UNIVERSITAT DE BARCELONA

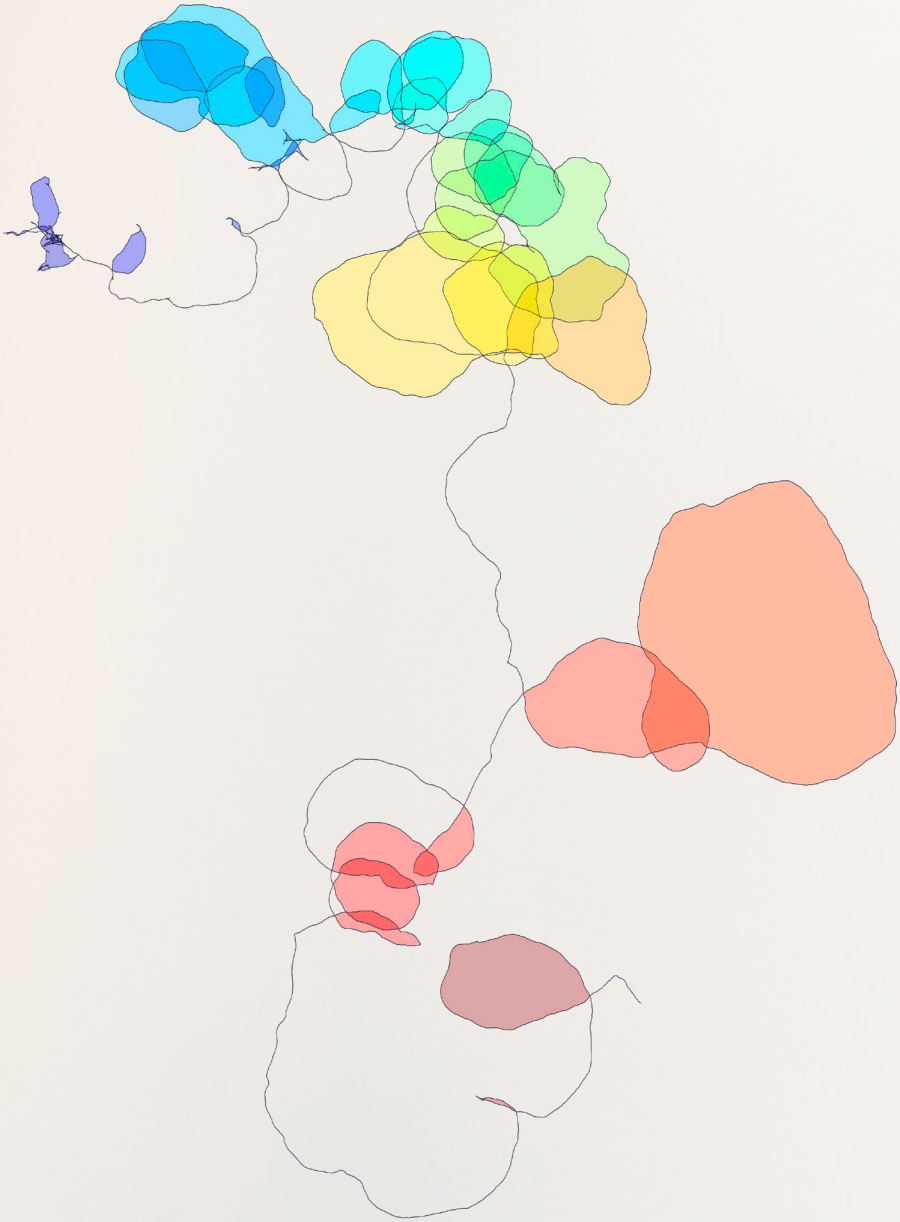
Anomalous diffusion and search behavior in *Caenorhabditis elegans*

Roger Lloret Cabot

ANOMALOUS DIFFUSION AND SEARCH BEHAVIOR IN

CAENORHABDITIS ELEGANS

Roger Lloret Cabot



UNIVERSITAT DE BARCELONA

Anomalous diffusion and search behavior in *Caenorhabditis elegans*

Roger Lloret Cabot

2024

Lloret-Cabot, R., 2024. Anomalous diffusion and search behavior in *Caenorhabditis elegans*. PhD thesis. Universitat de Barcelona. 202 p.

Tesi Doctoral

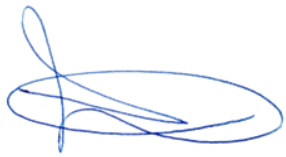
UNIVERSITAT DE BARCELONA

CEAB
exploring aquatic ecosystems

Facultat de Biologia, Universitat de Barcelona
Programa de Doctorat d'Ecologia, Ciències Ambientals i Fisiologia Vegetal

Anomalous diffusion and search behavior in *Caenorhabditis elegans*

Memòria presentada per Roger Lloret Cabot per optar al
grau de Doctor per la Universitat de Barcelona



Roger Lloret Cabot

Departament d'Ecologia Integrativa d'Aigues Continentals
Centre d'Estudis Avançats de Blanes (CEAB)
Consell Superior d'Investigacions Científiques (CSIC)
Març 2024

Director de tesi
Dr. Frederic Bartumeus Ferré
Centre d'Estudis Avançats de Blanes
(CEAB)

Tutor de tesi
Dr. Alejandro Aguilar Vila
Universitat de Barcelona

A la Mare i el Pare,
a la Flàvia, als meus germans,
i a tots els que m'heu acompanyat en aquest camí.

Agraïments

Tot arriba. Aquesta tesi és de lluny la tasca més llarga i costosa que he fet mai. Tan llarga que es fa difícil saber quan vaig començar a orbitar al seu voltant, i tan costosa que es fa difícil saber quan deixaré de fer-ho. Certament, això és una mica igual. El que és important és el procés que he viscut—definitivament no estacionari—, els mecanismes que l'expliquen, i l'entorn de persones que l'han fet possible. Tinc la sensació d'haver estat immers en un procés superdifusiu, similar al que van viure els 126 individus de *C. elegans* que la meva apreciada Mia Panlilio va anar dipositant metòdicament a les plaques d'agar dels experiments que he analitzat. M'he passat la major part del temps invertint els meus esforços localment, revisitant i donant voltes sobre els mateixos conceptes, fins que eventualment l'acceleració del procés ha començat a ser evident. I que segueixi.

A principis del 2015 vaig començar aquest camí motivat per una pregunta que m'interessa profundament i que va ser la principal causa per passar d'estudiar el mar i les Balenes (gràcies, Àlex per aquells magnífics dos anys), a estudiar el moviment i les estratègies de cerca. En aquell moment l'enfocava d'una manera, podem dir, filosòfica, per la influència del meu germà Ferran. Estava preocupat en com reconciliar el concepte de llibertat amb el de coneixement o ciència. Bàsicament, em preguntava si la llibertat és una aparença o és real—l'equació, ai la pregunta, funciona igual substituint llibertat per atzar—i en aquest context de preguntes existencials em vaig trobar, per esdeveniments aparentment atzarosos, el Dr. F. Bartumeus (el meu director i amic Fede). Ara, per mi és bastant obvi que tot allò d'atzar no va tenir pràcticament res. Dios los cría y ellos se juntan diuen.

En aquells seminaris de dispersió i acoblament del plàncton bentos (gràcies, Simone per aquelles meravelloses classes) en Fede em va ensenyar per primera vegada el concepte de "random search". Ens explicava a mi i als meus companys de classe que en certs contextos els animals executen cerques aleatòries per trobar menjar o altres recursos d'interès, i que una estratègia aleatòria pot ser igual o millor que una estratègia determinista. Certa o no, la idea de "random search" és meravellosa i extremadament útil per aprendre. Aquí, per tant, trobo la primera de tantíssimes coses que t'haig d'agrair Fede. T'agraeixo profundament que m'obrissis les portes del teu lab, a mi que en prou feines sabia encendre l'ordinador, i que m'ensenyessis amb una paciència infinita i una dedicació incansable tot el que m'has ensenyat. Que sense embuts és pràcticament tot el que sé en l'àmbit científic. Recordo el pòster al minibrains, l'estada a Leiden, on vaig conèixer al Ghandi, en Van de Koppel, en Sims, en Fryxel i en Simon... recordo l'EcoLets, mare meva l'Ecolets... la visita d'en Will, l'estada a Dresden amb en Rainer i l'Alexei, i també

el congrés a Menorca, on vaig conèixer en Vicenç i en Gonzalo. Quina època més maca. Queda lluny, però el record és nítid. També és nítid el record dels primers passos en programació: fes-me la matriu mare d'aquests tracks, em deies. Allà vaig prendre consciència de la importància de "l'ordre comportamental". Per aconseguir C, primer has d'executar A i després B. Si no, no funcionarà (aquest verb és realment bonic: que una funció faci el que ha de fer). Programar t'obliga a identificar les tasques que s'han d'executar i a ordenar-les en una seqüència lògica. Jo no sabia fer ni una cosa ni l'altra, i és admirable la paciència que vas tenir per ensenyar-me'n. Recordo els atacs de riure al teu despatx amb els noms que posava als objectes. Crec que poques vegades he rigut tant, i tan bé. Recordo les sessions de "brain storming" (Simone, la curvatura!), les discussions a la pissarra i una llista infinita d'anècdotes i vivències de tota mena fins a arribar a les sessions interminables i esgotadores que has encabit sense dilació a la teva agenda per ajudar-me a acabar això que vam començar fa tant de temps. Sempre t'estaré agraït Fede.

Dani Campos, a tu també t'haig d'agrair moltes coses. Has estat una persona clau en el meu procés d'aprenentatge. I com en Fede, també és admirable la paciència i diligència que has tingut per respondre totes les preguntes. Bateries de preguntes recurrents que t'he anat fent des del primer dia. He après molt amb totes elles. T'imagino com un rellotger, muntant el que ha estat, i és, la meva joguina preferida: el model de Langevin. Feia poc que havia llegit el paper de Lenz et al. (2013) i somiava en tenir un model de moviment per reproduir les trajectòries de *C. elegans*. Allò va ser un regal de reis. Igual que la idea de "random search", la idea de "random walk" és meravellosa, i l'equació de Langevin, en especial la que tu vas plantejar amb una persistència en la velocitat i un biaix en la curvatura (la màgia està en els girs), és la manera més bonica de descriure aquesta idea. Per mi, que no domino el camp de la física, aquest model és un pou sense fons i una font d'aprenentatge inesgotable. T'agraeixo molt que m'hagis ensenyat i acompanyat en l'apassionant món de la física estadística.

William Ryu and Mia Panlilio, it would have been a great pleasure to meet you in person and to have had the opportunity to learn more closely from you. I have spent many years dreaming and working towards a stay that could not come true—I hope, however, our paths will cross at some point in the future—. For me, you are a source of inspiration and a clear example of how things in research should be done. Whenever I address a scientific task, especially in the laboratory, I think of you and on how you would have approached it. When I can think like that, I know the outcome will be good. Since I started exploring the dataset you generated, I understood the magnitude of your effort. The pinpoint accuracy of the measurements, the highly-detailed resolution of the trajectories, the beauty in

the simplicity of the experiment, and in the chosen model organism, and also in the observed complexity of its movement patterns. Like the Langevin model Dani gave me, working with your data has been a masterful lesson. The one that grounds me and keeps me down to earth. Something very necessary when developing theory. I have tried to respect the full dataset strictly, and avoid falling into the trap of taking the easiest path. I always believed that, if possible (and is always possible), I should work with all of its contents and not apply unnecessary simplification techniques. Finally, Mia, I would like to let you know the great impact you have had on my research life. Just looking at the schedules of your experiments makes it clear that you are unique. I have given everything I could to take your 126 *C. elegans* trajectories as far as possible. I hope and wish that you and Will would like and enjoy reading this work.

Joan Garriga, gran amic i company de despatx durant tots aquests anys. Tu també has estat una peça clau en tot això. Sempre disposat a donar-me un cop de mà i a explicar-me les coses que no entenia de la manera més didàctica possible. T'agraeixo molt la gran quantitat de temps que has dedicat a ajudar-me professionalment i personalment. A respondre tots els meus dubtes de programació, de matemàtiques, de ciència, de relacions humanes i també de la vida. El nostre despatx i el Leov sempre seran un lloc especial per tot el que hi hem viscut junts. Espero que quan llegeixis aquesta tesi t'hi vegis reflectit i t'agradi el que hi trobis. En gran part és gràcies a tu. Moltíssimes gràcies per tot.

Alfonso Pérez Escudero. Muchas gracias por haber hecho que recuperara el amor por la ciencia, las ganas de seguir intentándolo y sobre todo la fe en mí. Muchas gracias por haber creído tanto y demostrarlo sin tapujos, por haberme dejado soñar y por permitir que persiguiera mis objetivos aun sabiendo que había caminos mejores. Efectivamente, el error en su justa medida es la mejor lección. Muchas gracias por haberme enseñado otra manera de trabajar, por hacerme levantar la cabeza, por siempre apuntar alto y por enseñarme a trazar el plan perfecto para conseguirlo. Muchas gracias por esperarme, de verdad. Muchas gracias. Sin lugar a dudas Toulouse es lo mejor que ha ocurrido durante mi doctorado.

Jofre Carnicer i Alex Gómez-Marín (y también a ti Alfonso), moltes gràcies per haver format la meva comissió de seguiment, i haver estat sempre disposats ajudar-me en tot el que us he demanat. També Biel Obrador, moltes gràcies pels consells que m'has donat per enfocar de la millor manera possible aquest final de tesis tan complicat.

I per acabar els agraïments a les persones que han contribuït explícitament a què aquesta tesi sigui el que és, no puc deixar d'agrair la seva ajuda a l'Àlex Aguilar, a qui sempre guardaré un record especial. Gràcies per haver-me fet de tutor i per haver-me obert les portes del món científic. Estudiar la migració del rorqual

comú va significar complir un somni per a mi i, malgrat que el meu doctorat m'ha allunyat força d'aquests orígens, sento que ara puc contribuir significativament en l'estudi dels cetacis i els grans mamífers marins. Em fa molt feliç veure't a la primera plana de la meva tesi.

Els agraïments a l'Àlex evoquen inevitablement el record del Màster en Ciències del mar. Com he deixat més o menys entreveure a l'inici, pensar en l'origen de la tesi em genera una barreja de nostàlgia i alegria a parts iguals. Alegria perquè aquells dos anys de màster van ser la millor època de la meva vida. I nostàlgia perquè ja no tornaran. Allà, a la Universitat de Barcelona, algun dia entre setmana de l'octubre del 2012 vaig conèixer a la Marta Turon. Concretament a classes d'oceanografia dinàmica. La Marta és l'amiga que tothom voldria. La que m'ha acompanyat en totes les fases d'aquest procés. Ens hem vist créixer i ens hem ajudat mútuament. Haver-te conegut i haver compartit tantes vivències amb tu és de les millors coses que m'han passat. Gràcies per ser com ets i per ser-hi sempre Marta.

Al màster també vaig conèixer la Sílvia Frias, i (re)conèixer en Pol Capdevila i l'Anna Lloveras. Poc després la Jana i en Raül. Magnífic veure com les nostres vides ja estaven interconnectades abans de coincidir literalment en temps i espai. A en Pol a la Sílvia i a l'Anna (i a la Marta, que a partir d'aquí és a tot arreu) els hi agraeixo aquelles trobades que s'allargaven fins altes hores de la nit (segurament jo ja completament adormit) on, entre moltes altres coses, es cuinava el desig de progressar en el món científic i pensàvem en com aconseguir-ho. Els camins que hem seguit han estat força diferents del que imaginàvem—bé, el teu, Pol, no tan diferent—, però el que és segur, és que gràcies a vosaltres vaig trobar l'impuls necessari per decidir la direcció dels meus passos i aterrar al Centre d'Estudis Avançats de Blanes, on m'he desenvolupat com a científic.

El CEAB, apartat del soroll de la metròpolis barcelonina ha estat la meva casa durant pràcticament els últims deu anys. Hi he viscut experiències magnífiques tant en l'àmbit científic com personal. I d'altres de no tan bones. Però el que em quedo sense cap dubte és la gent que hi he conegut. Amics amb els qui he compartit moments inoblidables des dels inicials sopars de Nadal fins a les grans festes del CEAB. La Jana, amb qui no puc parar de riure (encara recordo dient-te que acabava de trobar les meves claus perdudes a dins la butxaca de l'abric). En Raül, el meu germà bessó de tesi—a Girona tu i jo ja apuntàvem maneres—amb qui compartim amor infinit cap als Manel. En Mateu, que és bo amb tot el que fa, ja sigui ballant, jugant a pàdel, escrivint una tesi o fent articles, però que sobretot és bo sent un bon amic. En Xavi, amb qui també hem rigut fins a plorar i que fa massa temps que no veig. En Jorge, que és la generositat en persona i que com en Mateu, millora tot el que toca. En Joan (aka Juanito) i en Xevi, sempre disposats a donar-me un cop de mà al laboratori o a on faci falta. La Cris, que sempre

em farà un forat per xerrar una estona i saber com vaig, igual que en Vicente (mi apreciado compañero), la Cèlia i en Jordi Boada amb qui un dia muntarem consultories LOBO. En Mario, també clar exponent de què significa ser generós. I finalment, la Mar Jambou i en Pol Fernández, als qui tinc una especial estima pel que va significar conèixer-los, en el moment en què vaig fer-ho. Gràcies a vosaltres l'última etapa d'aquest procés (i.e., la travessia del desert) ha sigut molt menys dura. Sou dos amics que ja m'acompanyareu sempre. A tots vosaltres, moltes gràcies per haver fet, d'una manera o d'una altra, que la part divertida d'aquesta experiència hagi sigut inoblidable.

