be seen in Fig. 3.7, and these clearly visualise how E1 has less statistics than the other experiments. An interesting phenomenon is how the clusters of points tend to be horizontally arranged for the slow regime, whereas those from the fast regime tend to be vertically arranged. Slow tracks have therefore less deviation from the mean values of the PDFs just discussed, whereas fast tracks are much more diverse and tend to be clustered based on their duration instead of specific velocity values. This kind of clustering near short tracks for the fast regime suggests a motion based on short burst of speed. During these bursts the specimens tend to maintain a constant speed, so various mean values

are possible for the velocity given the high variability possible. On the contrary, for the slow regime, the tracks tend to be longer and therefore the mean velocity tends to be evened towards specific points, meaning a less erratic motion.
The scatter plots in the fast regime manifest longer tails for various average velocities. The presence of longer jumps is perfectly compatible with a Lévy flight-like motion, and shows that when individuals relocate to a new area they tend to do it at high velocities. The jumps also tend to be more frequent when food is present, as it can be clearly seen by the plot for E3f, where the frequency of longer jumps is much higher than that of E4f. E1f also presents a tail, but its scarcity of trajectories is even more evident in this plot. Even so its tail shows a number of points comparable with that of E2f, so it is not unreasonable to suppose a similar behaviour. In the presence of food the creatures may be incentivised to explore more space, while this costly (in terms of energy) activity may not be encouraged in an environment where food is scarce, paving way for the suggestion of energy preservation strategies. This difference between space exploration, track duration and distance travelled will be the main topic of a following analysis.

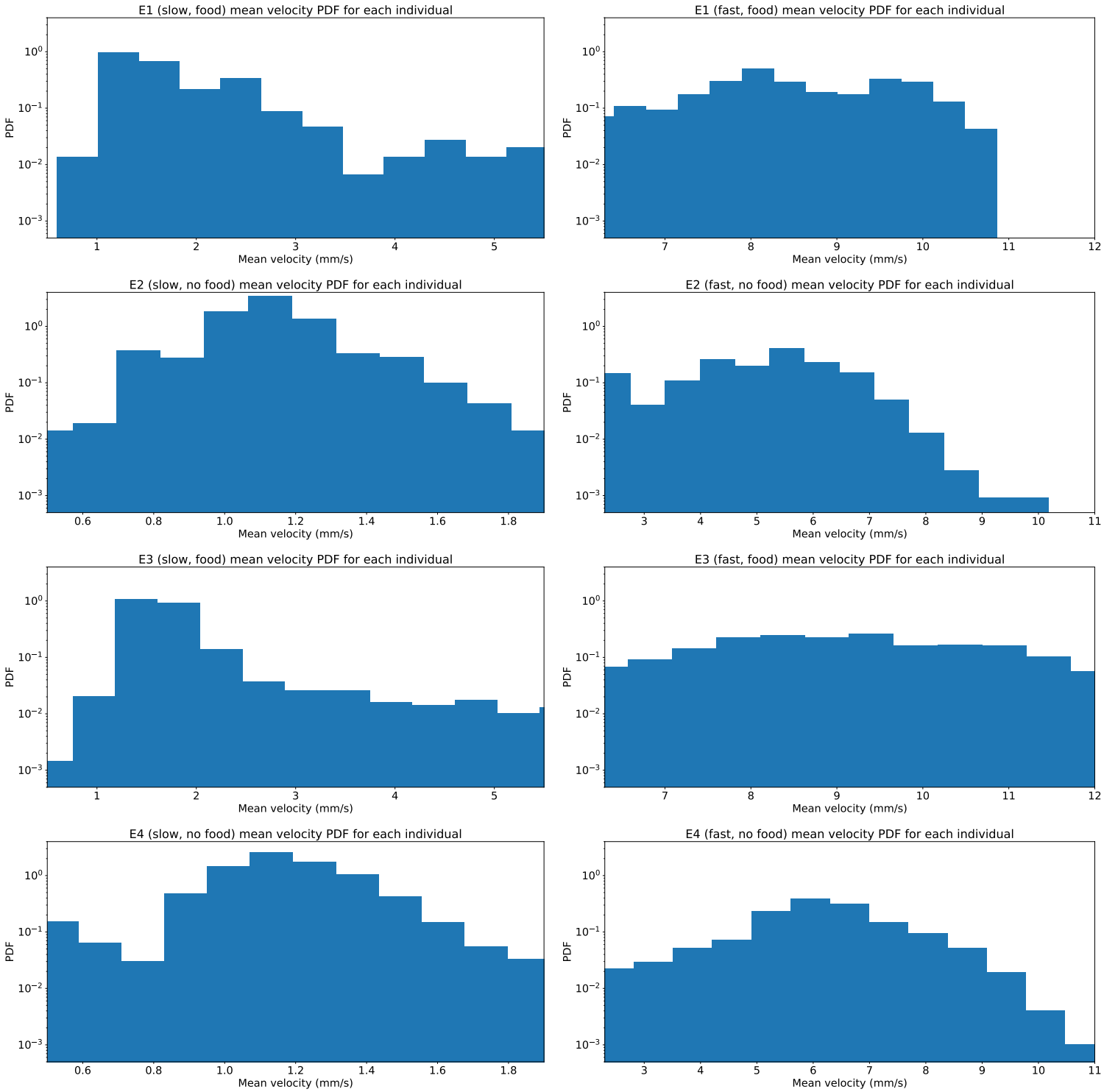
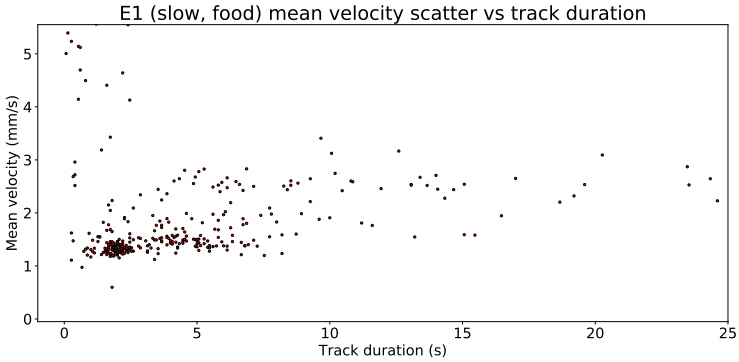
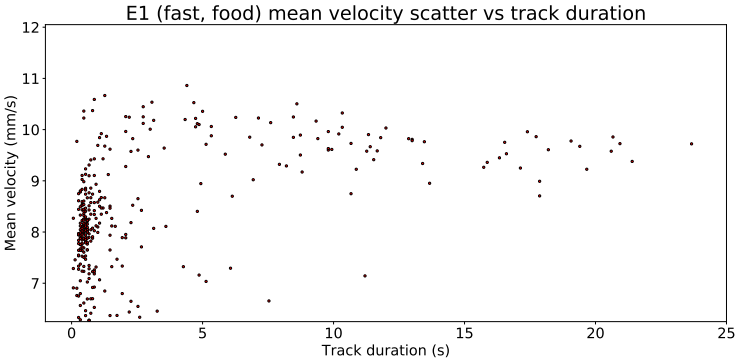
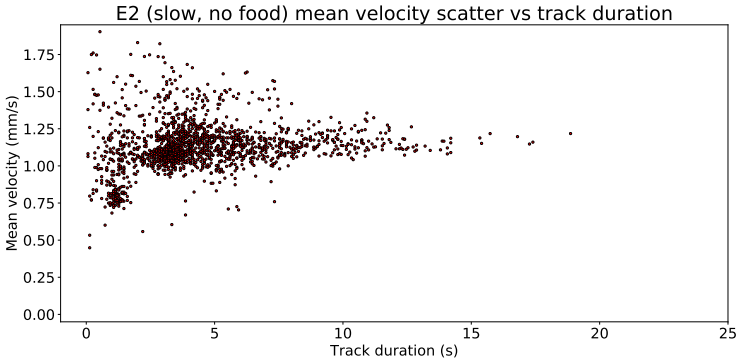
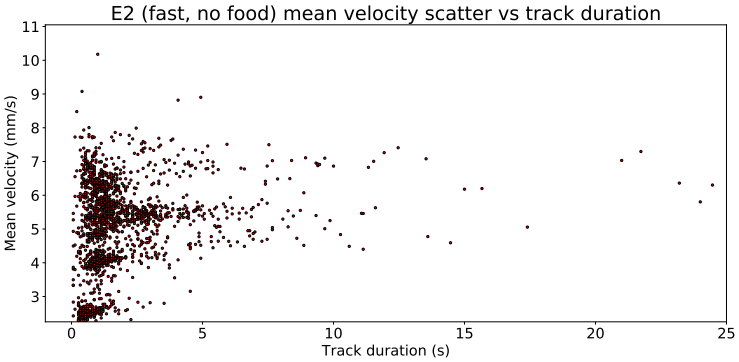


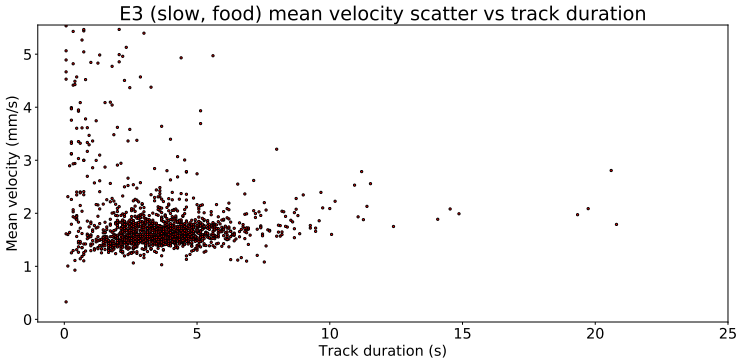
Figure 3.6: Individual mean velocity logarithmic *PDF*s for slow regime (left) and fast regime (right) of all experiments (ordered top to bottom). These distributions describe the mean velocity of a single individual over its entire track, as opposed to those in Figs. 3.3, 3.4 where the *PDF*s are constructed over all velocities for all specimens. By analysing the differences between these *PDF*s and the aforementioned ones it is possible to deduce that studying the behaviour of the individual is not indicative of the behaviour of the population, as many key points from the distributions differ, creating two totally different motion regimes.