Al CEAB (o gràcies al CEAB) també he conegut moltes altres persones a qui m'estimo i admiro. Amics, coneguts i saludats que diem en català, i que en tots ells sempre he notat el desig que m'anessin bé les coses. L'Adrià, l'Arnau, en Jan, la Maria López, en Rudi, la Marta García, la Nerea, en Víctor Osorio, l'Ibor, la Marta Mora, en Jesús, la Laura, la Marta Pardo, la Catu, en Jesus Bellver, l'Anaïs, l'Alid, en Tom, l'Aurelle, l'Elisabet en Jordi Pagès, la Berta, en Jolle, la Susana, en Xavi Torras— Xavi, aquelles sortides amb bici van ser genials—, en Dani Oro, en David Alonso, la Meritxell Genovart, gràcies Meritxell per haver-me ajudat tant amb els experiments amb les formigues, la Montserrat Soler, la Marta Alaman, la Carmela, la Núria, l'Àngel, en Francesc Coloma, en Xavi Roijals, en Ramón, la Maria Àngels Puig, l'Emili en Marc Ventura, i en Xavier Turon. Moltes gràcies a tots per aquests anys.

I per començar a enfocar el final, voldria agrair l'ajuda i el suport incondicional a les persones que formen part de la meva vida més personal. A la Laia, la Bet i en Bernat, sabeu que amb vosaltres va començar el que abans descrivia com la millor època de la meva vida i malgrat la intermitència temporal, retrobar-nos sempre em carrega les piles per seguir remant. Als meus amics Alberto, Ortega, Ferran, Aimar, Matías, Edgar, Viñas, Alba, Cris, Tania, Marc i Àlex, gràcies per acollir-me com un més a la vostra colla.

Als meus amics de sempre, en David, en Nil, l'Ian, en Pol, l'Anna, l'Oriol, en Guillem, l'Alba en Ramiro i en Pau. Compartir la vida i créixer amb vosaltres és de les millors coses que tinc, sens dubte. Gràcies per haver-me ajudat amb tot el que heu pogut a sortir d'aquest forat i gràcies per respectar la meva absència, i també la meva presència-absència, durant aquests últims mesos. Per fi he acabat la tesi, amics.

A en Joan, a la Present i a l'Òscar. A la Conxi, la Laura, en Marçal, la Carla i als meus estimats nebots, en Narcís, la Joana, la Juliana i la Queralt, gràcies per fer la meva vida més feliç.

A la meva estimada Flàvia. La persona amb qui trobo l'amor, la serenor i la calma tan necessàries per ser feliç. Gràcies per haver cregut sempre en mi i en què podia fer-ho. Sense el teu suport no ho hauria aconseguit.

Al meu germà Ferran, per donar-me l'esperit crític i l'amor pel saber. Al meu germà Martí, per guiar-me sempre. Al meu germà Jordi, per demostrar-me que tot és possible. Als meus germans, Marc i Arnau, per haver-me ensenyat a no rendir-me mai. I a la meva germana Anna, per fer-me tocar de peus a terra i ensenyar-me a relativitzar la magnitud de la tragèdia.

Finalment, i els més importants de tots, gràcies a la Mare i el Pare. Gràcies per haver format aquesta família i haver dedicat la vostra vida a fer-nos créixer. Aquesta tesi és per vosaltres.

Abstract

In this thesis, our focus was on unraveling how *Caenorhabditis elegans* samples its environment in the absence of gradients and food resources. We analyzed data from a relocation experiment in which individual worms, including wild-type isolates and defective mutants, were displaced from a food patch to an empty arena in tightly controlled environmental conditions. We characterized the spreading patterns of the worms and analyzed their movement dynamics at various temporal scales, ranging from milliseconds to hours. Our findings revealed an intricate, time-dependent spreading process, likely driven by internal states and memory relaxation due to the absence of clear external stimuli. Notably, *C. elegans* exhibited superdiffusive spreading behavior, prompting us to investigate the mechanisms behind this phenomenon and its adaptive significance in search contexts. To this end, we devised a classification algorithm that enabled us to characterize *C. elegans* trajectories into its elementary motor behaviors: sharp turns and crawls. Leveraging these insights, we developed a mechanistic model of movement that reproduces *C. elegans* search trajectories in great detail, shedding light on the emergence of superdiffusion. Additionally, we evaluated the search performance of various *C. elegans*-inspired movement models across multiple spatial scales, assuming two distinct search conditions: symmetric (all targets distant) and asymmetric (near and far targets) scenarios. Through this exercise, we quantitatively assessed the contribution of both superdiffusion and turning dynamics to the search performance of *C. elegans*. Finally, in an effort to reinforce our theoretical results with a real case study, we conducted a series of experiments with *C. elegans* individuals foraging in patchy landscapes with bacteria. In these experiments, we employed both homogeneous (regular) and heterogeneous (aggregate) landscapes, mirroring the search conditions evaluated in the synthetic simulations. Our findings confirmed that turning dynamics and superdiffusion are key elements of search behavior in *C. elegans*, fundamentally shaping its search performance in all landscape configurations analyzed. Furthermore, our results suggest that these two mechanisms have the potential to influence the ecological fitness of the species in the natural environment. While our conclusions are primarily based on the specific experimental conditions analyzed, which involve a sharp transition from resource-plentiful to empty arenas, we believe that the findings we present offer valuable insights into both the understanding of *C. elegans* search behavior and the adaptive value of superdiffusion and turning dynamics in search processes.

Resum

En aquesta tesi ens hem centrat a entendre com *Caenorhabditis elegans* explora un entorn sense menjar ni gradients ambientals. Hem analitzat el moviment d'aquest animal en un experiment de reubicació, en el qual vàrem traslladar individus de diverses soques tant salvatges com mutants, des d'una placa de petri amb menjar, a un espai sense menjar i amb condicions de temperatura i humitat molt controlades. Per desenvolupar la nostra recerca hem caracteritzat detalladament els patrons de dispersió de *C. elegans*, basant-nos en les trajectòries obtingudes i també en la dinàmica del seu moviment mesurat a diverses escales de temps, des dels milisegons fins a l'hora. Els nostres resultats indiquen que *C. elegans* dispersa segons un procés no estacionari, probablement modulat pels seus estats interns i no pas per estímuls externs com per exemple gradients ambientals. Cal destacar que els individus de *C. elegans* estudiats varen exhibir un patró de dispersió superdifusiu, fet que va motivar-nos a investigar en profunditat els mecanismes fisics que expliquen l'emergència d'aquest fenomen i també quina és la seva rellevància adaptativa en contextos de cerca. Per fer-ho, vàrem desenvolupar un algoritme de classificació per caracteritzar les trajectòries de *C. elegans* en els seus comportaments motrius elementals. Basant-nos en els resultats obtinguts, vàrem desenvolupar un model mecanicista del moviment de *C. elegans* que ens ha permès entendre quins mecanismes físics generen superdiffusió. Seguidament, ens vàrem centrar a caracteritzar la seva eficiència en condicions de cerca simètriques, on tots els recursos es troben lluny, i en condicions de cerca asimètrica, on es poden trobar recursos tant a prop com lluny. A través de simulacions virtuals, vàrem quantificar la contribució de la superdifusió i de la dinàmica de girs en l'eficiència de cerca de *C. elegans*. Finalment i amb la intenció de reforçar les nostres conclusions teòriques vàrem fer una sèrie d'experiments de cerca on individus de *C. elegans* podien buscar, trobar i consumir menjar distribuit segons un paisatge homogeni (regular), utilitzat per estudiar la condició simètrica, o un paisatge heterogeni (agregat), utilitzat per estudiar la condició asimètrica. Els resultats obtinguts varen confirmar que tant la dinàmica de girs com la superdifusió són dos elements clau en el comportament de cerca de *C. elegans*, ja que determinen la seva l'eficiencia de cerca. Més encara, els nostres resultats suggereixen que aquests dos mecanismes podrien arribar a influir l'exit reproductiu de l'espècie a la naturalesa. Tot i que les conclusions obtingudes en aquesta tesi són relatives a condicions experimentals expecífiques, creiem que els resultats presentats i les seves interpretacions, ofereixen un coneixament valuós per entendre el comportament de cerca de *C. elegans* i el valor adaptatiu de la superdifusió en els processos de cerca.

Contents

Chapter 1

Introduction

Search, purposeful movement to find things, is one of the most fundamental tasks living beings must cope with. From single bacteria to majestic blue whales, organisms explore their surrounding environment through movement directed to locate the resources of interest. Indeed, living and reproducing is contingent upon succeeding in this task. Consequently, we expect a diverse array of search strategies to exist across the three domains of life, each shaped by specific ecological contexts and evolutionary histories (Bell, 1991; Stephens et al., 2007; Viswanathan et al., 2011; Méndez et al., 2014). However, despite its pervasive nature and significance, our understanding of the generative mechanisms and evolutionary drivers of search behavior remains limited (Bartumeus and Levin, 2008; Bartumeus et al., 2016).

While various biological mechanisms, such as perception and learning, are well-documented in animals for acquiring and maintaining relevant environmental information (Real, 1981; Bell, 1991; Turchin, 1998), it is imperative to recognize the role of uncertainty in search processes (Viswanathan et al., 2011; Méndez et al., 2014; Bartumeus et al., 2016). Uncertainty regarding the precise location of resources is inherent to the concept of search itself. So, even though the ability to successfully solve search problems is tightly linked to information processing, the latter is required to be partial, either biased or inaccurate. Organisms, regardless of their cognitive capabilities, tend to leverage past experiences and learn, after several iterations, to optimize their search strategies (Bell, 1991; Stephens et al., 2007). Particularly in resource-limited environments (Bartumeus et al., 2003, 2010). Consequently, organisms engaged in search behavior adjust their motor responses based on prior expectations about resource distribution. The unfolding of a basic search template, and its further adjustment, is particularly crucial in weakly informative searches. Also, this adjustment may not necessarily be a swift process, as all information about resource location needs to be mapped *in-situ* through a sampling

strategy. It is precisely the sampling strategy used to collect the initial information the one that is more closely connected to the idea of search.

Despite the importance of understanding animal search behavior from first principles, laboratory experiments focusing on the concept of search in cue less environments are scarce in the literature. Similarly, studies aimed at elucidating the role of uncertainty and its potential adaptive significance are limited (Berg, 1983; Denny and Gaines, 2002; Bartumeus et al., 2005; Moore et al., 2017). Instead, much of the research effort has been directed toward understanding exploitative-information mechanisms, such as chemotaxis (Saeki et al., 2001; Berg and Berry, 2004; Pierce-Shimomura et al., 2005), thermotaxis (Ryu and Samuel, 2002) or path integration (Kim and Dickinson, 2017), to put some ecologically relevant examples. This bias may stem from both the relative ease of interpreting and manipulating these mechanisms (look for gradient climbing), and the clear optimal reference models in biological systems. In consequence, great advancements has been made on investigating how organisms use and exploit available information, but there remains a relative paucity of research addressing how organisms actually deal with uncertainty, or in other words, exploit the lack of information.

From a foundational perspective, search phenomena might be viewed within a three-dimensional space, comprised of three different axes: the amount of information, the distance to targets, and the energy or time available to search. The primary and most intuitive axis is the information axis. At its lowest extreme, when information is entirely absent, the search process is known to be a pure search (Bell, 1991; Viswanathan et al., 2011; Campos et al., 2014; Méndez et al., 2014; Bartumeus et al., 2016). On the contrary, with complete information, the need for search vanishes as the precise location of resources becomes known. In this circumstance, directed movement (i.e., ballistic and oriented motion) toward them is the optimal motor response. Following resource detection, various ecological phenomena ensue, including prey capture and handling, resource utilization, and even post-recruitment processes, among others (Stephens et al., 2007; Mischiati et al., 2015). However, these phenomena are tangential to the search process, and thus lie beyond the scope of this thesis. Therefore, search behavior unfolds within the continuum between zero and complete information, with the extent of information inversely influencing the intensity of search efforts.

Second, we might consider the distance-to-target axis. Importantly, spatial dimensions should not be considered in absolute terms but rather conceptualized as the set of Euclidean distances from the searcher position to the resources or targets explored. Thus, the distribution of resources in a given landscape matters from the relative perspective of the searcher, a distance-to-target perspective. Analogous to the information axis, extreme scenarios offer valuable insights. When the distance

to targets is effectively zero, searches cease to exist. This may occur either upon locating the targets through a completed search process, or by residing in a location already occupied by the targets. The latter scenario is exemplified in various natural occurrences such as animal nests, parental care in bird species, or mammalian lactation. Notably, in humans and other primates, maternal provisioning of breast milk to infants eliminates the need for either party to engage in resource search, as the food resource is readily available. Similarly, if targets are positioned at an infinite distance—though this is a purely theoretical scenario—searches effectively cease to exist too. In this context, searching would be akin to navigate an environment devoid of targets, as reaching infinity is impractical. Therefore, it is in the continuum of these two extremes where the distance-to-target perspective becomes relevant.

Finally, we find the third axis, related to search time allocation and energy availability. Indeed, the search is a spatial sampling process that unfolds on time and, specially in the context of biology, this time is inherently finite. While one can theoretically envision search processes unfolding over infinite time spans, and we acknowledge this approach can offer valuable insights (Viswanathan et al., 2011), living systems must contend with finite energy reservoirs. In scenarios where time is unlimited, search outcomes may seem arbitrary, with seemingly suboptimal strategies appearing effective due to the eventual discovery of targets, regardless of their time-constrained statistical efficiency. Conversely, when time or energy costs are relevant, certain strategies may prove detrimental while others emerge as highly effective. Therefore, energy constraints significantly shape our understanding of both search, and optimal search behavior. Indeed, the impact of different search strategies needs to be framed into adequate ecological scales. The latter are different for different organisms and ecological contexts (i.e. mating, feeding, breeding site selection). The ecological implications of search behavior, makes searches an excellent tangible framework for assessing natural selection and evolution. To further illustrate its relevance let us put a familiar example:

Imagine we are late in a first date, so that we need to find our car keys as soon as possible. The first thing we do is going straight to the key hanger at the entrance of our home, and try to find them were they were supposed to be, but unfortunately they are not there. We then go to the living room and perform a local search on the table, each of the chairs, and the sofa. But the keys are not there. Even, we go to the key hanger again, and try to find them where they were supposed to be. We search in our bed and in the living room again. Repeating the same actions at least twice until, nobody knows why, we touch our leg and realize we had the keys in our pocket. However, at that time, we are being very late.

Our conclusion then, must be that our search behavior was largely suboptimal. However, by analyzing our actions we can glean valuable insights. Specifically, three key points emerge from this example:

- Information is very valuable: I wish I had known I had the keys in my pocket—however elementary it may seem, a searcher need to know if the distance to sough targets is zero—.
- Wrong "information", or more precisely having a wrong model of the world, is disastrous. Let us say, bad decisions are expensive.
- Information processing i.e., the ability to change our model of the world and accept that the keys were not where they were supposed to be, required us to repeat the same actions more than once. Additionally, these suboptimal actions (spatial revisits) enabled us to made sure we did not overlook the keys in a place we already checked.

Having learned these valuable lessons, it is reasonable to anticipate that living organisms possess highly developed sensory systems to acquire unbiased and accurate information during searches (Stephens et al., 2007). Indeed, this expectation aligns with reality, as evidenced by the remarkable capabilities of complex sensory organs such as the eyes, ears and nose, but more rudimentary sensory machinery, geared towards the same objective, can be observed in the form of thermal, chemical or photo sensors in simple organisms as bacteria, amoebas, plants, and small animals such as *Caenorhabditis elegans* (Ueda et al., 1975; Takahashi and Scott, 1993; Berg and Berry, 2004; Pierce-Shimomura et al., 1999). Similarly, we should also anticipate the existence of mimicry species capable of deliberately mislead predators (or preys) by providing false information (Stephens et al., 2007). This phenomenon is particularly prevalent in cryptic species, where deception plays a crucial role in search, survival and reproduction.