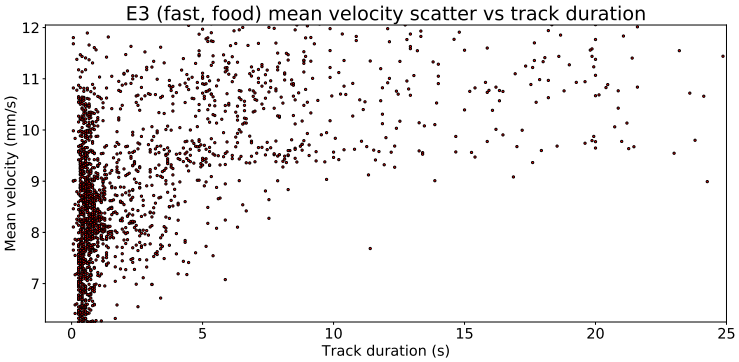


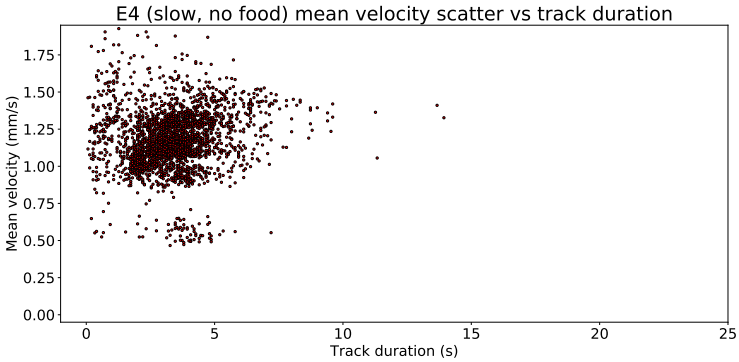












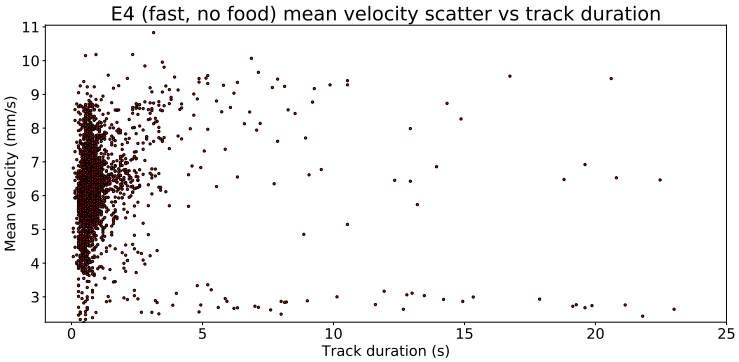


Figure 3.7: Scatter plot for mean velocity of an individual over its entire trajectory versus the corresponding track duration for slow regime (left) and fast regime (right) of all experiments (ordered top to bottom). In the slow regime the velocities tend to be clustered around specific points, while in the fast regime this is not the case and instead the velocities tend to form vertical lines, suggesting that the duration of the track may be less dependent on the mean velocity chosen. This type of behaviour may be explained by a motion based on longer and more consistent tracks for the slow regime, against a more erratic behaviour made of short bursts of almost constant speed in the fast regime. The fast regime also presents longer tails, which are compatible with jumps from a Lévy flight model, showing that longer jumps, albeit rarer, are possible.

3.4 Track Durations and Track Lengths

The next step of the analysis is a classification of the distribution of the tracks durations and lengths, where the hypothesis is that of a power law decay for both magnitudes in the fast regime, and of an exponential decay for the slow regime. The choice of these functions is justified by the proposed model as follows: for the slow regime the use of an exponential suggest energy consumption, meaning that in this regime the creatures tend to remain for shorter times and travel shorter distances, suggesting a transitory regime of rest; for the fast regime the power-law follows the Lévy flight model, suggesting efficient space exploration to identify points of interest.
Assuming a Poisson error on the bin content for the normalised histograms, the results of the curve fitting were tested using the Kolmogorov-Smirnov (KS) test. The symbol D is used to identify the statistic, while the symbol p is used to identify the confidence level (these are explained and expanded upon in appendix A.1). The results of the analysis are reported in Tabs. 3.3 for for the track durations and 3.4 for the track lengths, while the PDFs and the curve fits can be found respectively in Figs. 3.8, 3.9.
For the slow regime the PDFs tend to have, both in track duration and length, a small rise before starting an exponential fall. Regarding track duration all experiments share similar β coefficients for the exponential decay and have a satisfactory KS probability, which means that the null hypothesis can not be rejected with a confidence level of over 18.6%. For the track lengths E1s and E2s show less convincing confidence levels, however this may be due to the lower data quantity, as E3s and E4s show instead higher confidence levels for the curve fits. This shows that long tracks (both in time and space) at slow velocities are extremely rare, confirming the idea of a transitory regime. The initial rise however may be evidence that once the specimen decides to slow down it is not useful to instantly start a new fast trajectory, allowing instead some time to pass.
For the fast regime, regarding track durations, the α coefficients are substantially different between the food experiments and the no food experiments, with the latter being larger (in absolute terms) indicating that trajectories tend to last longer in the presence of food. The confidence levels for the fit parameters are very high for the experiments without food, whereas they are lower for the experiments with food, most notably E3f, where the confidence is only 63%. This trend is repeated in the case of the α coefficients for the tracks lengths, where once again E3f show a confidence level of only 58%. The other p values are instead above 90%, even for E1f. Most notably, the α coefficients for track durations of the four experiments are compatible with the corresponding ones for the track lengths within the error. From a behaviour standpoint this confirms that longer jumps (both in space and time) are possible, and are more frequent when food is present, once again suggesting energy conservation strategies when food is lacking, while promoting exploration when food is commonly found.

Experiment	Fit	Fit parameter	KS test p	KS test D
E1s	$y = C\exp\left[\beta\left(t - t_0\right)\right]$	$\beta = (-0.31 \pm 0.02)\, s^{-1}$	0.946	0.153
E1f	$y = C\left(t - t_0\right)^{\alpha}$	$\alpha = -1.3 \pm 0.3$	0.856	0.178
E2s	$y = C\exp\left[\beta\left(t - t_0\right)\right]$	$\beta = (-0.44 \pm 0.04)\, s^{-1}$	0.932	0.143
E2f	$y = C\left(t - t_0\right)^{\alpha}$	$\alpha = -2.9 \pm 0.3$	0.996	0.098
E3s	$y = C\exp\left[\beta\left(t - t_0\right)\right]$	$\beta = (-0.65 \pm 0.09)\, s^{-1}$	0.816	0.222
E3f	$y = C\left(t - t_0\right)^{\alpha}$	$\alpha = -0.96 \pm 0.11$	0.626	0.205
E4s	$y = C\exp\left[\beta\left(t - t_0\right)\right]$	$\beta = (-0.79 \pm 0.06)\, s^{-1}$	0.814	0.222
E4f	$y = C\left(t - t_0\right)^{\alpha}$	$\alpha = -1.72 \pm 0.14$	0.975	0.115

Table 3.3: Type of curve fit, parameters and KS test coefficients for track duration of each experiment, divided by regime. The functions chosen for the curve fits were exponential and power-law decay. The first one was applied to the slow regime to check compatibility with an energy consumption model, while the latter was applied to the fast regime to check compatibility with a Lévy flight jump model. This model could suggest search patterns in conformity with the foraging hypothesis in the case of higher velocities, while lower velocities can be considered a transitory regime. The C parameter is just a normalisation constant, the t_0 parameter is in the range $[-0.8, 0.8]\, s$ for the fast experiments and $[-4, 1.5]\, s$ for the slow experiments, but its covariance could not be safely estimated so it is not reported.

Experiment	Fit	Fit parameter	KS test p	KS test D
E1s	$y = C\exp\left[\beta\left(x - x_0\right)\right]$	$\beta = (-0.21 \pm 0.02)\, mm^{-1}$	0.707	0.231
E1f	$y = C\left(x - x_0\right)^{\alpha}$	$\alpha = -1.2 \pm 0.2$	0.986	0.145
E2s	$y = C\exp\left[\beta\left(x - x_0\right)\right]$	$\beta = (-0.24 \pm 0.03)\, mm^{-1}$	0.741	0.239
E2f	$y = C\left(x - x_0\right)^{\alpha}$	$\alpha = -2.7 \pm 0.3$	0.925	0.153
E3s	$y = C\exp\left[\beta\left(x - x_0\right)\right]$	$\beta = (-0.37 \pm 0.07)\, mm^{-1}$	0.863	0.268
E3f	$y = C\left(x - x_0\right)^{\alpha}$	$\alpha = -0.78 \pm 0.12$	0.576	0.192
E4s	$y = C\exp\left[\beta\left(x - x_0\right)\right]$	$\beta = (-0.57 \pm 0.03)\, mm^{-1}$	0.965	0.199
E4f	$y = C\left(x - x_0\right)^{\alpha}$	$\alpha = -1.65 \pm 0.12$	0.986	0.145

Table 3.4: Type of curve fit, parameters and KS test coefficients for track length of each experiment, divided by regime. These results are compatible with those found in Tab. 3.3, suggesting that track duration and track length are regulated by the same underlying laws. Notably. the α coefficients are compatible within the error for track length and duration in each experiment. The C parameter is just a normalisation constant, the x_0 parameter is in the range $[-5, 5]\, mm$ for the fast experiments and $[-2, 0.5]\, mm$ for the slow experiments, but its covariance could not be safely estimated so it is not reported.

Several attempts were made to compute the hazard functions for the track durations in the fast regime with the relation (1.25), however the results were inconclusive as the PDFs and CDFs appear to be too fragmented to produce relevant results. Attempts to use more refined bins however increased fluctuations too much, producing unusable PDFs.

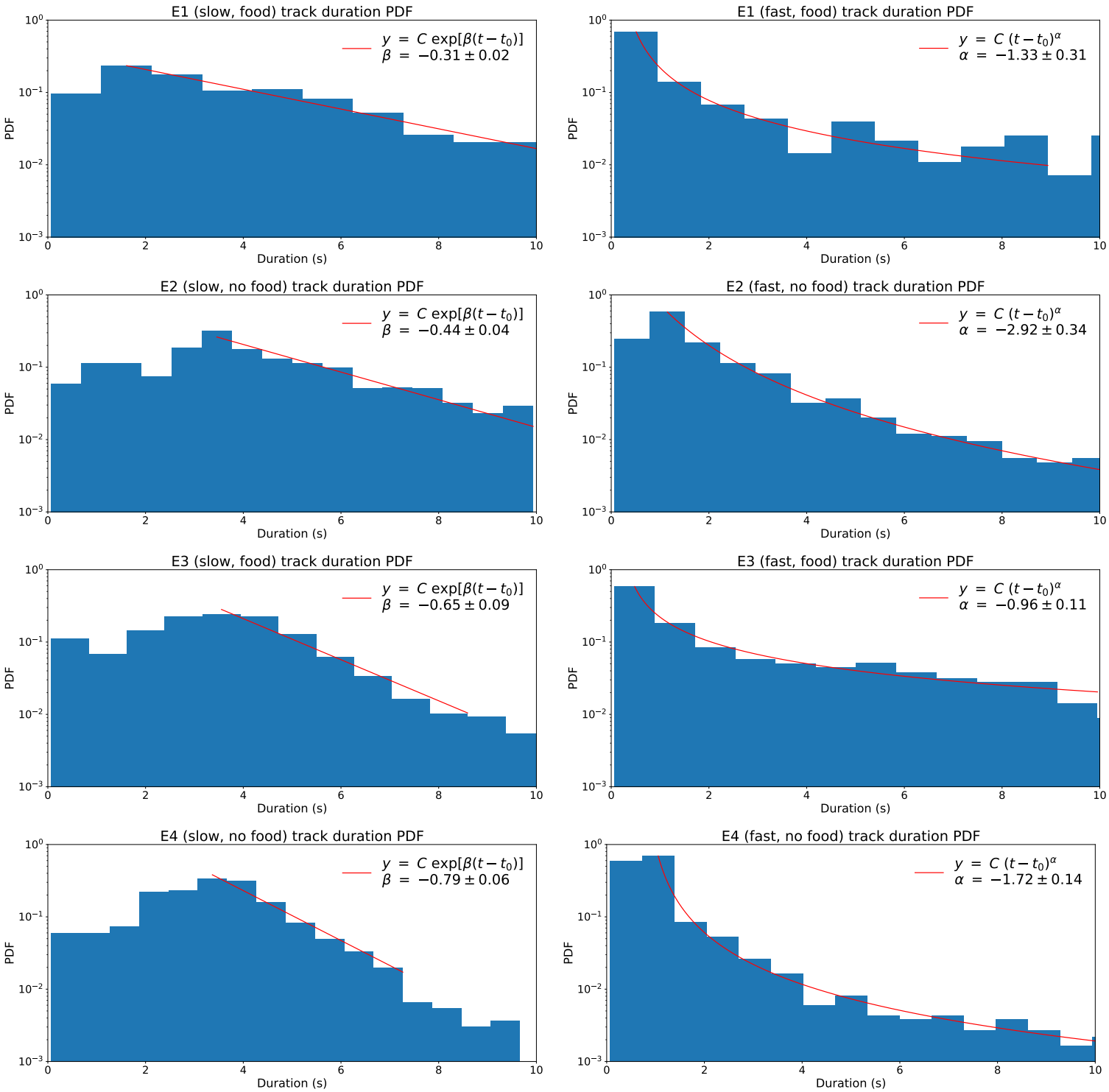


Figure 3.8: Track duration logarithmic PDFs with their respective curve fits for slow regime (left) and fast regime (right) of all experiments (ordered top to bottom). The legend shows the curve used and reports the main parameter. Each fit was prolonged from the start of the main fall in the PDF to the point where the fluctuations became too relevant. The exponential and power-law PDFs were used to see compatibility with a model of short travels for the slow regime and Lévy jumps for the fast regime.

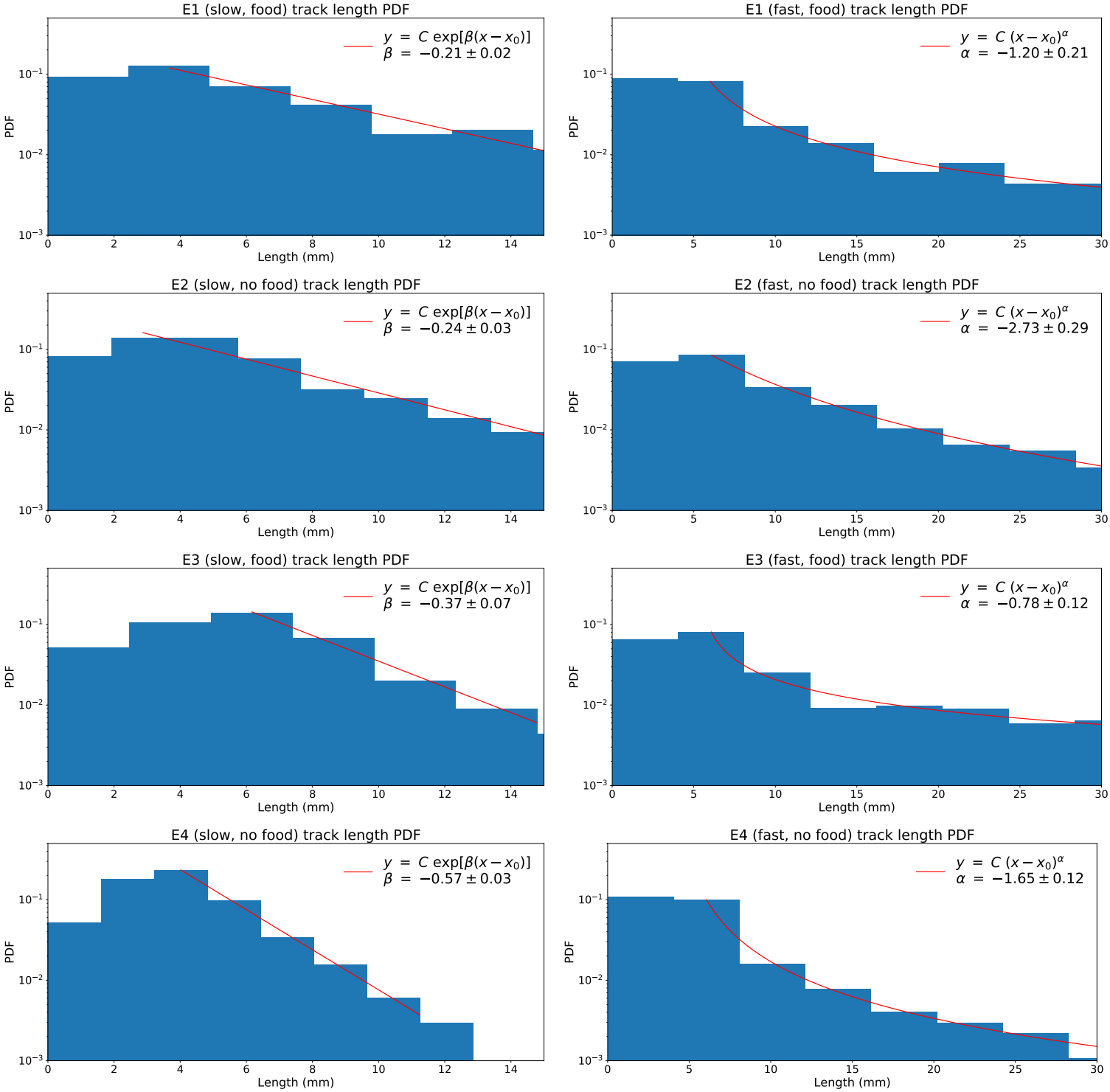


Figure 3.9: Track length logarithmic PDFs with their respective curve fits for slow regime (left) and fast regime (right) of all experiments (ordered top to bottom). The legend shows the curve used and reports the main parameter. Each fit was prolonged from the start of the main fall in the PDF to the point where the fluctuations became too relevant. These PDFs show consistent behaviours with the ones that can be found in Fig. 3.8, suggesting that the model of short trips for the slow regime and Lévy jumps for the fast regime may explain the behaviour of the specimens as a series of alternating active exploration phases (power-law decay) and transitory slow phases (exponential decay).