1.1 Anomalous diffusion and search tradeoffs

A search is essentially a sampling strategy unfolded over time and space. As with many other facets of life, the crux of the matter lies in the tension between taking good and wrong decisions. In the context of search behavior, these decisions revolve around whether and how to heed specific cues. Determining the pace of exploration and the degree of spatial overlap, required to enhance detection probability, are central considerations. As well as deciding how to allocate sampling effort. Many of these facets determining the efficacy of a search strategy can be lumped in one simpler idea, which is deciding whether to explore nearby or instead looking for

better resources elsewhere. This dilemma, known as the intensive-extensive search tradeoff (Kramer and McLaughlin, 2001; Bazazi et al., 2012; Bartumeus et al., 2014; Pradhan et al., 2019), is a particular case of the more general exploitation-exploration tradeoff, and underscores the fundamentals of any search process (Hills et al., 2015; Bartumeus et al., 2016). Its adequate balance depends on both the available information and the prior expectations. If there are good reasons to expect targets are nearby, the optimal behavior should be trying to find them nearby. That is, being spatially local and intense. Contrary, if targets are known to be faraway, executing an extensive search and relocate fast would be the optimal choice. The interesting scenario occurs when no information about target location is available, meaning that no *a priori* assumptions on whether targets are near or far away can be made. In this context, balancing the search tradeoff becomes crucial in order to unfold an optimal search behavior (Bartumeus et al., 2014, 2016). On the contrary, any information gain, would decrease the need of balancing such a tradeoff. We might ask ourselves, however, what does exactly mean to balance the intensive extensive tradeoff. How can we assess if a search pattern is more or less spatially intense? To address this question, researchers have focused on analyzing movement in both living organisms and random walk models (Turchin, 1998; Viswanathan et al., 2011; Méndez et al., 2014).

At the population and macroscopic level, animal movement is often viewed as a diffusion process characterized in terms of the mean square displacement (MSD) (Turchin, 1998; Okubo et al., 2001; Méndez et al., 2014). This magnitude describes the spreading or dispersal pattern of a group of moving objects over time—in this case animals—. Typically, three different regimes are distinguished depending on whether the MSD grows linearly with time (normal diffusion), supra-linearly (superdiffusion), or infra-linearly (subdiffusion). Whenever the MSD grow in a non-linear fashion, either faster (superdiffusion) or slower (subdiffusion) than in normal diffusion, we recognize it as anomalous, and classify the spreading process in the realm of the anomalous diffusion. Importantly, studying the origin and effects of anomalous diffusion can help to understand the intricacies of the intensive-extensive tradeoff and search behavior in general (Bell, 1991; Viswanathan et al., 2011; Méndez et al., 2014; Bartumeus et al., 2016), as the diffusive regime describes how sampling effort is allocated in both time and space.

At the individual level, the focal point of investigation lies in the search trajectory. The technological advancements in movement tracking devices such as GPS and biologgers, which are becoming increasingly affordable and compact, have propelled the field of movement ecology. This revolution has facilitated the documentation of animal movement in natural habitats across a broad spectrum of scales and ecological contexts (Nathan et al., 2008; Kays et al., 2015; Tucker

et al., 2018). Researchers have leveraged these data to explore the properties of trajectories and population dispersal, revealing prevalent instances of anomalous diffusion and diverse dynamics of movement intensity and extent. Nonetheless, some datasets pose challenges in discerning the underlying drivers of observed patterns and whether search processes contribute to them at all (Benhamou, 2007; Benhamou and Collet, 2015; Pyke, 2015).

Despite the ongoing technological revolution, which holds promise for bridging this gap in the future, achieving controlled conditions in field setups remains challenging. Of particular concern is the difficulty in determining the quantity and nature of information animals possess while in motion, as well as the underlying purpose behind observed movement patterns (Sims et al., 2008; Humphries et al., 2010; Jager et al., 2011). Delving into search behavior necessitates greater control over environmental variables, including the quantity of accessible information, sensory capabilities, and internal states (such as motivation) of organisms. In this regard, model organisms emerge as optimal candidates for investigating specific concepts related to search behavior.

1.2 *C. elegans* as a model organism

Model organisms are species that are extensively studied to understand particular biological phenomena. They are chosen for their ease of manipulation, rapid reproduction, and similarity to other organisms. Some examples of animal model organisms include: fruit flies (*Drosophila melanogaster*), mouse (*Mus musculus*), zebrafish *Danio rerio*, or the roundworm *Caenorhabditis elegans*.

Among these model organisms we chose *C. elegans* for studying search behavior due to several compelling reasons. This organism offers significant advantages in both the feasibility of experiments and the interpretation of results. It is simple enough to be manipulated at both the genetic and neuronal levels, yet complex enough to exhibit a rich repertoire of behaviors, including complex exploratory patterns and even basic associative learning and memory conditioning (Rankin et al., 1990; Ardiel and Rankin, 2010; Flavell et al., 2020). Furthermore, *C. elegans* is well-characterized and one of the most extensively studied animals, providing a robust foundation for our research (Brenner, 1974; Meneely et al., 2019).

In contrast to other animals with high-cognitive skills and complex brains and sensory systems, *C. elegans* allows us to control and manipulate its available information to its bare essentials: internal states, memory, and *in-situ* learning. This simplicity enabled us to focus specifically on the mechanisms underlying search behavior without the confounding effects of extraneous factors such as fear responses. Additionally, the availability of finely tuned mutant strains enhanced the precision

and scope of our experiments, allowing us to investigate specific aspects of search behavior, such as hunger sensation modulation, and natural *versus* strain-related variability on search behavioral templates.

Chapter 2

Objectives

This thesis aims to provide a thorough understanding of *Caenorhabditis elegans* movement under minimal information gradients or fluxes. Throughout the thesis, we assumed *C. elegans* movement in these circumstances is mainly triggered as a search response to the lack of food or relevant environmental information, like thermal or chemical gradients. In order to approach such research, we subdivided this general objective into the following specific objectives:

1. Characterize the spreading pattern of *C. elegans* in the context of relocation experiments and diffusion theory.
2. Understand *C. elegans* movement across scales, identifying the mechanistic links between microscopic, mesoscopic, and macroscopic trajectory patterns.
3. Quantitatively asses the search efficiency of *C. elegans* movement in key limiting case scenarios.

Below, we provide a summary description of the thesis chapters and how they are associated to each specific objective (Ob):

- **Chapter 4 -** *Hurst exponent decomposition of C. elegans search trajectories* (Ob1)
  This chapter offers a first overview of how *C. elegans* explores space when searching for food in low-information conditions. Our aim was characterizing the spreading pattern of the strains under study and delineate the main differences between them. To do so, we computed their mean square displacement (MSD) and determined the phenomenological origin of their diffusive regimes. More specifically, we assessed the contribution of non-stationary dynamics, long-ranged correlations and extreme events, on the scaling exponent of the MSD.

- **Chapter 5 -** *Temporal unfolding of sharp turns and crawls cause the non-stationary spreading of C. elegans search* (Ob2)
  The main objective of this chapter is to provide a detailed description of *C. elegans* elementary motor behaviors: sharp turns and crawls, in terms of their most relevant movement features. To do so, we developed a classification algorithm that allowed us to categorize *C. elegans* trajectories into sharp turns and crawls, and characterize their associated distributions of speed, angular velocity, duration, net displacement and turning angle, along with their temporal unfolding.

- **Chapter 6 -** *Memory decay drives superdiffusion in C. elegans search behavior* (Ob1, Ob2)
  This chapter aims to determine how *C. elegans* is able to superdiffuse even when it shows a systematic curvature bias in the direction of motion, apparently more compatible with the subdiffusive regime. In order to achieve that, we developed a model of movement of *C. elegans* that allowed us to understand which physical mechanisms, operating at the microscopic scale, determine the diffusive regime of the mean square displacement.

- **Chapter 7 -** *An ordered mind: C. elegans search behavior outperforms equivalent random but stationary movement patterns* (Ob3)
  This chapter aims to address two complementary objectives. Firstly, to evaluate the impact of superdiffusion on the search performance of *C. elegans*. Secondly, to elucidate the role of turning dynamics in *C. elegans* search behavior, particularly investigating whether the presence of a curvature bias represents a mere biological constraint or instead holds adaptive value. To achieve these objectives, we devised a series of search efficiency simulations comparing *C. elegans*-inspired movement models with equivalent random walks. In addition, we conducted a series of foraging experiments, mirroring the search conditions of the synthetic simulations, to both reinforce and complement our theoretical results.

Chapter 3

Experimental data

Abstract

Here, we present the experimental data used throughout the thesis. The data-set comprises 126 trajectories of *Caenorhabditis elegans* individuals moving in a virtually unbounded bare arena of 24.5×24.5 cm, after being growth in standard conditions (Brenner, 1974). The development of the tracking system, execution of the experiments, and data collection were carried by Mia Panlilio under the supervision of William S. Ryu, at Ryu Lab (Department of Physics and the Donnelly Centre, University of Toronto, 60 St George St., Toronto, Canada M5S1A). The experiments were based in prior *C. elegans* works (Stephens et al., 2008; Salvador et al., 2014) and were especially designed to promote a free transition from exploitative to exploratory movement patterns without having the effect of physical disturbances such as plate walls, or reactive movement behavior such as thermotaxis (Ryu and Samuel, 2002) or chemotaxis (Pierce-Shimomura et al., 1999). Our primary objective was to assemble a high-resolution data-set that allowed us to unravel mechanistic links between microscopic movements dynamics and macroscopic trajectory patterns, such as the mean square displacement. In the following sections, we will delve into the specifics of the strains used, the experimental design, and the methodology employed to acquire the trajectories.

3.1 Strains

Worms were maintained at the assay temperature (21 ± 0.1 °C) and fed on *Escherichia coli* OP50 lawns as described in Brenner (1974). N2 (Bristol), CB4856 (Hawaiian), and DA609 npr-1(ad609) were acquired from the *Caenorhabditis* Genetics Center. VC125 tyra-3(ok325) was kindly provided by M. de Bono (Cambridge University, Cambridge, UK). The experiments were performed at the Ryu Lab (U. Toronto) by Mia Panlilio and William S. Ryu over 3-months, with strain randomization. Daily randomization was used between N2 (n = 48), CB4856 (n = 30), DA609 (n = 24); and VC125 (n = 24).

3.2 Worm tracker

The tracking system was composed of a GigE vision camera (Manta G201B-30fps, Allied Vision Technologies; lens MMS R-4 and MMS OBJ-11, Edmund Optics) mounted on a modified computer numerical control machine (Zenbot CNC 2424). A temperature-controlled aluminum plate constituted the stage. To achieve uniform illumination during tracking, a ring of 12 red LEDs were fixed below the lens through a custom-designed Delrin attachment. A glass mirror was placed on the tracker stage such that light, transmitted downwards through the NGM assay plate, was reflected upwards towards the *C. elegans* individual. To minimize potential heating, a current controller was integrated into the LED circuit and synchronized to the camera shutter, thus achieving a stroboscopic effect. Water was circulated through channels within the aluminum stage to maintain NGM temperatures at 21 ± 0.1 °C, confirmed by an infrared camera (Isotemp 3016D, Fisher Scientific; ICI 7640 P-Series). The entire apparatus was contained in an acrylic enclosure to avoid external perturbations and NGM desiccation over time. Custom machine vision control software was developed in LabVIEW (NIVision and NI-Motion packages, National Instruments, 2012). Images were recorded in JPEG format at 30 Hz (1600 x 1234 pixels, 525 pixels $\cdot$ mm^{-1}). Basic image processing from gray-scale to binary PNG images and calculation of in-frame center-of-mass was performed in LabVIEW, post-acquisition. Trajectory reconstruction was done using custom scripts in MATLAB (MathWorks, r2011b).

3.3 Behavioral assay

Large NGM plates (24.5×24.5 cm, Square BioAssay Dish, Corning Inc.) were prepared without bacterial lawns. Each plate contained 200 mL of NGM. Those that were not used within 24 hours of pouring were stored at 4 °C, and no plate older than one week was used in any experiment. Before tracking, each plate was placed on the stage for 0.5-1 hours to ensure that the NGM was at thermal equilibrium during experiments. Following each experiment, plates were cleaned, ethanol-washed, and reused. Assayed worms were well-fed, young adult hermaphrodites. For each replica, an individual was manually transferred from the culture plate to a 1 mL wash of M9 buffer, where it was allowed to swim for 1-2 minutes. The worm was then placed on the assay plate and tracking began but the first few minutes (always less than 5 minutes) are discarded as the worm initially behaves sightly different, and the tracking system requires an adaptation time. Images were acquired for 90 min or until the worm crawled to the edge of the assay plate (Figure 3.1).

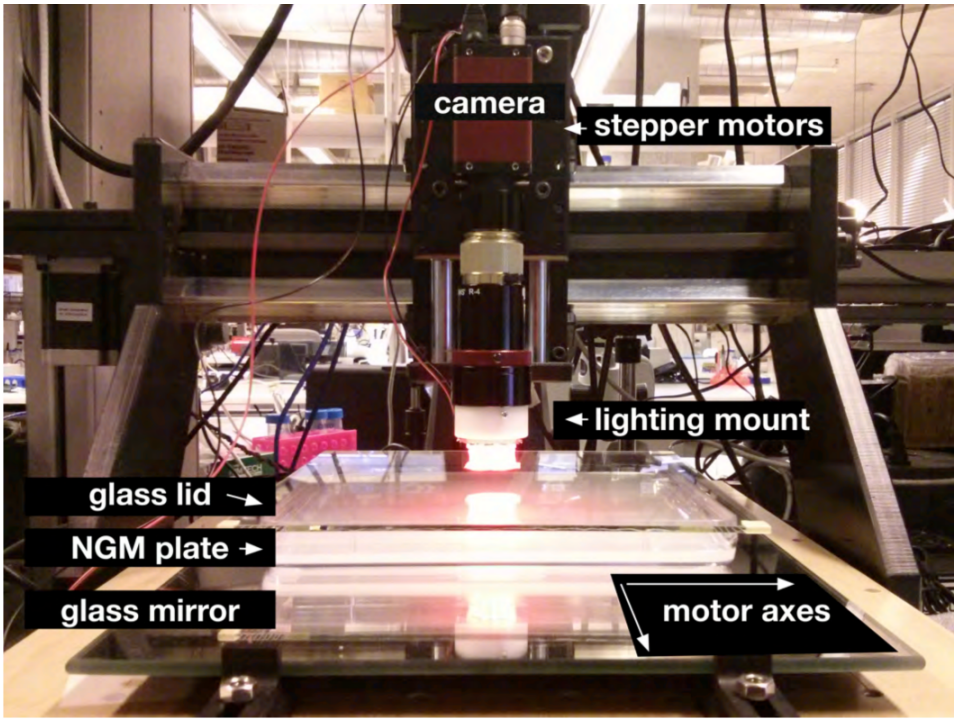


Figure 3.1: Worm tracker utilized to record high-resolution *C. elegans* trajectories in a virtually unbounded bare arena. Image source: Mia Panlilio, 26-08-2013 (Panlilio and Ryu, 2013).

3.4 Trajectories

Trajectories (Figures 3.2-3.8) were reconstructed using the in-frame center-of-mass and CNC stepper motor coordinates. The time series of the coordinates $x(t)$ and $y(t)$ were smoothed using a third degree polynomial across a 1-second sliding window ($\pm$ 15 frames), as in Stephens et al. (2010). Speed was calculated by taking the first time derivatives of the smoothing polynomial. Similarly, angular velocity was calculated as the time derivative of the turning angle, computed at the trajectory frame-rate (33Hz). The turning angle measures the change of direction between two successive vectors (or three consecutive locations), which in this case span an overall duration of 0.099 seconds. In order to differentiate left (positive) to right (negative) turns, we scaled the turning angle from $-\pi$ to π before computing the angular velocity.