3.5 Displacements

The next step of the analysis was to characterise how far from the origin the specimens travel while moving. Displacement for a track is defined as the euclidean distance

$$X(T) = \|\vec{x}(T) - \vec{x}_0\| \tag{3.4}$$

between the considered point at a fixed time $\vec{x}(t = T)$ and the starting point $\vec{x}(t = 0) \equiv \vec{x}_0$. The PDFs for displacements at the end of each track are shown in Fig. 3.10. The slow regime distributions are not particularly interesting and are only reported to remark the expected fact that the specimens tend to travel less distance when going slower, as probabilities fall quickly to 0 after 10 mm for all experiments. This is an ulterior confirmation of the fact that the slow regime is just transitory.
The fast regime, on the other hand, shows rather interesting properties as the decays for the displacement PDFs appear to be exponential, as shown in Tab. 3.5. These curve fits show high confidence levels in the KS tests, especially for E3f and E4f, which are statistically more relevant. The fact that the fall of the displacement is exponential (as opposed to the power-law fall of the travelled distance) suggests that the trajectories cannot be considered straight in space, as the creatures do not just swim away from the origin in straight lines. This suggests a high interest in exploring space at high velocities, meaning the creatures are actively looking for points of interest (food, mates, ...).

Experiment	Fit	Fit parameter (mm^{-1})	KS test p	KS test D
E1f	$y = C\exp[\beta(x - x_0)]$	$\beta = -0.19 \pm 0.02$	0.823	0.177
E2f	$y = C\exp[\beta(x - x_0)]$	$\beta = -0.23 \pm 0.02$	0.878	0.182
E3f	$y = C\exp[\beta(x - x_0)]$	$\beta = -0.20 \pm 0.01$	0.990	0.134
E4f	$y = C\exp[\beta(x - x_0)]$	$\beta = -0.39 \pm 0.04$	0.988	0.164

Table 3.5: Type of curve fit, parameters and KS coefficients for Displacement in the fast regime of each experiment. The C parameter is just a normalisation constant, the x_0 parameter is in the range $[-3.3, -0.8]\,mm$, but its covariance could not be safely estimated so it is not reported.

In order to better explore this phenomenon another analysed quantity was the mean displacement, i. e. instead of computing the displacement at the end of each trajectory, this time the mean was computed over all possible trajectories as

$$\langle X \rangle (T) = \frac{1}{N} \sum_{i=1}^{N} X_i(T) \ , \tag{3.5}$$

where N is the number of tracks with a duration greater or equal to T and the index i identifies the displacement at T for the i-th track. This of course meant that the set of available data from long enough tracks decreased rapidly as T increased, as is shown by the yellow curves in Fig. 3.11, representing the percentage of tracks that were long enough to be counted at that specific time. These curves for the fast regime can be interpreted as the complementary CDF for the track duration, i. e. $\overline{P} = 1 - P$, where P is the CDF.
The mean displacement (red curve in the same figure) shows interesting progressions that need further comments. Independently of the regime the first portion of the mean displacement follows a linear law with parameters reported in Tab. 3.6. These have extremely high confidence levels and the A coefficient of the linear law gives the mean velocity of the specimens for the first portion of the track. The linear portion is really short for E1s compared to the other three experiments, but again this may be due to the lower quantity of available data. Even so, E1s and E3s show mean velocities that, although not completely compatible, are quite similar. The mean velocities in E2s and E4s, which are instead compatible, are lower than those of the two food experiments. This slightly faster mean velocity could be explained by the presence of food in E1 and E3.

Experiment	Fit	Fit parameter (mm/s)	KS test p	KS test D
E1s	$y = Ax + B$	$A = 1.361 \pm 0.006$	1.000	0.125
E1f		$A = 6.75 \pm 0.20$	1.000	0.044
E2s	$y = Ax + B$	$A = 1.089 \pm 0.001$	1.000	0.043
E2f		$A = 4.79 \pm 0.03$	1.000	0.013
E3s	$y = Ax + B$	$A = 1.345 \pm 0.005$	1.000	0.038
E3f		$A = 7.56 \pm 0.21$	1.000	0.038
E4s	$y = Ax + B$	$A = 1.17 \pm 0.07$	1.000	0.067
E4f		$A = 5.94 \pm 0.08$	1.000	0.045

Table 3.6: Mean displacement linear fit angular parameter for all experiments. The KS test shows really good confidence levels for the coefficient A, that represents the mean velocity of the specimens in the first portion of the track. The B parameter is compatible with $0mm$ for all experiments.

Notably E2s, E3s and E4s have long lasting linear phases for the mean displacement. The data is displayed until the time at which the sample of available trajectories falls under 5% of the total, and while in E3s the relation ceases being linear at 12% of the total data, for E2s and E4s it stops right as the noise given by the single trajectories becomes too relevant, suggesting that the linear law may in fact continue if it wasn't for the lack of data.
The fast regime also has systematically higher mean velocities for the experiments with

food, but the linear law also stops being the correct description of the system way sooner than the corresponding slow regimes, as the line stops at 47% of the total data for E1f and at 60% for E3f, while it stops at 38% for E2f and 27% for E4f. The actual time duration of the phases appears to be similar for experiments with food and without, with a duration between 1 s and 2 s for all experiments. This suggests that when going slow the specimens tend to travel in straight trajectories, whereas the movement becomes more convoluted when going fast, meaning a more confident space exploration.
In the same figure the blue lines represent the mean travelled distance between each pair of points, that tends to remain constant until the moment where the fluctuations become too relevant for the lack of tracks. This quantity is of course systematically higher for the fast regime, and it is also higher in the experiments where food is present. As the mean travelled distance remains approximately constant over time, this confirms that the displacement is not slowed down by the specimen diminishing its velocity, but by other effects. These effects should mainly be imputable either to the specimen having to change direction due to the physical barrier of the aquarium, or to the specimen deliberately altering its course to explore the environment. These last ones will be expanded upon in the 3-*Dimensional Trajectories* section, where the analysis will be centred on space exploration.

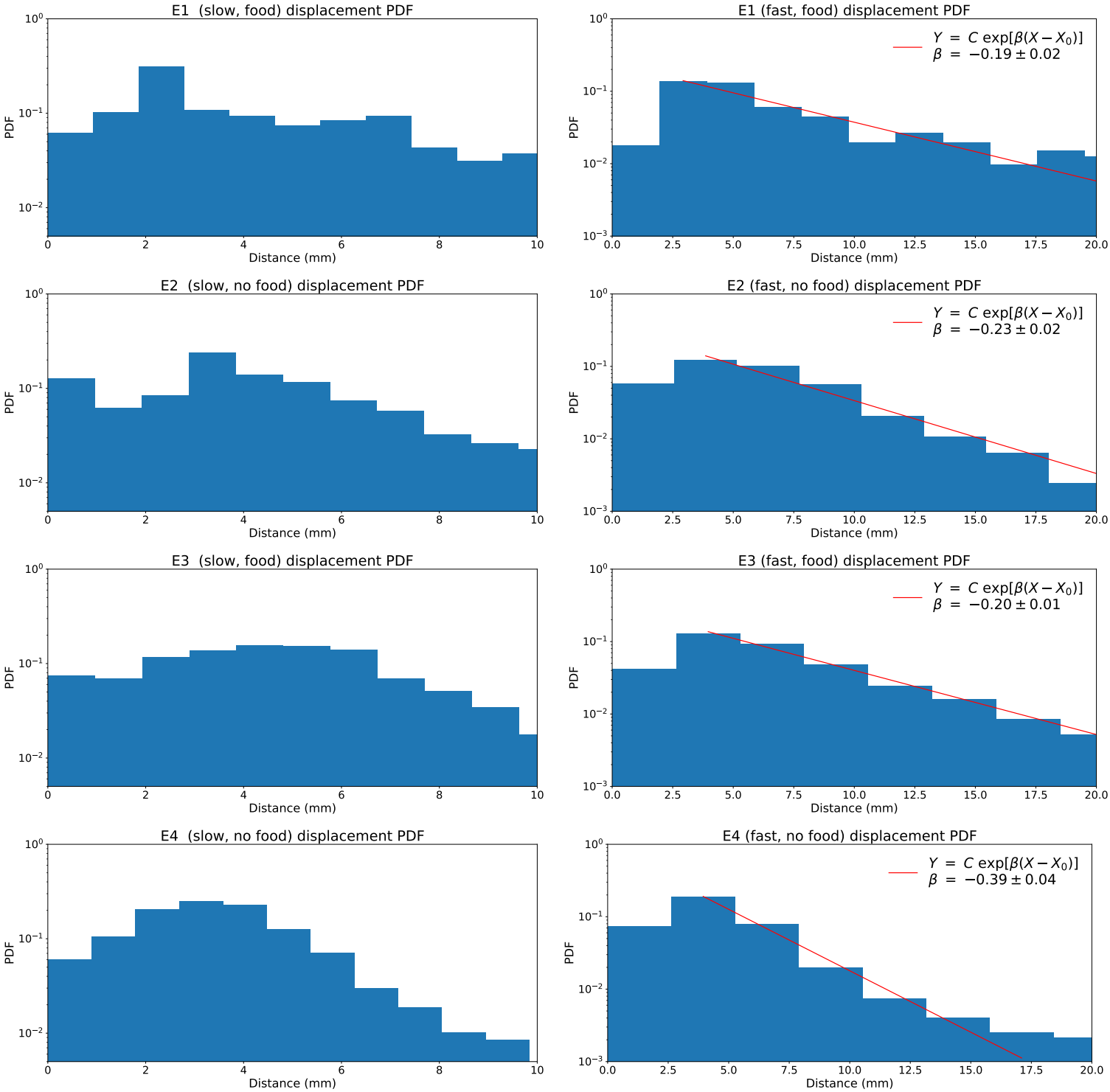


Figure 3.10: Logarithmic *PDF*s for displacements at the end of the tracks with their respective curve fits in slow regime (left) and fast regime (right) for all experiments (ordered top to bottom). For the fast regime the legend shows the used curve and reports the main parameter. The exponential decay of the displacement, when compared to the power-law decay of the track length found in Fig. 3.9, suggests that the specimens do not just travel in straight lines at high velocities but instead tend to explore their environment looking for food or mates.

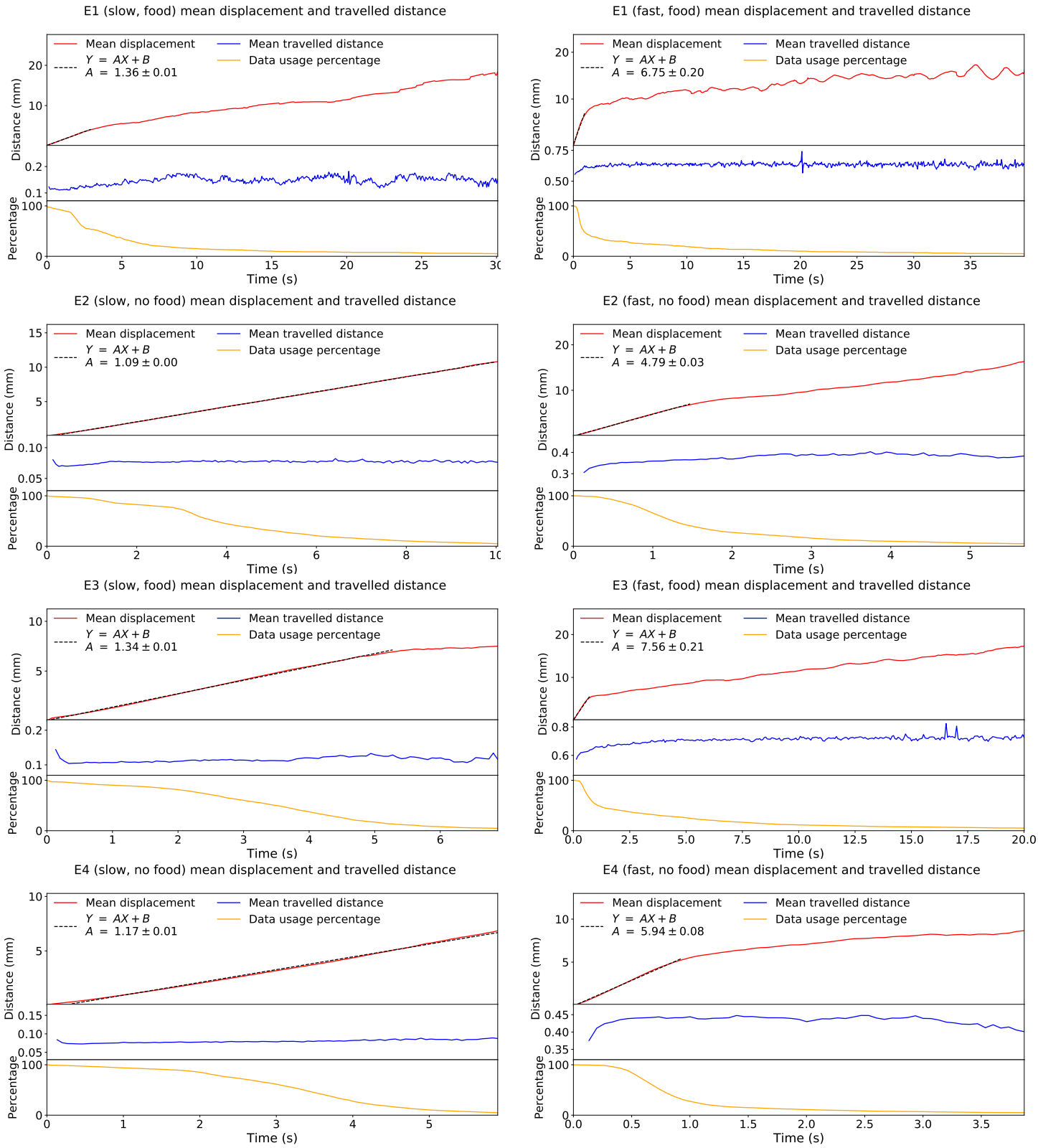


Figure 3.11: Mean displacement (red curve), mean travelled distance between each step (blue curve) and percentage of tracks at least as long as the time (yellow curve). The angular coefficient of the linear fit for the mean displacement represents the mean velocity of each regime for all the experiments. The yellow curves can be thought of as the complementary CDF for the track duration and it was used as a reference to show how much data was being considered when computing the average values. The abscissae axis in each figure is prolonged until the percentage of active data falls under 5%, after which the fluctuations on the average values caused by each individual become non-neglectable.

3.6 3-Dimensional Trajectories

The last step of the analysis was to classify the behaviour of the specimens in the 3-dimensional space. As shown in Fig. 3.2 the trajectories can be quite elaborate and hard to follow, especially the longer ones, showing various and interesting swimming patterns, that would require too many pages to be shown in their entirety.
In order to classify the behaviour of the specimens a Frenet-Serret frame was associated to every point of each trajectory, an example of which can be seen in Figs. 3.12, 3.13. This frame allows to investigate the trajectories using both the curvature κ and the torsion τ, whose PDFs can be seen in Fig. 3.14. The individual trajectories manifest diverse structures, so in order to globally classify them the curvature and torsion were chosen as indicators since they can easily quantify both the rectifiability of the trajectories and give an insight into space exploration. The division in fast and slow trajectories was kept, as the main interest of this work is to characterise the behaviour of the specimens in different motion regimes. The results from the fits on the curvature PDFs are reported in Tab. 3.7.

Experiment	Fit	Fit parameter	KS test p	KS test D
E1s	$y = C\exp\left[\beta\left(x - x_0\right)\right]$	$\beta = (-0.372 \pm 0.015)\, mm$	0.991	0.147
E1f	$y = C\left(x - x_0\right)^{\alpha}$	$\alpha = -1.3 \pm 0.2$	0.478	0.214
E2s	$y = C\exp\left[\beta\left(x - x_0\right)\right]$	$\beta = (-0.319 \pm 0.001)\, mm$	0.999	0.108
E2f	$y = C\left(x - x_0\right)^{\alpha}$	$\alpha = -2.00 \pm 0.12$	0.516	0.201
E3s	$y = C\exp\left[\beta\left(x - x_0\right)\right]$	$\beta = (-0.427 \pm 0.014)\, mm$	0.993	0.136
E3f	$y = C\left(x - x_0\right)^{\alpha}$	$\alpha = -1.5 \pm 0.3$	0.738	0.167
E4s	$y = C\exp\left[\beta\left(x - x_0\right)\right]$	$\beta = (-0.317 \pm 0.004)\, mm$	0.999	0.111
E4f	$y = C\left(x - x_0\right)^{\alpha}$	$\alpha = -2.01 \pm 0.14$	0.951	0.134

Table 3.7: Type of curve fit, parameters and KS coefficients for curvature of both regimes in each experiment. The C parameter is just a normalisation constant, the x_0 parameter is in the range $[-1.2, 1]\, mm^{-1}$ for the fast experiments and $[-2, 0.8]\, mm^{-1}$ for the slow experiments, but its covariance could not be safely estimated so it is not reported.

The slow regime shows a clear exponential decay in curvature for all experiments, indicating that tracks tend to be instantaneously rectilinear and rectifiable, as the curvature radius κ^{-1} quickly tends to infinity. While the slow regime is not particularly interesting to study, it was still reported to show that the decay is faster than that of the corresponding fast regime, that appears to be a power law. The exponents in the slow regime tend to be slightly larger (in absolute terms) for experiments performed in the presence of food, yet the type of motion appears to be the same for all experiments.