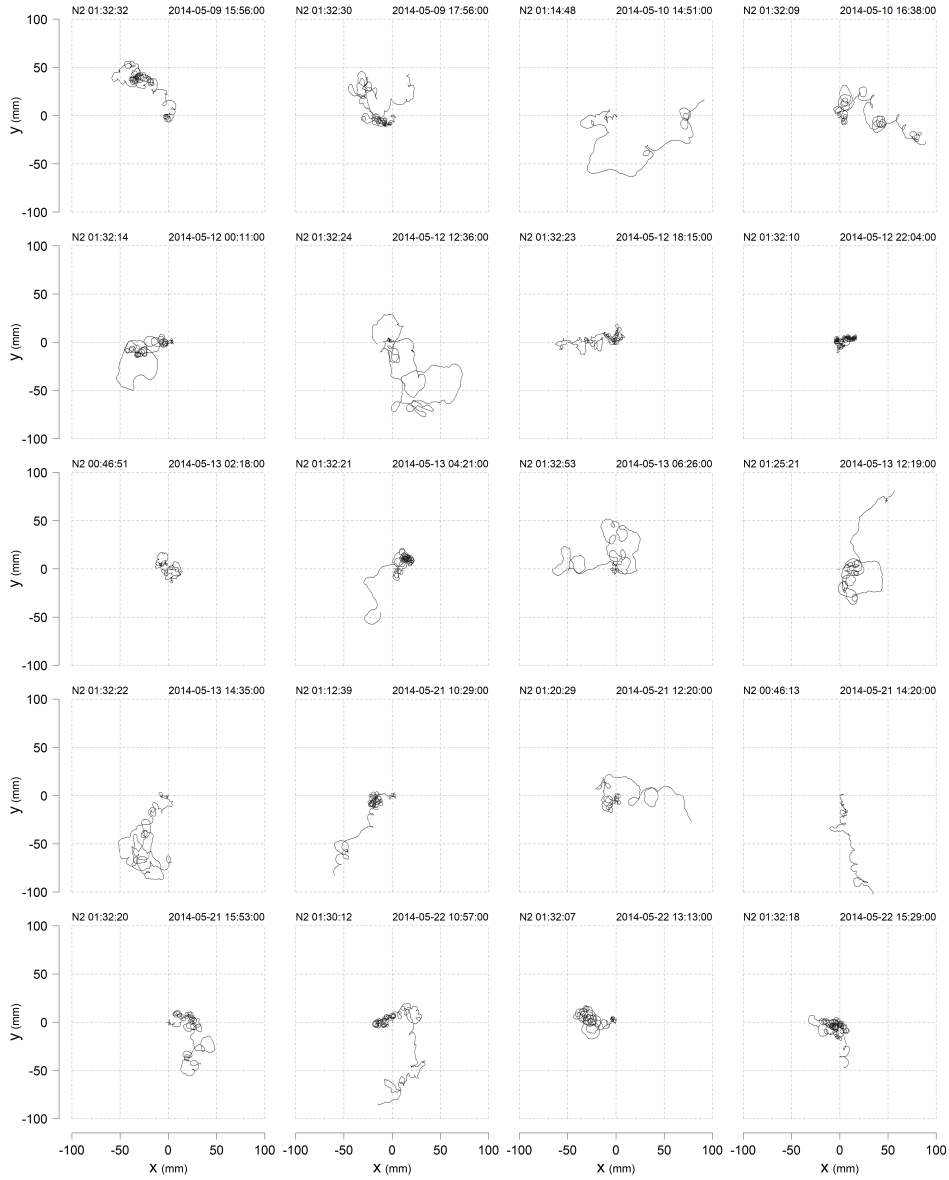


Figure 3.2: Search trajectories of isolated *C. elegans* individuals of the strain N2 during a relocation experiment without food. Prior to relocation, worms were maintained at the assay temperature (21 ± 0.1 °C) and fed *E. coli* OP50 lawns. The size of the experimental arena was 24.5 × 24.5 cm, and the tracks were recorded ad 33Hz during 90 minutes or until the worm reached the edge of the plate. All experiments were conducted by Mia Panlilio under the supervision of William S. Ryu at the Ryu Lab, Department of Physics and the Donnelly Centre, University of Toronto (60 St George St., Toronto, Canada M5S1A).

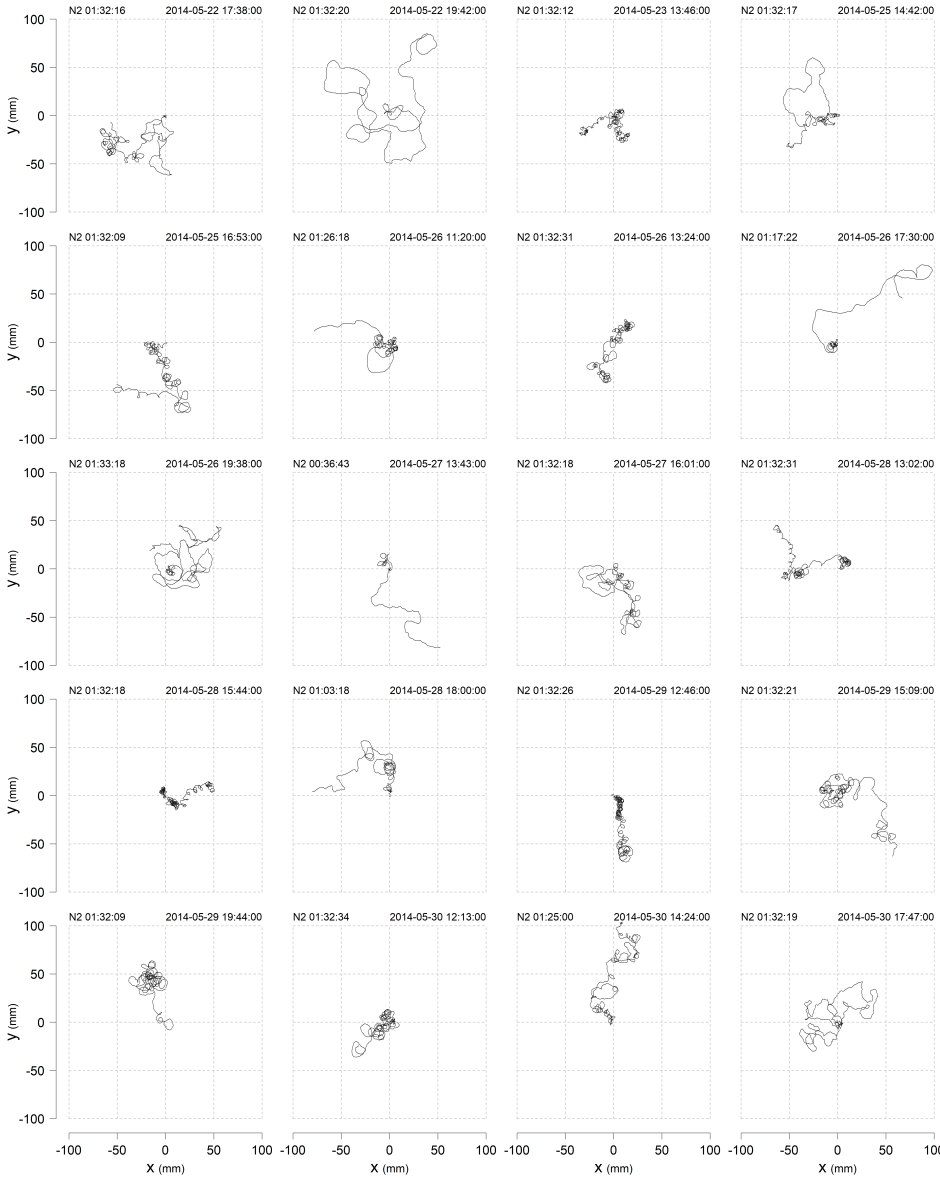


Figure 3.3: Search trajectories of isolated *C. elegans* individuals of the strain N2 during a relocation experiment without food. Prior to relocation, worms were maintained at the assay temperature (21 ± 0.1 °C) and fed *E. coli* OP50 lawns. The size of the experimental arena was 24.5×24.5 cm, and the tracks were recorded ad 33Hz during 90 minutes or until the worm reached the edge of the plate. All experiments were conducted by Mia Panlilio under the supervision of William S. Ryu at the Ryu Lab, Department of Physics and the Donnelly Centre, University of Toronto (60 St George St., Toronto, Canada M5S1A).

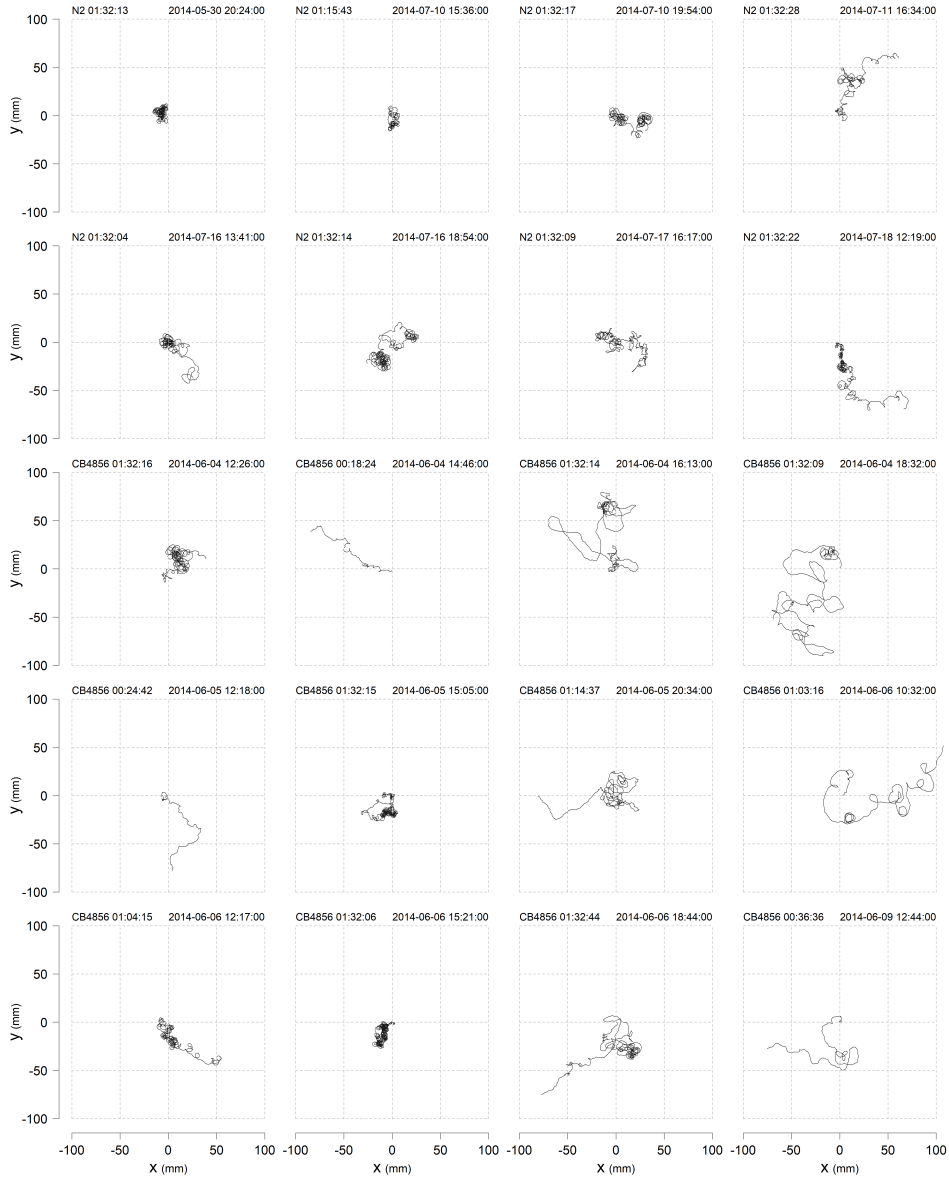


Figure 3.4: Search trajectories of isolated *C. elegans* individuals of the strains N2 and CB4856 during a relocation experiment without food. Prior to relocation, worms were maintained at the assay temperature (21 ± 0.1 °C) and fed *E. coli* OP50 lawns. The size of the experimental arena was 24.5×24.5 cm, and the tracks were recorded ad 33Hz during 90 minutes or until the worm reached the edge of the plate. All experiments were conducted by Mia Panlilio under the supervision of William S. Ryu at the Ryu Lab, Department of Physics and the Donnelly Centre, University of Toronto (60 St George St., Toronto, Canada M5S1A).

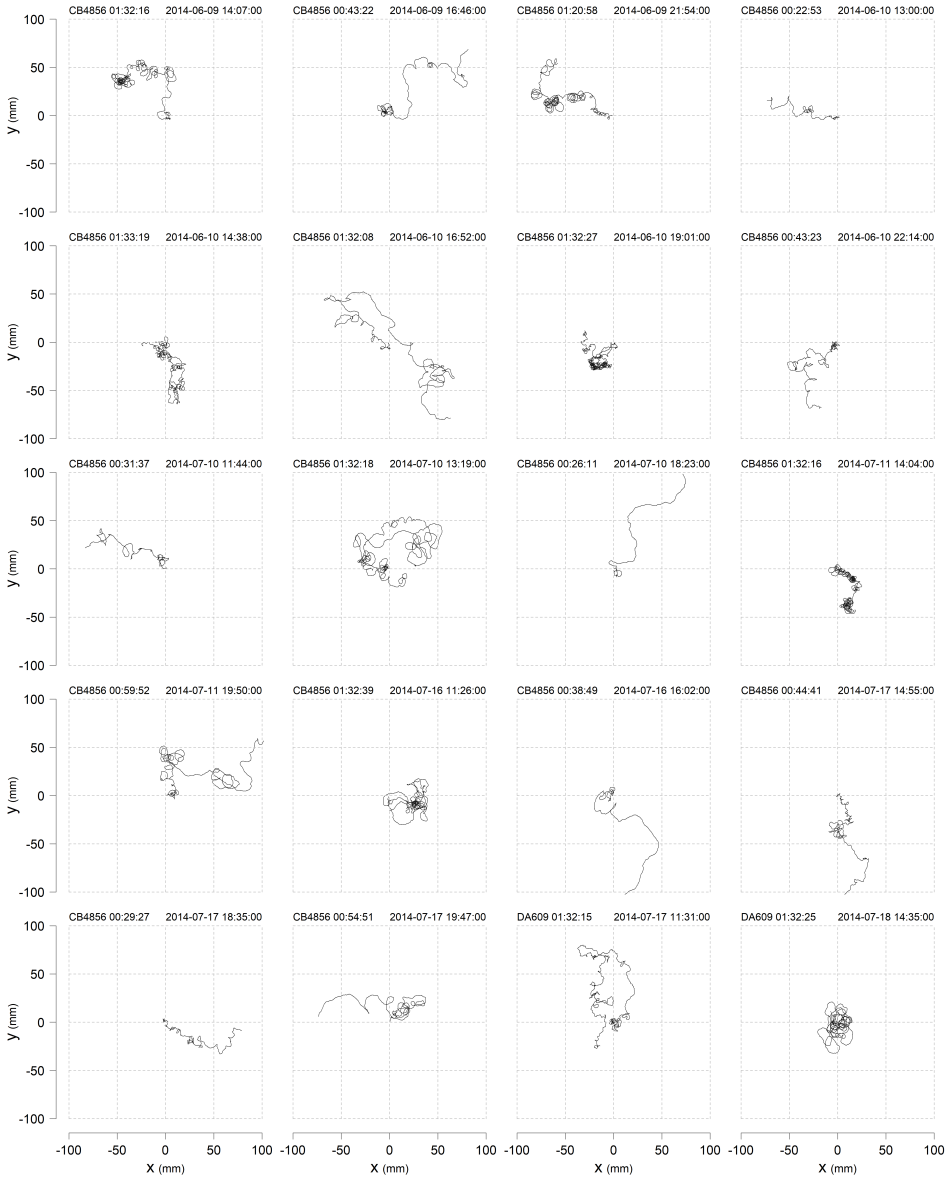


Figure 3.5: Search trajectories of isolated *C. elegans* individuals of the strains CB4856 and DA609 during a relocation experiment without food. Prior to relocation, worms were maintained at the assay temperature (21 ± 0.1 °C) and fed *E. coli* OP50 lawns. The size of the experimental arena was 24.5 × 24.5 cm, and the tracks were recorded ad 33Hz during 90 minutes or until the worm reached the edge of the plate. All experiments were conducted by Mia Panlilio under the supervision of William S. Ryu at the Ryu Lab, Department of Physics and the Donnelly Centre, University of Toronto (60 St George St., Toronto, Canada M5S1A).

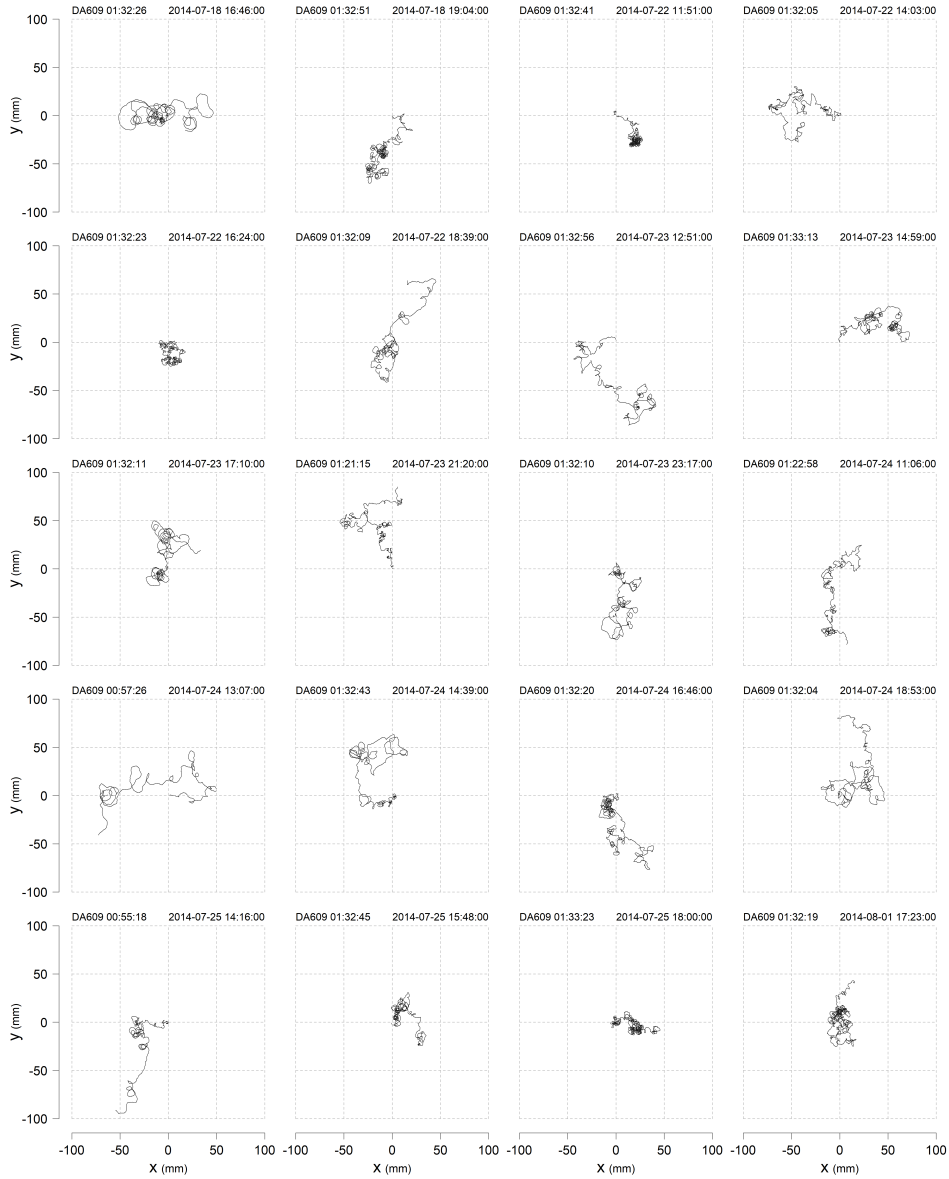


Figure 3.6: Search trajectories of isolated *C. elegans* individuals of the strain DA609 during a relocation experiment without food. Prior to relocation, worms were maintained at the assay temperature (21 ± 0.1 °C) and fed *E. coli* OP50 lawns. The size of the experimental arena was 24.5 × 24.5 cm, and the tracks were recorded ad 33Hz during 90 minutes or until the worm reached the edge of the plate. All experiments were conducted by Mia Panlilio under the supervision of William S. Ryu at the Ryu Lab, Department of Physics and the Donnelly Centre, University of Toronto (60 St George St., Toronto, Canada M5S1A).