The fast regime shows instead a clear distinction between the experiments with food and those without. While the decay is a power-law for all experiments, the two food experiments E1f and E3f manifest smaller compatible exponents (in absolute terms), while their counterparts E2f and E4f show a compatible inverse quadratic decay. In the no food experiments E2f has a low confidence level while E4f, the statistically more relevant experiment, shows a 98.8% confidence. For the two food experiments the confidence levels are instead both low, with E1f especially bad, but probably due to its smaller quantity of data.
The final step before formulating a hypothetical motion scheme for the specimens is to analyse the PDFs for the torsion τ, which are also found in Fig. 3.14. For the slow regime E2s, E3s and E4s show fairly consistent behaviours with a small central peak and two slightly taller, yet still low, peaks to the sides. The differences between maximums and minimums in these PDFs are however always smaller than 0.05, so they can safely be regarded as almost constants. For the slow regime the specimens therefore have a near constant probability to change their osculating plane, with a slightly higher chance of steeper or null changes. This type of behaviour may be an explanation for the dives often found in these types of trajectories, however these dives are not exclusive to the slow regime so further investigation may be required. E1s shows inconsistent behaviour with the other experiments, showing instead a quick fall compatible with the fast regime, but this is most likely due to the low quantity of available trajectories, that also manifest longer average durations.
For the fast regime the PDFs have a distinct peak centred at the origin that quickly falls to 0, suggesting that, at higher velocities, the specimens are more likely to maintain a planar motion relative to their trajectory. This behaviour is consistent independently of the presence of food, with the only difference being that the experiments in the presence of food have taller peaks.
Combining the information from both κ and τ, slow trajectories seem to be mostly linear and either ascending or descending, coherently with what was proposed in the previous sections and almost regardless of the presence of food. The fast trajectories on the other hand tend to differentiate on environments where food is present, since curvatures tend to fall to 0 more slowly in the latter case, meaning that space exploration is incentivised, while it quickly becomes discouraged when no food is found. Behaviourally these may yet again be evidences of a will to preserve energy when resources become scarce, while investing more in exploration when a return is expected.

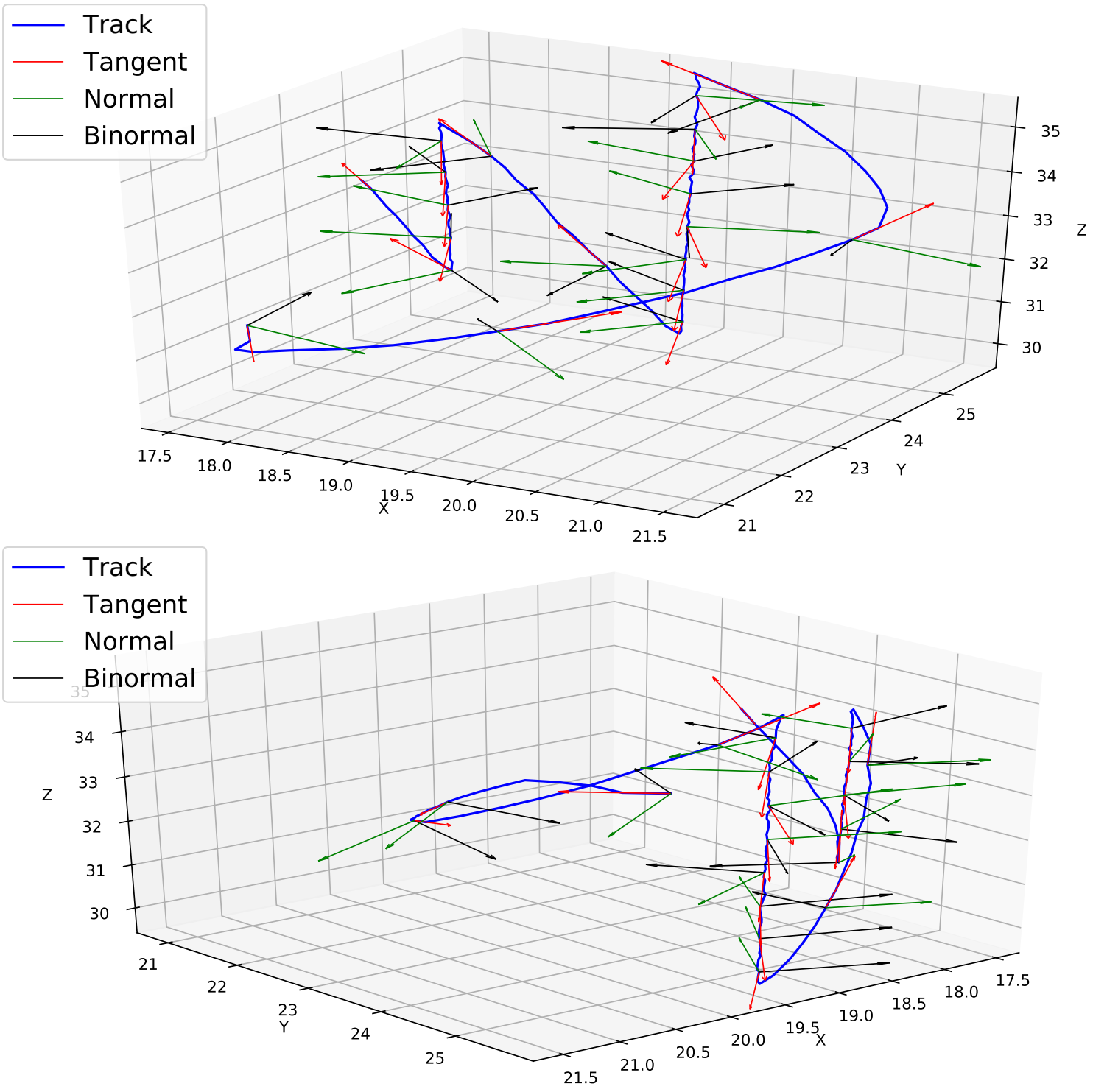


Figure 3.12: Two different views of the same sample trajectory from E1s, manifesting two distinct upwards dives. In order to maintain legibility of the figure only some points show the frame, where the red unitary vector is the tangent, the green one is the normal and the black one is the binormal. Due to the data set being discrete and sometimes with quick changes, as clearly shown in the upward dives, the normal and tangent may be lacking orthogonality as the derivatives used to compute them are just approximations.