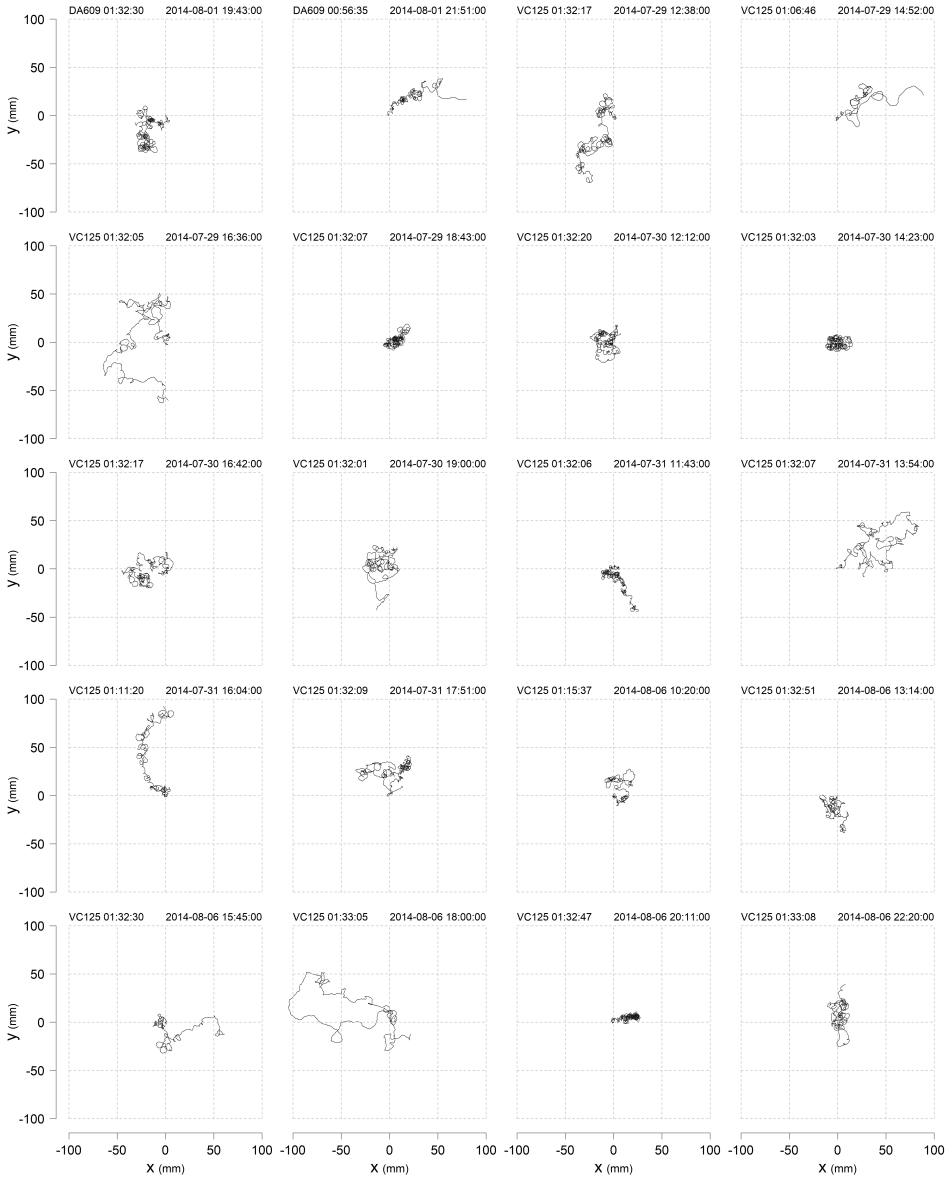


Figure 3.7: Search trajectories of isolated *C. elegans* individuals of the strains DA609 and VC125 during a relocation experiment without food. Prior to relocation, worms were maintained at the assay temperature (21 ± 0.1 °C) and fed *E. coli* OP50 lawns. The size of the experimental arena was 24.5 × 24.5 cm, and the tracks were recorded ad 33Hz during 90 minutes or until the worm reached the edge of the plate. All experiments were conducted by Mia Panlilio under the supervision of William S. Ryu at the Ryu Lab, Department of Physics and the Donnelly Centre, University of Toronto (60 St George St., Toronto, Canada M5S1A).

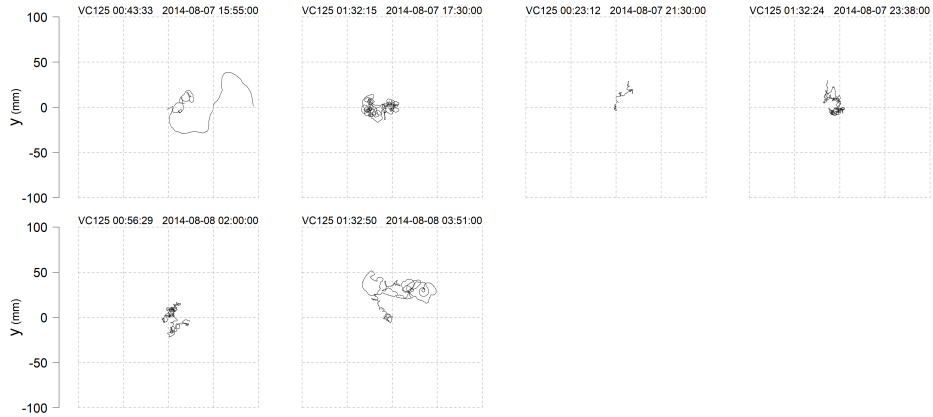


Figure 3.8: Search trajectories of isolated *C. elegans* individuals of the strain VC125 during a relocation experiment without food. Prior to relocation, worms were maintained at the assay temperature (21 ± 0.1 °C) and fed *E. coli* OP50 lawns. The size of the experimental arena was 24.5 × 24.5 cm, and the tracks were recorded ad 33Hz during 90 minutes or until the worm reached the edge of the plate. All experiments were conducted by Mia Panlilio under the supervision of William S. Ryu at the Ryu Lab, Department of Physics and the Donnelly Centre, University of Toronto (60 St George St., Toronto, Canada M5S1A).

Chapter 4

Hurst exponent decomposition of *C. elegans* search trajectories

Abstract

Anomalous diffusion pervades the natural world and has profound influence on how we comprehend transport processes and animal movement behavior. In movement ecology, the crucial consideration lies in determining whether anomalous diffusion persists across relevant behavioral and ecological scales. In such cases, delving into a thorough understanding of the generative mechanisms of anomalous diffusion becomes worthwhile. Here we used the Hurst exponent decomposition method developed by Vilk et al. (2022) to discern sources of anomalous diffusion in *Caenorhabditis elegans* trajectory data. Noteworthy, this method quantifies the relative contribution of (i) time-dependent dynamics, (ii) long-range correlations, and (iii) punctuated extreme events, on the anomalous diffusion exponent of the mean squared displacement (MSD). In essence, this approach provides a mean to quantifying how *C. elegans* trajectories deviate from the conditions conducive to normal diffusion.

4.1 Introduction

The spreading pattern of moving animals is often expressed as a diffusion process (Turchin, 1998; Okubo et al., 2001; Méndez et al., 2014) in terms of the mean square displacement (MSD). This magnitude quantifies the amount of space covered over time by a population of moving objects, in this case animals, and is computed according to:

$$\mathrm{MSD}(t) = \frac{1}{N}\sum_{i=1}^{N} |\mathbf{r}_i(t) - \mathbf{r}_i(0)|^2 \ , \tag{4.1}$$

where N is the number of individuals; $\mathbf{r}_i(t)$ is the position vector of the individual i at time t; and $\mathbf{r}_i(0)$ is the initial position vector of that same individual (Figure 4.1).

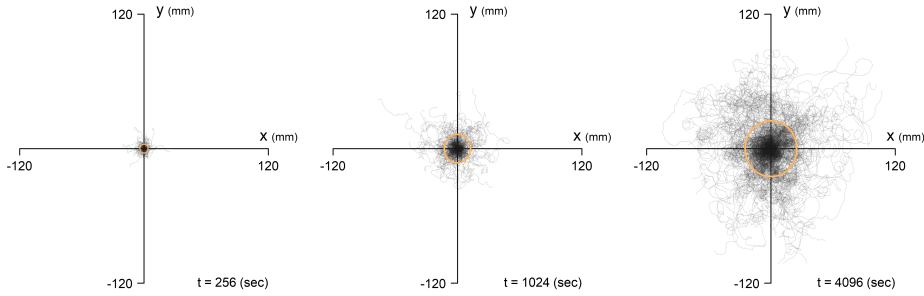


Figure 4.1: Spreading pattern of 126 *C. elegans* trajectories, including wild type (N2, CB4856) and mutant strains (DA609, VC125), at three different times. Orange circles correspond to the MSD, which is considered the area explored by the worm population.

The analysis of the mean square displacement emerged in the context of statistical mechanics and the study of particle motion, from which Einstein (1905); von Smoluchowski (1906) and Langevin (1908), derived that the MSD of free Brownian particles scale linearly with time according to:

$$\mathrm{MSD}(t) = 2nDt \ , \tag{4.2}$$

where D is the diffusion constant and n is the dimension of the space $n = 1, 2, ..., N$. This equation describes the so-called normal diffusion, or Fickian diffusion, and sets the basis to model more complex transport processes involving anomalous diffusion, which occurs when the MSD scales non-linearly with time. In these special cases, Eq. (4.2) is generalized as a power law of the form:

$$\mathrm{MSD}(t) = 4D_\alpha t^\alpha \ , \tag{4.3}$$

where Eq. 4.3 is expressed in a two-dimensional space, α is the anomalous diffusion exponent, and D_α is the generalized diffusion coefficient, of physical dimension $\left[\mathrm{L}^2 \times \mathrm{T}^{-\alpha}\right]$ (Metzler et al., 2014). Alternatively, the same equation is generally expressed in terms of the Hurst exponent according to $\mathrm{MSD}(t) \sim t^{2H}$, where $H = \alpha/2$ (Feder, 1988; Viswanathan et al., 2011; Metzler et al., 2014; Méndez et al., 2014). When $\alpha > 1$ or $H > 0.5$, the system exhibits superdiffusion, meaning the area covered per unit time increases with time. When $\alpha < 1$ or $H < 0.5$, the system exhibits subdiffusion, as the area covered per unit time decreases with time. Finally, when $\alpha = 1$ or $H = 0.5$, we recover the normal diffusion, as the MSD scales linearly with time.

Since Eq. (4.3) is a power law, the most common approach to estimate the parameters α and D_α, is to display the MSD curve in a log-log plot, and fit a simple or piece-wise linear model depending on the observed regimes (Bartumeus et al., 2005). In either case, when Eq. (4.3) is expressed in logarithmic form, the slopes correspond to the scaling exponent α, and the intercepts (in exponential form and divided by 4), to the diffusion coefficient D_α also referred as the effective diffusion coefficient (Helms et al., 2019).

In general, discussions about the MSD are centered on the scaling exponent and the mechanisms enabling anomalous diffusion ($\alpha \neq 1$ or $H \neq 0.5$). On this regard, the study of random walks, first introduced by Pearson (1905), has become a solid theoretical framework for comprehending the emergence of both diffusion and anomalous diffusion processes (Langevin, 1908; Uhlenbeck and Ornstein, 1930; Montroll and Weiss, 1965; Berg, 1983; Viswanathan et al., 2011). Random walk theory has evolved significantly over time and continues to be a cornerstone in the analysis of diverse phenomena, from the seemingly random motion of inanimate particles in a fluid (Langevin, 1908), to the intricate patterns of foraging animals (Viswanathan et al., 2011; Méndez et al., 2014). At the core of this theory lies the central limit theorem (Laplace, 1785; Lyapunov, 1901; Lévy, 1935; Feller, 1935; Fischer, 2011), a fundamental concept demonstrating that any random walk, no matter its complexity, will inevitably converge toward normal diffusion ($\alpha = 1$) if the following conditions are met:

i) A stationary distribution of displacements. Meaning all displacements follow the same probability density function (PDF).

ii) Independence of individual displacements. Indicating each step in the random walk is uncorrelated with the previous and future ones.

iii) Displacements with a finite variance. Implying the distribution of the mean displacement lengths is well-defined and behaved. Therefore, better estimates

of the mean displacement length are obtained as the number of steps in the walk (time) increases.

Conversely, anomalous diffusion emerges when at least one of these conditions is broken. Pure anomalous diffusion occurs when this breakage is permanent, while transient anomalous diffusion occurs when the breakage has a finite duration (Méndez et al., 2014). For instance, continuous time correlated random walks, or simply, persistent random walks (Langevin, 1908; Fürth, 1920; Uhlenbeck and Ornstein, 1930; Méndez et al., 2014), break condition (ii) until the correlation timescale of displacements is surpassed. In consequence, their MSD curves manifest an early transient superdiffusive regime with a relatively sharp transition towards normal diffusion (see Figure 4.2). Likewise, truncated Lévy walks, which are random walks with upper-bounded power-law distributed displacements (Bartumeus et al., 2005; Viswanathan et al., 2011), break conditions (ii) and/or (iii) for varying durations depending on the width and the scaling exponent of their generative power laws. Therefore, their MSD also transitions toward normal diffusion once the influence of their heavy tails stops driving the statistics of the mean displacement length.

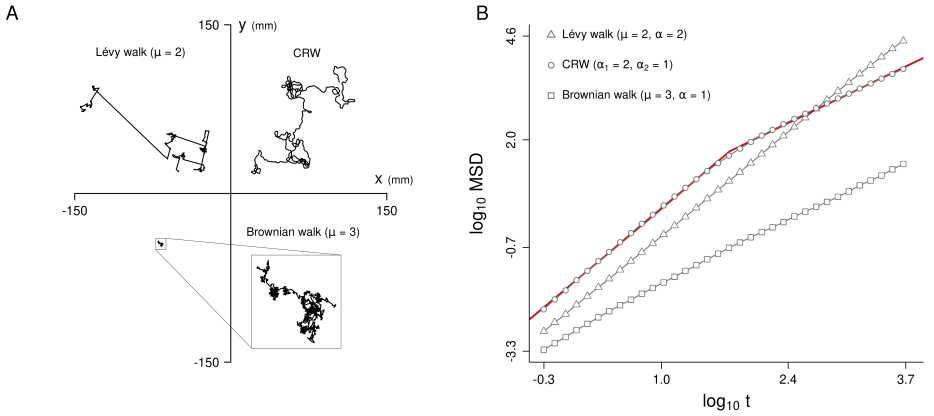


Figure 4.2: Parameter estimation of the mean square displacement (MSD). (A) Illustrative examples of the Lévy walk, the continuous time correlated random walk (Langevin model) and the Brownian walk. (B) MSD curves of the random walks in A with the corresponding linear and piece-wise linear models fitted to estimate the anomalous diffusion exponent α and the effective diffusion coefficient D_α.