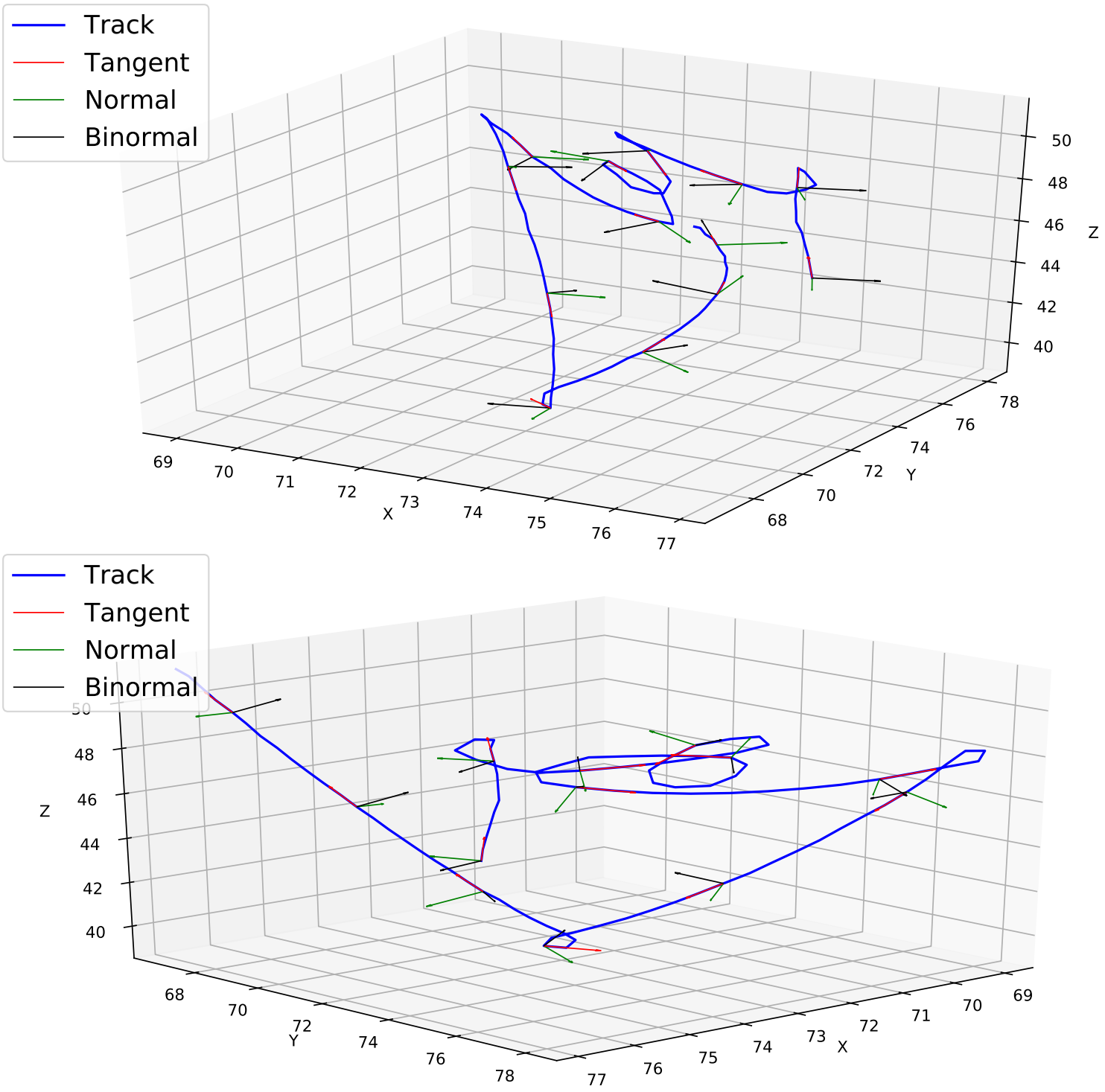
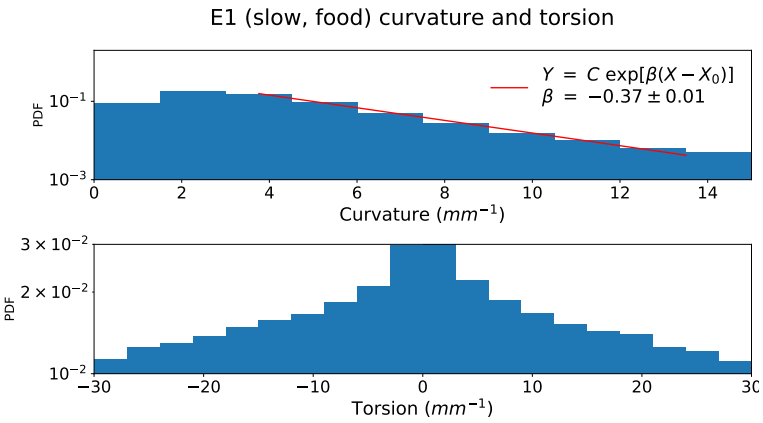
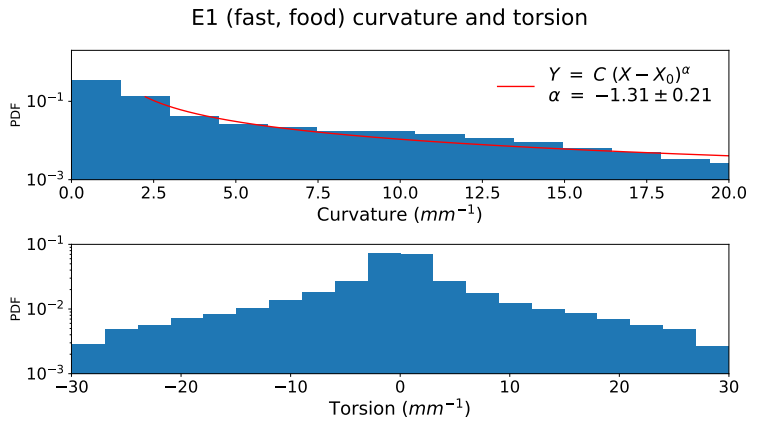
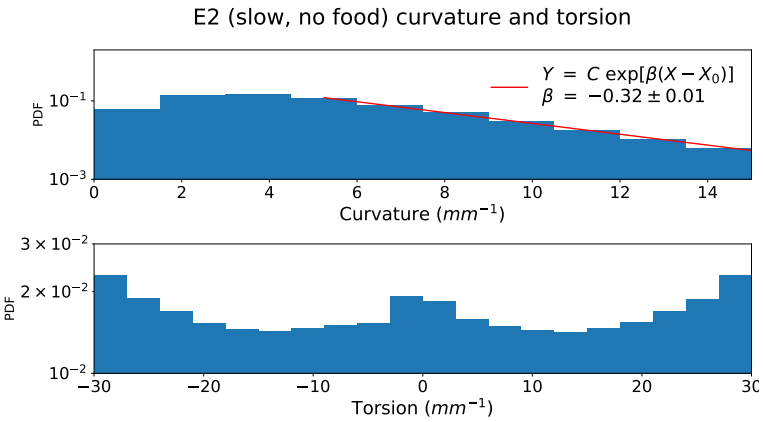
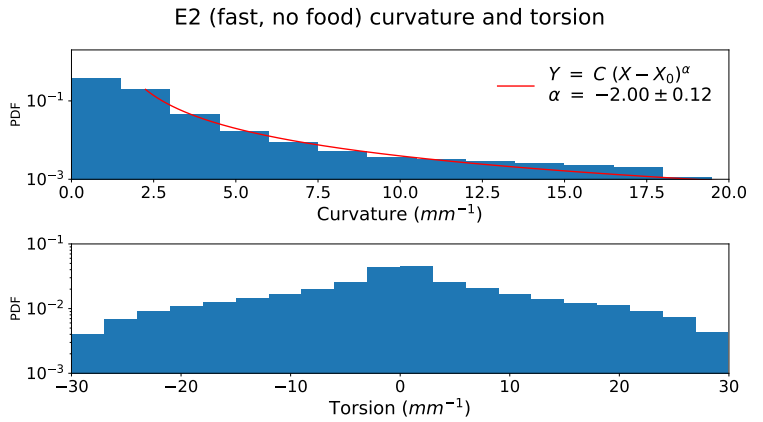


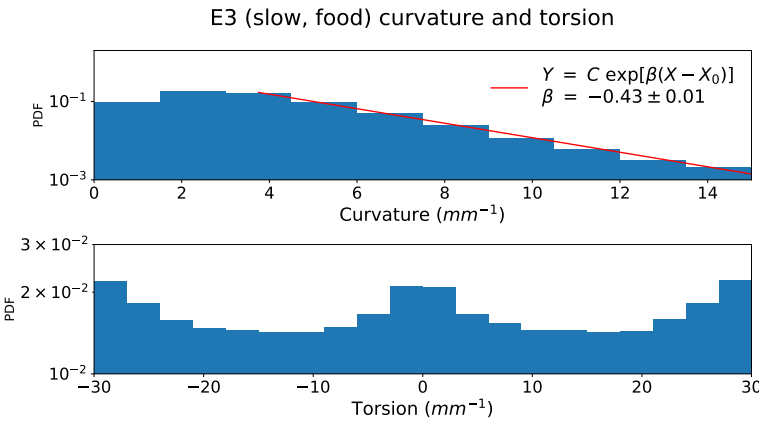
Figure 3.13: Two different views of the same sample trajectory from E3f, manifesting loop structures and turns. In order to maintain legibility of the figure only some points show the frame, where the red unitary vector is the tangent, the green one is the normal and the black one is the binormal. The considerations on orthogonality expressed in Fig. 3.12 also apply here.

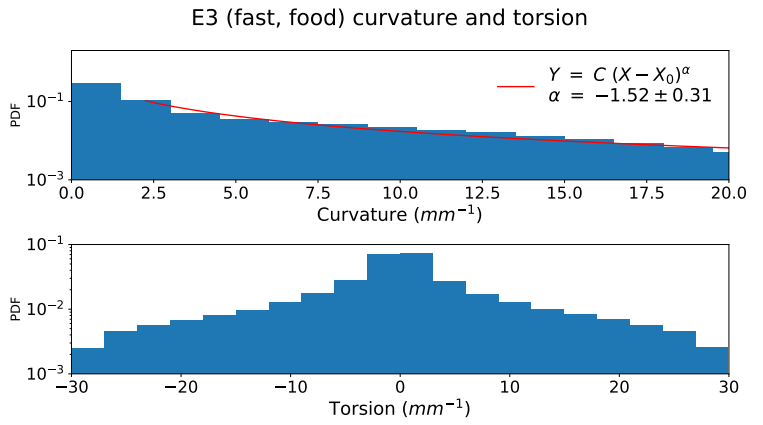


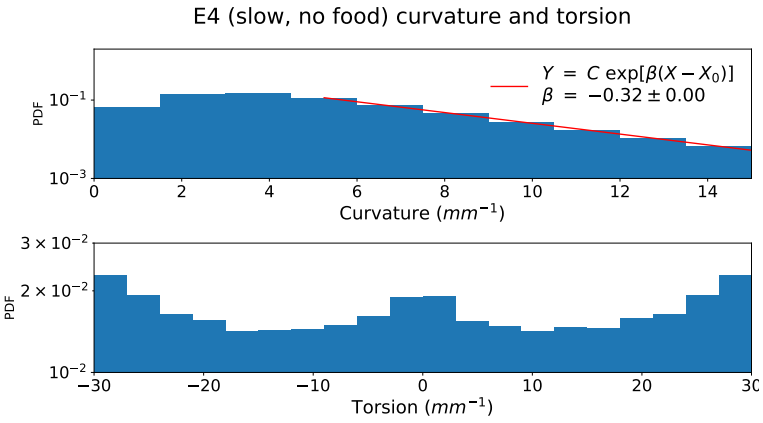












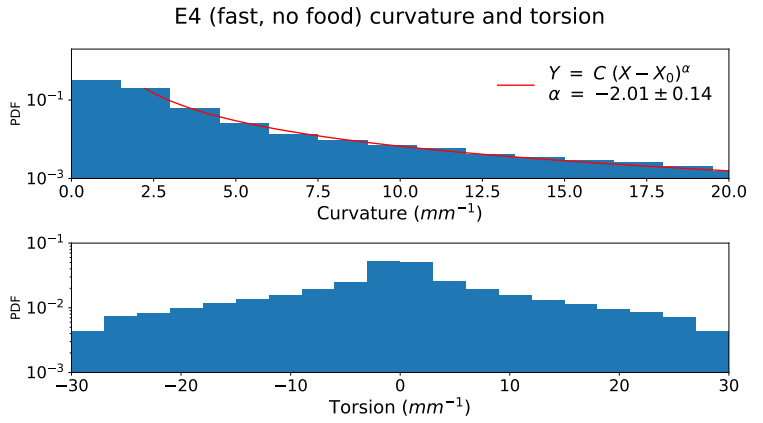


Figure 3.14: Logarithmic *PDF*s for curvature and torsion in slow regime (left) and fast regime (right) for all experiments (ordered top to bottom). The legend reports the curve fit with the main parameter for the curvatures. The exponential decay for the curvature in the slow regime suggests mostly straight trajectories, while the near-constant torsion (with the exception of E1s) suggests upwards and downwards trajectories. The opposite is true for the fast regime, where the power law decay for the curvature is indication of more space exploration which however seems to be more planar.