Anomalous diffusion, whether in a transient or persistent form, pervades the natural world and has profound influence on how we comprehend transport processes and animal movement behavior. In the realm of biology, it manifests across multiple organizational levels, spanning from single cells to large animals, and its

impact extends to fundamental processes such as gene expression, tumor proliferation and animal migration (Viswanathan et al., 1996; Golding and Cox, 2006; Dieterich et al., 2008; Ariel et al., 2015; Vilk et al., 2022), among other significant examples (Méndez et al., 2014). From the standpoint of movement ecology, the crucial consideration lies in determining whether anomalous diffusion persists across relevant behavioral and ecological scales. Exploring whether anomalous diffusion at these scales carries any adaptive significance becomes pivotal. In such cases, delving into a thorough understanding of the generative mechanisms of anomalous diffusion becomes worthwhile. To that end, the integration of statistical physics frameworks with species-specific biological insights offers a holistic perspective, enabling a nuanced interpretation of animal movement patterns and spreading processes. Nevertheless, comprehending the primary drivers of anomalous diffusion in animal movement data poses a formidable challenge. Limiting factors, such as short trajectories, irregular sampling rates, and the intricate interplay of various behaviors, commonly impede the identification of key mechanisms underlying anomalous diffusion.

To address these challenges, Vilk et al. (2022) have introduced a comprehensive method for discerning potential sources of anomalous diffusion in trajectory data, building upon prior works (Mandelbrot and Wallis, 1968; Chen et al., 2017; Meyer et al., 2017, 2018, 2022; Aghion et al., 2021). Noteworthy, this method quantifies the relative contribution of (i) time-dependent dynamics, (ii) long-range correlations, and (iii) punctuated extreme events, on the anomalous diffusion exponent of the MSD. In essence, this approach provides a mean to quantifying how trajectories deviate from the conditions conducive to normal diffusion, and by extension conducive to the central limit theorem. Namely, being stationary, uncorrelated, and featuring a finite variance displacement lengths.

As an initial step towards comprehending the key components of the *C. elegans* search strategy, we applied this methodology on the *C. elegans* dataset detailed in Chapter 3. This dataset comprised 126 trajectories of both wild type isolates (N2 and CB4856) and defective mutant strains (DA609 and VC125) individuals moving in a virtually unbounded bare arena, after being fed *ad libitum* in a Petri dish scale *Escherichia coli* OP50 food patch. The primary objective of this chapter is obtaining a first overview of the processes driving the spreading of *C. elegans* when searching for food in a food-less and cue-deprived (i.e., lack of gradients) landscape. Specifically, we aimed at characterizing the MSD curves of each strain under study, and determine the relative contribution of (i) non-stationary dynamics, (ii) long-range correlations, and (iii) punctuated extreme events on the overall anomalous diffusion exponent α.

4.2 Materials and Methods

All data used in this chapter were extracted from a set of relocation experiments displacing worms from food to non-food condition, described in Chapter 3. Well-fed *C. elegans* individuals were displaced and gently placed in a bare arena of 24.5×24.5 cm, in which we recorded its movement during the following 90 minutes, or until the animal reached the edge of the plate. In total 126 trajectories of 4 different strains were recorded: N2 ($n = 48$), CB4856 ($n = 30$), DA609 ($n = 24$); and VC125 ($n = 24$).

Trajectories were reconstructed from the center-of-mass of the worms and then smoothed using a third degree polynomial across a 1-second sliding window, like in Stephens et al. (2010). Speed was calculated by taking the first time derivative of the smoothing polynomial and the angular velocity was calculated as the time derivative of the turning angle, computed at the trajectory frame-rate (33Hz). Extreme speed values and consecutive observations with the exact same coordinates, speed or turning angle were assumed to be errors of the tracking system, as the tracker always generated some noise when repositioning the worm. The latter type of errors included zero exact values of speed ($v = 0$), which should not be considered locomotion pauses nor quiescent states as defined in (Costa et al., 2019). Indeed, locomotion pauses occurred occasionally in some individual tracks, but they were associated with fine-scale head or body movements that affected the worm's center-of-mass and thus, caused the speed to be greater than 0. Those pauses were not specifically characterized in our analysis. In total, less than 0.17% of data were filtered out. All analyses that follow were performed at the individual level, unless otherwise stated, and were programmed using custom scripts in R (R Core Team, 2023).

4.2.1 Decomposing anomalous diffusion from trajectory data

As discussed earlier, anomalous diffusion occurs when at least one of the conditions leading to the central limit theorem is not met (Méndez et al., 2014) (Figure 4.3). Expanding on this idea, Mandelbrot and Wallis (1968) introduced two scaling exponents to describe stochastic processes that break conditions (ii), which require independence of individual displacements, and (iii), which require displacements to have a finite variance. The Joseph exponent J indicates a failure of condition (ii), while the Noah exponent L indicates a failure of condition (iii). Specifically Mandelbrot and Wallis (1968) designated as the Joseph effect ($J \neq 0.5$) the fact that periods of high or low precipitation can be extremely long, and by Noah effect ($1/2 \leq L < 1$), the fact that precipitation can be wild and extreme indeed. In the context of random walks, these phenomena can be understood as prolonged

periods of specifically oriented displacements (Joseph effect, Figure 4.3B), and exceptionally fast and long displacements (Noah effect, Figure 4.3C).

Adhering to the biblical nomenclature introduced by Mandelbrot and Wallis (1968), Chen et al. (2017) expanded the characterization of scaling exponents of stochastic processes with the so-called Moses effect that encompasses the failure of condition (i), stating the distribution of displacements must be stationary to satisfy the central limit theorem (Figure 4.3A). By introducing the Moses effect and its associated scaling exponent M, Chen et al. (2017) established the fundamental relation:

$$H = M + L + J - 1, \tag{4.4}$$

where H represents the Hurst exponent, and M, J, and L denote the scaling exponents associated with the Moses, Joseph and Noah effects, respectively. This equation provides a comprehensive mathematical framework for evaluating the impact of (i) non-stationary dynamics, (ii) long-range correlations, and (iii) punctuated extreme events on the overall anomalous diffusion exponent, here expressed in terms of the Hurst exponent H. Consequently, the very same relation applies for the anomalous diffusion exponent of the mean square displacement as: $\alpha = 2\,(M + L + J - 1)$.

4.2.2 Mean square displacement

We computed the mean square displacement as $\mathrm{MSD}(t) = \langle x^2(t) + y^2(t) \rangle$, where x and y are the positions with respect to the origin of coordinates $(0, 0)$, and $\langle \cdot \rangle$ denotes averaging over the ensemble of trajectories of a particular strain.

In all cases, both the anomalous diffusion exponent α and the effective diffusion coefficient D_α, were estimated by fitting a log-log piece-wise linear model with 4 regimes over the MSD curve, similar to Riahi et al. (2019). The fitting procedure was based on least-squares and was implemented using the R package *segmented* (Muggeo, 2003). Although the strain DA609 appeared to exhibit a different number of regimes, we found that 4 was a good compromise between simplicity and accuracy, as it allowed us to: i) obtain reliable and comparable estimates of the anomalous diffusion exponent α, specially on the long-term limit, and ii) capture the complexity of the MSD pattern observed in *C. elegans*. In our case, fitting a smaller amount of diffusive regimes would yield to an overall overestimation of the long-term anomalous diffusion exponent.

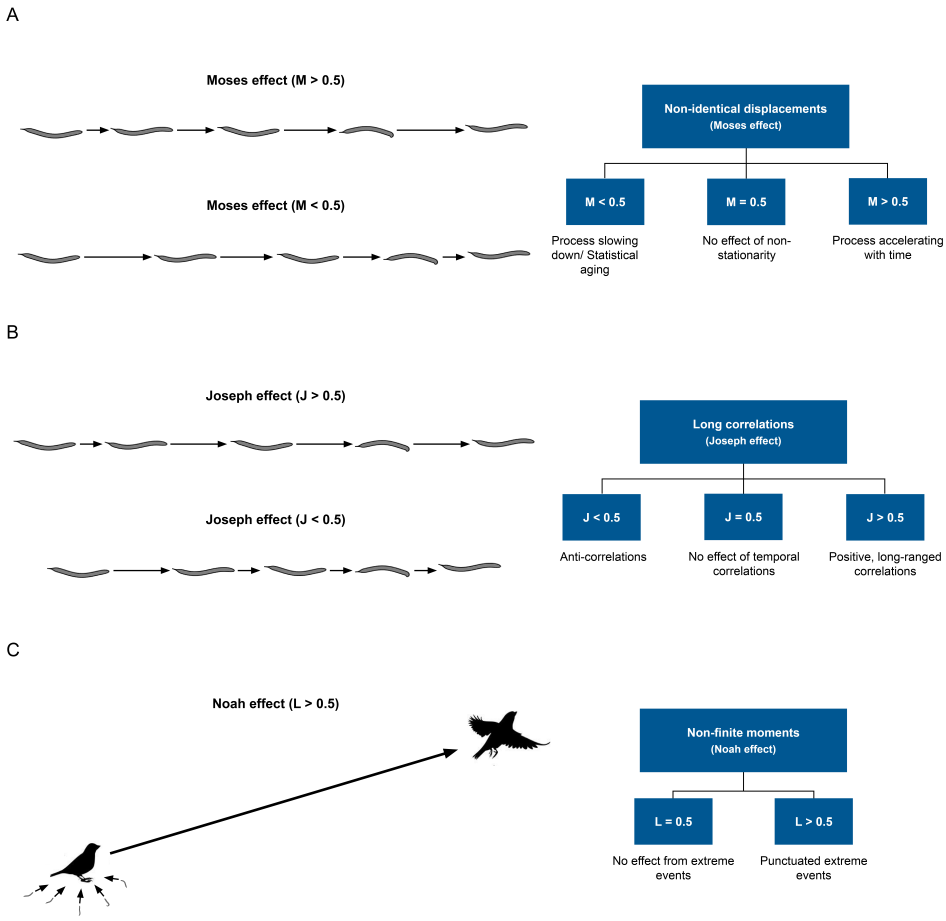


Figure 4.3: Classification of general mechanisms underlying anomalous diffusion, interpreted within the framework of *C. elegans* movement behavior and dispersal. (A) Non-identical displacements, or the Moses effect. Transitioning from local to global search patterns in *C. elegans*, as observed in classical relocation experiments without food, could give rise to this form of anomalous diffusion. (B) Long-correlations, or the Joseph effect. Directed motion and steering control mechanisms may contribute to the emergence of this phenomenon. (C) Displacements with non-finite moments, or the Noah effect. Anomalous diffusion characterized by multiple spatial and temporal scales, where individual displacements can shift the scaling of the mean square displacement (MSD) curve towards either subdiffusion (long pauses) or superdiffusion (rapid large displacements). In the context of *C. elegans* dispersal, transport via vectors such as house sparrows or common pigeons may contribute to the emergence of the Noah effect. The figure is adapted from Méndez et al. (2014) and Vilk et al. (2022). *C. elegans* images were courtesy of Alfonso Pérez-Escudero.

4.2.3 Ensemble time-averaged mean square displacement

Building on Meyer et al. (2022), we estimated the Joseph exponent (J) from the scaling exponent of the ensemble time-averaged mean square displacement (TAMSD). To compute this trajectory observable we followed Metzler et al. (2014), who calculate the TAMSD as $\langle\overline{\delta^2(\tau)}\rangle = \frac{1}{N}\sum_{i=1}^{N}\overline{|\mathbf{r}(t+\tau)-\mathbf{r}(t)|_i^2}$, where $\mathbf{r}$ is the position vector of a given individual i, so that $|\mathbf{r}(t+\tau)-\mathbf{r}(t)|_i^2$ is the squared displacement between two positions separated τ seconds in time, the over-line denotes averaging these squared displacements by τ, and finally N indicate the number of individuals in the ensemble, so that the time-averaged mean square displacements are averaged again at the population level. Here, it is important to note that τ does not represent the absolute timeline, but instead it represents a relative time, or a time lag. Therefore, in non-stationary processes $\mathrm{MSD}(t) \neq \langle\overline{\delta^2(\tau)}\rangle$. Additionally, it is also worth mentioning that the TAMSD needs to be computed using overlapping time windows. Meaning that the squared displacements associated to each τ should be computed point-wise, until reaching the point $t_i + \tau = T$.

4.2.4 Absolute increment displacements

Using methods outlined in Chen et al. (2017), we based the estimation of both the Moses (M) and the Noah (L) exponents on the scaling of the absolute increment displacements with time. To compute the absolute increment displacements of a trajectory, one first needs to calculate the time series of net displacements $d(t) = \sqrt{x^2(t) + y^2(t)}$, where $x(t)$ and $y(t)$ are the coordinates. Therefore, displacements are assumed to start from an origin position $t = 0$ with coordinates (0, 0). Next, discretize $d(t)$ in regular time stamps using a non-overlapping time window of an arbitrary size. In our case we employed a time window of $\Delta = 1$ seconds, and regularized $d(t)$ using linear interpolation (R package *stats*). Finally, we obtained the absolute increment displacements by computing $|\mathbf{d}_i| = |d_{i+1} - d_i|$ (see Figure 4.4), where bold vector notation indicates $\mathbf{d}_i$ can be either positive (departing from the origin) or negative (coming back to the origin), and the subscript i denotes the observation time.

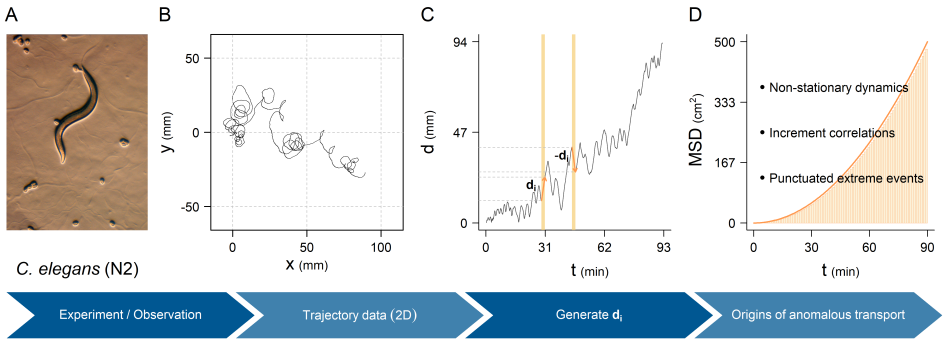


Figure 4.4: Schematic representation of the pipeline used to decompose anomalous diffusion in trajectory data. (A) Experimental design and model organism. Image source: Olga Simón, 17-07-2018, www.olgasimon.com. (B) Data collection and trajectory reconstruction. (C) Data analysis, and computation of the increment displacement $\mathbf{d_i}$. (D) Determining whether the spreading process is (i) time-dependent, (ii) correlated, (iii) prone to exhibit extreme fluctuations. This figure was adapted from Vilk et al. (2022).

4.2.5 Non-stationarity: Moses effect

The Moses effect, quantified through the scaling exponent M, characterizes anomalous diffusion driven by non-stationary or time-dependent dynamics (Chen et al., 2017; Aghion et al., 2021; Vilk et al., 2022). There are two primary methods for estimating M, both mathematically equivalent but potentially yielding different results depending on data structure. In our study, we adopted the original formulation by Chen et al. (2017), which defines M in terms of the median absolute displacement increments:

$$\text{Med}\left[\sum_{i=1}^{t/\Delta} |d_{i+1} - d_i|\right] \propto t^{\,M+1/2}\ , \tag{4.5}$$

where $\text{Med}[\,\cdot\,]$ denotes the median absolute displacement increment evaluated up to observation $i = t/\Delta$; Δ represents the time interval between consecutive observations; and d_i is the distance from the origin to observation i. Notably, the resulting exponent M is sensitive to the choice of Δ. Alternatively, M can be estimated from the absolute spreading velocity $|\mathbf{v}i| = |di + 1 - di|/\Delta$, as proposed by Aghion et al. (2021) and Vilk et al. (2022), (see Appendix A for a detailed description). However, in our data this approach tends to be noisier and may yield less accurate estimates of M.

4.2.6 Long-range correlations: Joseph effect

The Joseph effect characterizes anomalous diffusion arising from long-range correlations. This type of anomalous diffusion is quantified by the scaling exponent J, which can be estimated using various methods, including spreading velocity correlations (Chen et al., 2017; Aghion et al., 2021) or the ensemble time-averaged mean square displacement (TAMSD) (Meyer et al., 2022). Here, we estimated J from the asymptotic behavior of the TAMSD:

$$\langle\overline{\delta^2(\tau)}\rangle = \frac{1}{\mathrm{N}}\sum_{i=1}^{\mathrm{N}}\overline{|\mathbf{r}(t+\tau)-\mathbf{r}(t)|_i^2} \propto \tau^{\,2J}\ , \tag{4.6}$$

where $\mathbf{r}$ is the position vector of the different individuals i, the over-line denotes averaging by τ, and $\langle\,\cdot\,\rangle$ denotes averaging over an ensemble N of trajectories. It is important to remark here that J should be estimated at the long time limit but also assuming $\tau \ll T$, where T is the final time of the trajectory. Therefore, the success of this method might be compromised in short trajectory data. Ultimately, the Joseph effect occurs when $J \neq 1/2$, implying either positive ($J > 1/2$) or negative ($J < 1/2$) long correlations. The alternative approach to estimate J in terms of the spreading velocity correlations can be consulted in Appendix A.

4.2.7 Extreme events: Noah effect

The Noah effect, which is the generative process of anomalous diffusion associated with sudden extreme displacements or pausing times, is characterized by the latent exponent L. Like for the Moses effect we found the methodology proposed by Chen et al. (2017) is the best in terms of goodness of fit when trying to assess this effect in our data:

$$\mathrm{Med}\left[\sum_{i=1}^{t/\Delta}|d_{i+1}-d_i|^2\right] \propto t^{\,2L+2M-1}\ , \tag{4.7}$$

where all the notation is equivalent to Eq. (4.5). By definition $1/2 \leq L < 1$, so that, when decomposing anomalous diffusion, L should be the first exponent to estimate (Vilk et al., 2022). Otherwise, the goodness of fit with data might be compromised provided that L need to accomplish $1/2 \leq L < 1$. As illustrated in Figure 4.3, the Noah effect occurs whenever $L > 1/2$, indicating an abnormal high frequency of extreme events (which could indeed be a single event). An alternative version to estimate L in terms of the spreading velocity can be consulted in Appendix A.

4.2.8 Methods validation

We validated the implementation of the methods outlined from Section 4.2.1 to Section 4.2.7 by decomposing the Hurst exponent of synthetic trajectories generated using random walk models with predefined characteristics. Specifically, we utilized the Brownian walk (BW) as a baseline model, the Lévy walk (LW)(Viswanathan et al., 2011) to verify proper recovery of the Joseph exponent, and the Lévy flight (LF) (Viswanathan et al., 1996) to confirm recovery of the Noah exponent. Detailed descriptions of each model are provided in Appendix A.2.

Table 4.1: Hurst exponent decomposition for random walk models with known anomalous diffusion exponent. The reported regimes, expressed in seconds, are relative to the Noah exponent L, which constrains the entire analysis (Vilk et al., 2022). Therefore, estimates of the Moses M and the Joseph J exponents were obtained conditioned to these regimes. Two estimates of the Hurst exponent are provided: one derived from the fundamental relation $H_p = M + L + J - 1$, and the other directly from the mean square displacement $\mathrm{MSD}(t) \sim t^{2H_\alpha}$. The percentage error between H_p and H_α is calculated as: $100 \times |H_p - H_\alpha|/H_\alpha$. All estimates were obtained from ensembles of $\mathrm{N} = 10^4$ trajectories, except those for the Lévy flight, which were obtained from ensembles of $\mathrm{N} = 10^5$. Parameters used to run these simulations can be consulted in Figures A.1-A.3 from Appendix A.

Model	M	J	L	H_p	H_α	Error
BW	0.50	0.50	0.50	0.50	0.50	0.24%
$\mathrm{LW}_{\mu=1.1}$	0.50	1.00	0.50	1.00	1.00	0.00%
$\mathrm{LW}_{\mu=2}$	0.49	0.91	0.50	0.91	0.91	0.98%
$\mathrm{LW}_{\mu=3.1}$	0.50	0.52	0.50	0.52	0.56	7.36%
$\mathrm{LF}_{\mu=2}$	0.59	0.50	0.91	1.00	0.88	13.66%
$\mathrm{LF}_{\mu=3.1}$	0.50	0.50	0.53	0.53	0.51	4.28%

4.2.9 *C. elegans* inspired Langevin models

The basic Langevin equation

Based on Schienbein and Gruler (1993), we implemented a Langevin equation describing the dynamics of the velocity vector as a toy model of *C. elegans* movement. This model was fully developed in Chapter 6 and reads as follows:

$$\frac{d\mathbf{v}}{dt} = -\gamma(v - \bar{v})\mathbf{e}_h + \sigma_h \mathbf{e}_h + \sigma_\phi \mathbf{e}_\phi \ , \tag{4.8}$$

where γ is a propulsive force pulling the speed toward the mean speed $\bar{v}$, σ_h is Gaussian white noise acting in the direction of motion $\mathbf{e}_h$, and σ_ϕ is Gaussian white noise acting in the normal direction $\mathbf{e}_\phi$.

Langevin with increasing mean speed

Building on the Langevin model of Eq. (4.8), we incorporated non-stationary dynamics in the mean velocity assuming an exponential decay in increasing form, defined as:

$$\bar{v}(t) = V(1 - e^{-\lambda_v t}) \ , \tag{4.9}$$

where $\bar{v}(t)$ is a time-dependent mean speed, V is the maximum speed, and λ_v is the increasing rate of the mean speed. All other model parameters were the same as those shown in Eq. (4.8). The Hurst exponent decomposition of the resulting trajectories is shown in Figure 4.8.

Langevin with diminishing amplitude in angular velocity fluctuations

Similarly, we built a Langevin model with diminishing amplitude in angular velocity fluctuations assuming an exponential decay of the form:

$$\sigma_\phi(t) = \sigma_0 e^{\lambda_\phi t} + c \ , \tag{4.10}$$

where $\sigma_\phi(t)$ is Gaussian white noise acting in the normal direction of the velocity vector, with time-dependent amplitude, σ_0 is the initial amplitude of this Gaussian white noise, λ_ϕ is the decaying rate, and c is a constant (baseline) ensuring the model never reaches ballistic motion, which may occur when $\sigma_\phi = 0$. The Hurst exponent decomposition of the resulting trajectories is shown in Figure 4.9.

4.3 Results

4.3.1 Assessing anomalous diffusion in *C. elegans*

Here we present a comprehensive analysis of *C. elegans* spreading in terms of the mean square displacement (Section 4.2.2) and the Hurst exponent decomposition using methods detailed in Sections 4.2.5-4.2.7.

Mean square displacement

We characterized the mean square displacement (MSD) curves of the two wild types (N2, CB4856), and the two mutant strains (DA609, VC125) under study, estimating both the anomalous diffusion exponent α and the effective diffusion coefficients D_α in four distinct regimes. As shown in Figure 4.5 and reported in Table 4.3, all strains exhibited a complex spreading pattern with multiple transitions between

subdiffusive and superdiffusive regimes. Interestingly, the strain N2 exhibited a long-transient superdiffusive regime ($\alpha = 1.14$) that lasted for approximately one hour, while the strains CB4856 and VC125 ended up in the subdiffusive regime. The anomalous diffusion exponent α obtained for the strain DA609 also indicated long-transient superdiffusion. However, this result should be taken with caution as the slope of the last regime fitted in the MSD of the strain DA609 seem to be rather overestimated (Figure 4.5D).

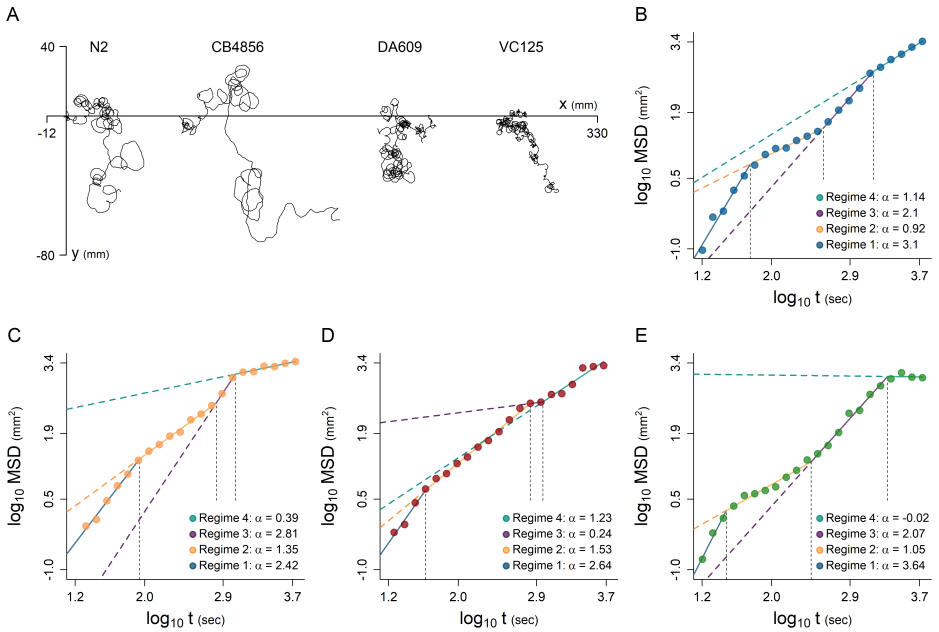


Figure 4.5: Mean square displacement (MSD) analysis of *C. elegans* trajectories recorded during a relocation experiment without food. (A) Illustrative trajectories depicting the movement patterns of the wild-type strains N2 and CB4856, along with defective mutants DA609 and VC125. (B-E) MSD curves for each strain: N2 (blue), CB4856 (orange), DA609 (red), and VC125 (green). In all cases the MSD curve was parameterized by fitting a log-log piece-wise linear model with four distinct regimes, and the estimates of the anomalous diffusion exponent α are reported. Ensemble sizes used for MSD computation were N = 48 (N2), N = 30 (CB4856), N = 24 (DA609), and N = 24 (VC215).

Table 4.2: Characterization of the mean square displacement (MSD) of *C. elegans* trajectories in a relocation experiment where worms where displaced from an environment with food to a bare arena. All estimates were obtained, from strain-averaged ensembles with N2 = 48, CB4856 = 30, DA609 = 24 and VC125 = 24 tracks. In the column Regime, time is expressed in seconds while in the column Duration it is expressed in hours, minutes and seconds. The MSD, predicted at the end of each regime, is expressed in mm^2, and the effective diffusion coefficient D_α in $mm^2\,s^{-1}$. The anomalous diffusion exponent α is dimensionless and is reported with the standard error. Finally, N $\geq$ reports the minimum number of tracks utilized to obtain the estimates of α, and D_α in each regime, as not all tracks started nor ended exactly at the same time.

Strain	Regime	MSD	$\alpha \pm \sigma_\alpha$	D_α	Duration	N $\geq$
N2	$1472 < t \leq 5405$	2638.3	1.14 ± 0.197	$3.5{\cdot}10^{-2}$	01:05:33	36
	$394 < t \leq 1472$	595.8	2.10 ± 0.197	$3.3{\cdot}10^{-5}$	00:17:58	48
	$56 < t \leq 394$	37.5	0.92 ± 0.118	$3.9{\cdot}10^{-2}$	00:05:37	48
	$16 < t \leq 56$	6.3	3.10 ± 0.197	$5.9{\cdot}10^{-6}$	00:00:40	2
CB4856	$1093 < t \leq 5404$	2696.2	0.39 ± 0.083	24.14^*	01:11:52	13
	$661 < t \leq 1093$	1451.7	2.81 ± 0.492	$1.1{\cdot}10^{-6}$	00:31:12	30
	$87 < t \leq 661$	352.9	1.35 ± 0.066	$1.4{\cdot}10^{-2}$	00:04:34	30
	$21 < t \leq 87$	22.8	2.42 ± 0.083	$1.2{\cdot}10^{-4}$	00:00:06	4
DA609	$937 < t \leq 5404$	2693.6	1.23 ± 0.154	$2.1{\cdot}10^{-2}$	01:14:28	21
	$672 < t \leq 937$	372.5	0.24 ± 0.913	18.45^*	00:04:25	24
	$42 < t \leq 672$	344.3	1.53 ± 0.083	$4.0{\cdot}10^{-3}$	00:10:29	23
	$18 < t \leq 42$	5.0	2.64 ± 0.289	$6.3{\cdot}10^{-5}$	00:00:24	2
VC125	$2147 < t \leq 5404$	1269.2	-0.02 ± 0.278	384.01^*	00:54:18	20
	$284 < t \leq 2147$	1295.5	2.07 ± 0.117	$4.1{\cdot}10^{-5}$	00:31:03	23
	$30 < t \leq 284$	19.6	1.05 ± 0.096	$1.3{\cdot}10^{-2}$	00:04:14	22
	$16 < t \leq 30$	1.8	3.64 ± 0.440	$1.9{\cdot}10^{-6}$	00:00:14	2

Note*: these abnormally high effective diffusion coefficients D_α appear in association with abnormally low anomalous diffusion exponents α (see Figures 4.5C-E).

Hurst exponent decomposition

We decomposed the Hurst exponent of the *C. elegans* trajectories in terms of the time scaling of squared increment displacements (Noah exponent L), absolute increment displacements (Moses exponent M) and ensemble time-averaged mean square displacement (Joseph exponent J) (Figure 4.6 and Table 4.3). Of note, the regimes observed in the MSD curves were not equivalent to the regimes observed when decomposing the Hurst exponent. Our results suggest the metrics used for the Hurst decomposition analysis was less variable across time, and converged

faster to specific scaling exponents. Therefore, caution must be taken as the Hurst decomposition framework assumes long-term stabilized MSD curves, and not MSD curves that continuously change the scaling exponent. Despite this, the analysis allows us to detect the Moses ($M > 0.5$) and the Joseph effects ($J > 0.5$) as the main drivers of anomalous diffusion in *C. elegans* (Figure 4.6 and Table 4.3).

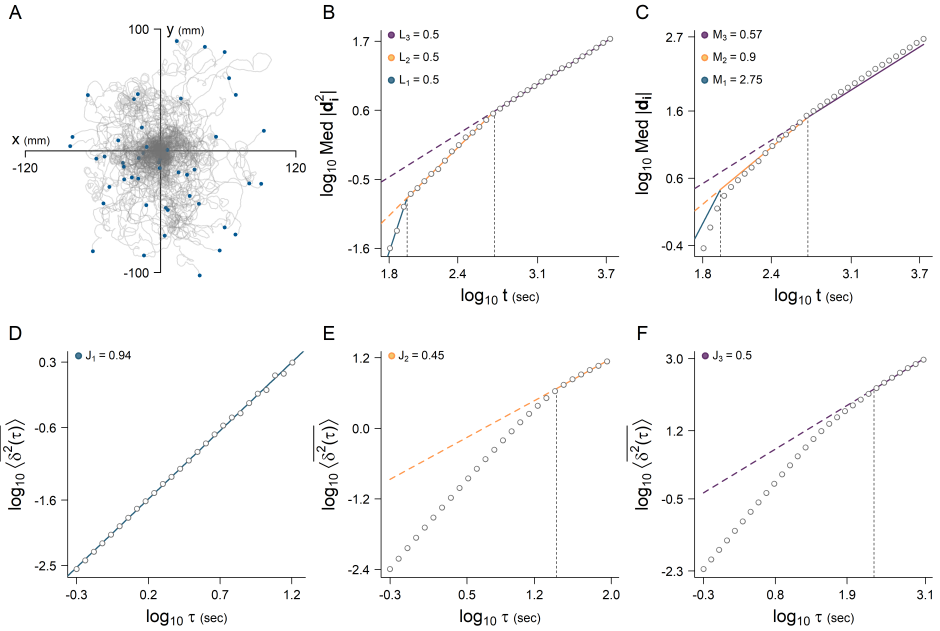


Figure 4.6: Decomposition of the Hurst exponent for *C. elegans* trajectories of the N2 strain recorded during a relocation experiment without food. (A) Ensemble of the 48 trajectories subjected to anomalous decomposition. Blue dots show end of individual trajectories. (B) Estimation of the Noah exponent L (extreme events) based on the time scaling of the median squared displacements (Eq. 4.7). Following (Vilk et al., 2022), regimes found for the Noah effect were utilized to determine both the Moses M and the Joseph J exponents. (C) Estimation of the Moses exponent M (non-stationary dynamics) based on the time scaling of the median absolute increment displacements (Eq. 4.5). It is noteworthy that the goodness of fit of the piece-wise model was constrained by the regimes identified for the Noah effect, depicted in panel B. (D-F) Estimation of the Joseph exponent J (long-range correlations) based on the ensemble time-averaged mean square displacements (TAMSD) (Eq. 4.6). Although the TAMSD refers to a relative time, we respected the regimes found in the data, as the value of M indicated the occurrence of non-stationary dynamics. Estimates of M and L were obtained by fitting a piece-wise linear model with three regimes on absolute increment displacements $|\mathbf{d}_i|$. For J, three regimes were fitted on the TAMSD $\overline{\langle \delta^2(\tau) \rangle}$, and the estimate of J was obtained from the third regime. The results obtained by applying the same protocol to other strains under study i.e., CB4856, DA609 and VC125 can be consulted in Table 4.3.

Table 4.3: Hurst exponent decomposition of the *C. elegans* trajectories under study. The reported regimes, expressed in seconds, are relative to the Noah exponent L, which constrains the entire analysis, as $1/2 \leq L < 1$ (Vilk et al., 2022). Therefore, estimates of the Moses M and the Joseph J exponents were obtained in all cases conditioned to these regimes. Two estimates of the Hurst exponent are provided: one derived from the fundamental relation $H_p = M + L + J - 1$, and the other directly from the mean square displacement $\text{MSD}(t) \sim t^{2H_\alpha}$. The latter estimate is presented only at the long-term limit (the last observed regime), as the regimes identified for the MSD do not align with those identified in the time scaling of squared increment displacements used to estimate L. The percentage error between H_p and H_α is calculated as: $100 \times |H_p - H_\alpha|/H_\alpha$. Finally, N $\geq$ reports the minimum number of tracks utilized to obtain the estimates of M, J, and L in each regime, as not all tracks started nor ended exactly at the same time.