4. Conclusions

To conclude it is finally possible to summarise the results and formulate a complete hypothesis on the mobility of *C. Furcatus*.
The motion is clearly divided into two regimes depending on the velocity, aptly named *slow* and *fast*. The threshold velocity for differentiating between these regimes seems to be heavily dependant on the presence of food, as the probability density functions show different structures tailored towards higher velocity values when food is found in the environment. For this reasons the velocity chosen as threshold was 2.1 mm/s for experiments conducted in the absence of food and 6 mm/s for those conducted in the presence of food.
When going slow specimens tend not to distinguish too much if food is present or not, and tracks are shorter both spatially and temporally, as shown by the exponential fall for both magnitudes. Curvature for tracks is also decreasing exponentially, meaning mostly straight trajectories that however tend to leave the normal plane of the track. This is also supported by the linearly increasing mean displacement of the specimens. For these reasons it is suggested that the slow regime may be mostly comprised of falls and dives and short periods of rest, constituting a transitory regime.
As velocity increases it becomes apparent that there is a clear distinction as to whether or not food is present in the environment. Tracks tend to be shorter and last less in the absence of food. The power law decay suggests a Lévy flight-like motion, a fact supported by the scatter of mean velocities versus track durations, that highlights the fact that trips tend to be comprised mostly of short bursts, but also manifest a non-neglectable tail of long lasting trajectories at various velocities.
The longer jumps appear to be more frequent when food is present, coherently with the smaller (in absolute terms) exponents of the corresponding probability density functions, meaning longer tails. Space exploration also follows a similar pattern, with curvatures decreasing as a power-law with a similar distinction between exponents, while torsions rapidly decrease to 0, suggesting convoluted, yet planar, trajectories, incentivising exploration of the environment.
The fact that exponents tend to be larger (in absolute terms) when there is no food seems to suggest that creatures tend to resort to energy conservation strategies when no immediate benefit in exploring is found, while an investment in energy is more easily committed with longer jumps and more space exploration when a return is expected.
Combining all this information the suggestion for the mobility is that of an alternation between slower steep upwards or downwards dives and subsequent quick bursts of small scale exploration, with a non-neglectable chance of relocation, starting a new cycle. The power-law properties of track duration, length and space exploration seem to suggest agreement with the Lévy flight foraging hypothesis, showing optimised searching cap-

abilities for points of interest in space. If resources become scarce specimens quickly resort to energy conservation strategies by exploring local space for longer periods of time before committing large quantities of energy in a longer jump to relocate.
To conclude, the elaborate swimming patterns of *C. Furcatus* show many interesting and non-trivial properties, so the proposed model is certainly a very simplified version. The emerging properties of even such simple creatures show that, despite strong individual differences, common behaviours may still emerge causing interesting patterns to appear at any scale of life.

A. Appendix

A.1 Kolmogorov-Smirnov Test

The Kolmogorov-Smirnov (KS) test is a nonparametric test that is used to compare either a sample with a PDF or two samples to determine whether or not they may have been extracted from the same distribution. Only the latter (two-sample KS test) was used in this thesis via the `scipy.stats` Python package (see the `ks_2samp` function in [Vir+20] for reference).
The first necessary step is to introduce the concept of null hypothesis, which is a general statement that two samples derive from the same population or that two phenomena coincide. For the specific application of the two-sample KS test the null hypothesis is that the two samples are drawn from the same PDF. The next necessary definition is that of the empirical cumulative distribution function ($ECDF$). Let $x_1, ..., x_n$ be the results of n measures ordered as

$$x_1 \leq ... \leq x_n \ ,$$

and let $P(x)$ be the CDF associated with that sample, then the $ECDF$ is

$$P_n(x) \stackrel{\text{def}}{=} \begin{cases} 0 & x \leq x_1 \\ \frac{k}{n} & x_k \leq x \leq x_{k+1},\ k = 1, ..., n-1 \\ 1 & x_n \leq x \end{cases} \quad . \tag{A.1}$$

To test the null hypothesis the KS test introduces a statistic. In the one-sample case the statistic for the KS test quantifies the distance between $P(x)$ and $P_n(x)$, in the two-sample case it quantifies between the empirical distributions of the two samples $P_n^1(x)$ and $P_m^2(x)$. Formally

$$D_n \stackrel{\text{def}}{=} \sup_x |P(x) - P_n(x)| \qquad \text{(one-sample)} \ , \tag{A.2}$$

$$D_{n,m} \stackrel{\text{def}}{=} \sup_x \left|P_n^1(x) - P_m^2(x)\right| \qquad \text{(two-sample)} \ . \tag{A.3}$$

It was proven that [Kol33]

$$\lim_{n \to +\infty} \sup_x |P(x) - P_n(x)| = 0 \ , \tag{A.4}$$

so if the null hypothesis is true the statistic tends to 0 as the number of samples increases. Via a probability distribution function called Kolmogorov distribution, which is

evaluated numerically by the aforementioned `ks_2samp` function, it is possible to obtain a confidence level for the null hypothesis to hold true given the value of D_n for the one-sample case, or $D_{n,m}$ for the two-sample case. The values returned from the package and referenced in this work are the statistic $D_{n,m} \equiv D$ and the confidence level p (sometimes also referred to as p-value).
In this thesis the second sample was always a set of values computed from the assumed PDF with the parameters obtained from a series of curve fits, and the sample size was fixed at $m = 5000$.

References

[Car20a] Cardinali, D. *CFurcatus*. https://github.com/Dzegheim/CFurcatus. Last update at the moment of writing: 18-11-2020. 2020.

[Car20b] Cardinali, D. *3DCurve*. https://github.com/Dzegheim/3DCurve. Last update at the moment of writing: 12-11-2020. 2020.

[Gar09] C. W. Gardiner. *Stochastic Methods A Handbook for the Natural and Social Sciences*. Springer Series in Synergetics. Springer, 2009. ISBN: 978-3-5407-0712-7.

[Ibe13] O. C. Ibe. *Elements of Random Walk and Diffusion Processes*. Wiley Series in Operations Research and Management Science. John Wiley & Sons Inc, 2013. ISBN: 978-1-1186-1809-7.

[BH00] D. Ben-Avraham, S. Havlin. *Diffusion and Reactions in Fractals and Disordered Systems*. Cambridge University Press, 2000. ISBN: 978-0-5116-0582-6.

[Kla+90] J. Klafter et al. 'Lévy walk approach to anomalous diffusion'. In: *Physica A: Statistical Mechanics and its Applications* 168.1 (1990), pp. 637–645. DOI: https://doi.org/10.1016/0378-4371(90)90416-P. URL: https://ideas.repec.org/a/eee/phsmap/v168y1990i1p637-645.html.

[Zan] N. Zanghì. 'Brownian Motion.' URL: https://www.ge.infn.it/~zanghi/FS/BrownTEXT.pdf.

[Vis+99] G. M. Viswanathan et al. 'Optimizing the success of random searches'. In: *Nature* 401.6756 (Oct. 1999), pp. 911–914. ISSN: 0028-0836. DOI: 10.1038/44831. URL: https://doi.org/10.1038/44831.

[Rap+09] E. Raposo et al. 'Lévy flights and superdiffusion in the context of biological encounters and random searches'. In: *Journal of Physics A: Mathematical and Theoretical* 42 (Oct. 2009), p. 434003. DOI: 10.1088/1751-8113/42/43/434003.

[Hum+12] N. E. Humphries et al. 'Foraging success of biological Lévy flights recorded in situ'. In: *Proceedings of the National Academy of Sciences of the United States of America* 109.19 (May 2012), pp. 7169–7174. ISSN: 1091-6490. DOI: 10.1073/pnas.1121201109. URL: https://europepmc.org/articles/PMC3358854.

[GBR15] R. Gallotti, A. Bazzani, S. Rambaldi. 'Understanding the variability of daily travel-time expenditures using GPS trajectory data'. In: *EPJ Data Science* 4.1 (2015). DOI: 10.1140/epjds/s13688-015-0055-z. URL: https://doi.org/10.1140/epjds/s13688-015-0055-z.

[Tur99] G. Turchetti. *Dinamica classica dei sistemi fisici*. Zanichelli, 1999. ISBN: 978-8-8080-9820-7.

[Pas+18] R. Pastore et al. 'Distinctive diffusive properties of swimming planktonic copepods in different environmental conditions'. In: *bioRxiv* (2018). DOI: 10.1101/343582. eprint: https://www.biorxiv.org/content/early/2018/06/10/343582.full.pdf. URL: https://www.biorxiv.org/content/early/2018/06/10/343582.

[Bia+13] G. Bianco et al. 'Unexpected Regularity in Swimming Behavior of Clausocalanus furcatus Revealed by a Telecentric 3D Computer Vision System'. In: *PLOS ONE* 8.6 (June 2013), pp. 1–11. DOI: 10.1371/journal.pone.0067640. URL: https://doi.org/10.1371/journal.pone.0067640.

[Vir+20] Pauli Virtanen et al. 'SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python'. In: *Nature Methods* (2020).

[Kol33] A. N. Kolmogorov. 'Sulla determinazione empirica di una legge di distribuzione'. In: *Giornale dell'Istituto Italiano degli Attuari* 4.1 (1933), pp. 83–91.