Strain	Regime	M	J	L	H_p	H_α	Error	N $\geq$
N2	$524 < t \leq 5405$	0.57	0.50	0.5	0.56	0.57*	0.33%	36
	$90 < t \leq 524$	0.90	0.45	0.5	0.85	-	-	48
	$64 < t \leq 90$	2.75	0.94	0.5	3.19	-	-	48
CB4856	$463 < t \leq 2353$	0.54	0.54	0.5	0.58	0.19*	$\gg 10\%$	13
	$119 < t \leq 463$	1.10	0.56	0.5	1.13	-	-	30
	$97 < t \leq 119$	5.24	0.91	0.5	5.65	-	-	30
DA609	$472 < t \leq 4705$	0.54	0.51	0.5	0.55	0.61*	2.24%	19
	$117 < t \leq 472$	0.76	0.64	0.5	0.90	-	-	24
	$64 < t \leq 117$	1.93	0.88	0.5	2.31	-	-	24
VC125	$474 < t \leq 4096$	0.61	0.57	0.5	0.68	-0.01*	$\gg 10\%$	18
	$102 < t \leq 474$	0.94	0.53	0.5	0.97	-	-	24
	$74 < t \leq 102$	4.04	0.85	0.5	4.40	-	-	24

Note*: these estimates of the Hurst exponent were obtained from the last observed regime of the mean square displacement curves. Specifically, the reported values are $\alpha/2$ and can be found in Table 4.2 and Figure 4.5.

C. elegans exhibits non-stationary dynamics in elementary movement metrics

To determine whether the Moses effect observed in *C. elegans* trajectories stemmed from non-stationary dynamics in the modulus of the velocity vector (i.e., the speed) and/or changes in its direction, we analyzed the temporal trend of the mean speed and the fluctuations in angular velocity. Specifically, we computed a running average of both speed and angular velocity fluctuations for each trajectory using a time window of 300 seconds. As depicted in Figure 4.7, both metrics exhibited non-stationary trends, particularly at the onset of the trajectories, when the preconditioning experienced by the worms was more recent. However, further research is necessary to ascertain how *C. elegans* modulates these metrics over time, as various mechanisms could explain these results.

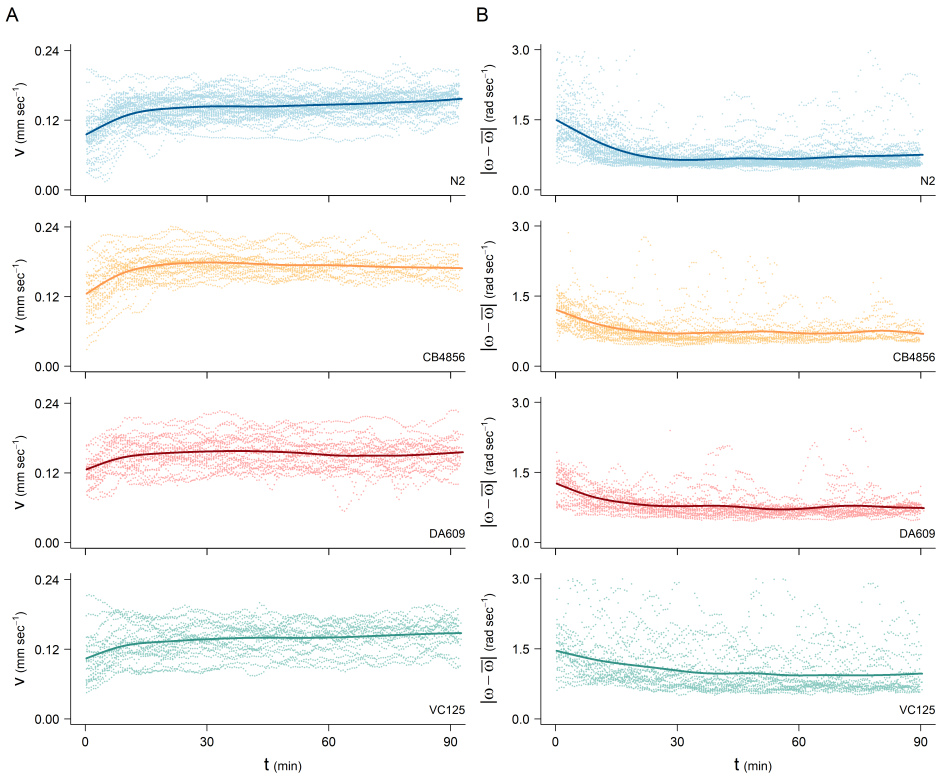


Figure 4.7: Average temporal trends of elementary movement metrics observed in the *C. elegans* trajectories under study. (A) Temporal series of speed. (B) Temporal series of absolute fluctuations around the mean angular velocity. For each metric, the variables were computed at the trajectory frame rate (i.e., 33 Hz) and then roll-averaged with a running window of 300 seconds. This process was applied to each individual trajectory, represented as dotted lines in the plots. Finally, we applied a smoothing spline to depict the overall trend of the variables at the level of strain.

4.3.2 Testing the Hurst exponent decomposition with non-stationary velocity

Finally, we characterized the mean square displacement (MSD) and decomposed the Hurst exponent of the trajectories generated with non-stationary Langevin models described in Section 4.2.9. Our results indicate that non-stationary dynamics in the velocity vector can produce complex MSD curves similar to those observed in *C. elegans* (Figures 4.8 and 4.9). Like with *C. elegans* data, we found that the regimes identified in the MSD did not align with the regimes observed in the time scaling of squared increment displacements, particularly when the non-stationary dynamics occurred solely in the directional component of the velocity

vector (Figure 4.9). Since estimates of both the Noah exponent L and the Moses exponent M are sensitive to the choice of Δ used to compute increment displacements (see Section 4.2.4), we explored a range of different values for Δ: 0.033, 0.33, 3.33, and 33.3. Notably, the best agreement between the MSD and Hurst exponent decomposition was obtained for $\Delta = 0.033$ (Figures 4.8 and 4.9), although it was still relatively poor. These inconsistencies may offer a potential explanation for the disparities observed between the MSD and the Hurst exponent decomposition in empirical data. However, further research is needed to validate this hypothesis and understand how *C. elegans* is able to generate non-stationary dynamics in the velocity vector.

Table 4.4: Hurst exponent decomposition for the two non-stationary Langevin models used to explore the phenomenological origin of anomalous diffusion in *C. elegans*. $\mathrm{LN}_{v(t)}$ incorporates an increasing trend in the mean speed, while $\mathrm{LN}_{\sigma_\phi(t)}$ incorporates a decreasing trend in angular velocity fluctuations as reported in Section 4.2.9. Regimes are expressed in seconds and are relative to the Noah exponent L (Vilk et al., 2022). Therefore, estimates of the Moses M and the Joseph J exponents were obtained in all cases conditioned to these regimes. Two estimates of the Hurst exponent are provided: one derived from the fundamental relation $H_p = M + L + J - 1$, and the other directly from the mean square displacement $\mathrm{MSD}(t) \sim t^{2H_\alpha}$. The percentage error between H_p and H_α is calculated as: $100 \times |H_p - H_\alpha|/H_\alpha$. All estimates were obtained from ensembles of $\mathrm{N} = 10^4$ trajectories.

Model	Regime (sec)	M	J	L	H_p	H_α	Error
$\mathrm{LN}_{v(t)}$	$1680 < t \leq 5405$	0.71	0.62	0.5	0.83	0.92	9.31%
	$208 < t \leq 1680$	1.04	0.71	0.5	1.25	-	-
	$64 < t \leq 208$	0.71	0.58	0.5	0.80	-	-
$\mathrm{LN}_{\sigma_\phi(t)}$	$64 < t \leq 5405$	0.50	0.68	0.5	0.69	1.34	$\gg 10\%$

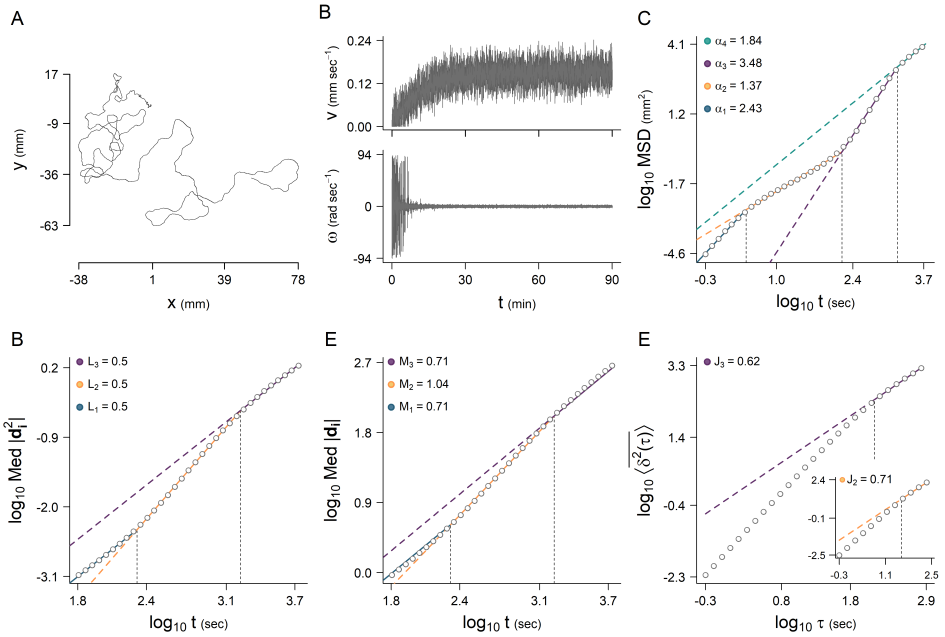


Figure 4.8: Hurst exponent decomposition of the Langevin model with non-stationary mean speed. (A) Illustrative example of the trajectories outputted by the model. (B) Temporal series of speed (upper panel) and angular velocity (lower panel) of the trajectory shown in panel A. (C) Mean square displacement (MSD) with the anomalous diffusion exponent α. (D) Time scaling of squared increment displacements with the Noah exponent L. Note the disparity with the regimes found in the MSD curve. (E) Time scaling of absolute increment displacements with the Moses exponent M. (F) Ensemble time-averaged mean square displacement with the Joseph exponents J fitted at the long term regime. All estimates were obtained from ensembles of $\mathrm{N} = 10^4$ trajectories of 90 minutes length. The trajectories were generated using the following parameters: (i) propulsive force $\gamma = 0.71$ sec^{-1}, (ii) maximum speed $V = 0.15$ mm $\cdot$ sec^{-1}, (iii) increasing rate of the mean speed $\lambda_v = 0.0015$, (iv) intensity of Gaussian white noise in the direction of motion $\sigma_h = 0.03$, and (v) intensity of Gaussian white noise in the normal direction $\sigma_\phi = 0.025$.

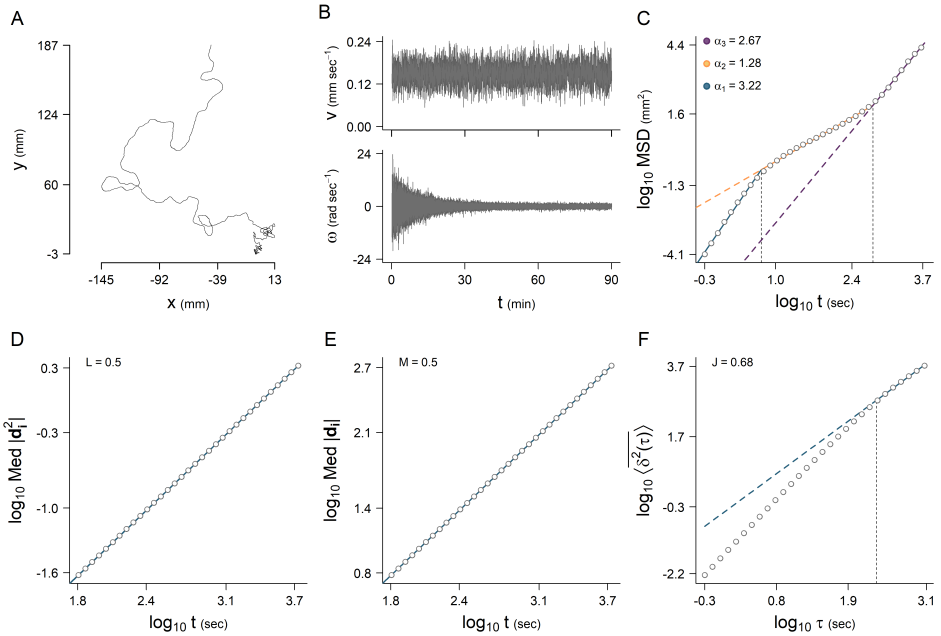


Figure 4.9: Hurst exponent decomposition of the Langevin model with non-stationary angular velocity fluctuations. (A) Illustrative example of the trajectories outputted by the model. (B) Temporal series of speed (upper panel) and angular velocity (lower panel) of the trajectory shown in panel A. (C) Mean square displacement (MSD) with the anomalous diffusion exponent α. (D) Time scaling of squared increment displacements with the Noah exponent L. Note the disparity with the regimes found in the MSD curve. (E) Time scaling of absolute increment displacements with the Moses exponent M. (F) Ensemble time-averaged mean square displacement, with the Joseph exponent J fitted at the long term regime. All estimates were obtained from ensembles of $\mathrm{N} = 10^4$ trajectories of 90 minutes length. These trajectories were generated using the following parameters: (i) propulsive force $\gamma = 0.71$ sec^{-1}, (ii) mean speed $\bar{v} = 0.15$ mm $\cdot$ sec^{-1}, (iii) intensity of Gaussian white noise in the direction of motion $\sigma_h = 0.03$, (iv) initial amplitude of the Gaussian acting in the normal direction $\sigma_0 = 0.15$, (v) decaying rate of the noise $\lambda_\phi = 0.0015$, and (vi) baseline intensity of the noise $c = 0.015$.

4.4 Discussion

We conducted a comprehensive analysis of the diffusive properties of *C. elegans* strains (N2, CB4856, DA609, VC125) in the context of search, worms being displaced from a food patch to a bare arena. The direct examination of the MSD curve revealed complex spreading patterns in a cue less environment. Over time, multiple spreading transitions occurred, depicting changes in the anomalous diffusion exponents, leading to shifts between superdiffusive, diffusive, and even subdiffusive regimes. Despite we can discretize the MSD curve in about 4 regimes it looks as if

the scaling exponent of the MSD in *C. elegans* appears to be a continuous function of time rather than adhering to a few established diffusive regimes.

In addition, when decomposing the Hurst exponent using the methods proposed by Aghion et al. (2021) and Vilk et al. (2022) we only could observe three distinct diffusive regimes for the different metrics used (i.e., absolute increment displacements and TAMSD) in order to understand potential generative processes: non-stationarity (Moses effect), long-range correlations (Joseph effect), extreme events (Noah effect). Part of the problem is that estimates of the scaling exponent α requires a large enough segment of data, and if one can guess there could be several exponents α involved in the spreading process one needs to look for a method to decide where are these segments or diffusive regimes. The other part of the problem is that in complex, non-stationary processes these exponents may be continuously changing across time despite we only can compute them segmenting or discretizing the MSD curves. Therefore, the scaling exponents can only be approximated, and maybe they not represent the ones that should be represented in the Hurst analysis framework and equations. Thus, depending on how measurements are done and the specific structure of the empirical data, anomalous diffusion revealed different patterns that do not need to be coincident. Amidst all this complexity, two conclusions were clear: (i) *C. elegans* trajectories under study showed superdiffusion and (ii) both non-stationarity dynamics ($M > 0.5$) and long range correlations ($J > 0.5$) in spreading velocity could explain such superdiffusion.

Building upon the literature on anomalous diffusion(Metzler et al., 2014; Vilk et al., 2022) and *C. elegans* locomotion (Hills et al., 2004; López-Cruz et al., 2019), one might hypothesize that (i) the Moses effect (non-stationarity) is associated with the temporal evolution of sharp turns and crawls (Pierce-Shimomura et al., 1999; Salvador et al., 2014), and (ii) the Joseph effect (long-range correlations) could arise either from a similar time-dependent mechanism or through stationary fractal-like reorientation patterns (Bartumeus and Levin, 2008). To test these hypotheses and unravel the primary source of anomalous diffusion we need to identify first sharp turns and crawls, and get estimates of their respective duration and time allocation across the trajectory. In other words, to get a more comprehensive understanding of anomalous diffusion one requires a much more explicit analysis of turning dynamics across scales.

Of note, our analysis is subject to a very specific experimental condition. What we observe is always connected to the specific ecological response of the worms in order to initiate a search process in an agar plate without food nor gradients, being previously displaced from a food patch. In real ecological setups, for example, *C. elegans* may reflect similar diffusive anomalies but maybe activating other mechanisms, or even may show new anomalies not observed in our data. For exam-

ple, worms may be induced to produce large-scale migration or homing behavior (Joseph effect) if gradients are available. Also worm movement could even show Noah effects (extreme jump events), as it is known they can travel in bird legs in dauer state (Frézal and Félix, 2015; Schulenburg and Félix, 2017). The impact of these type of phenomena in generating anomalous diffusion may depend on the frequency and scales involved in such a passive transport media. As another example, Vilk et al. (2022), reported super-diffusion and sub-diffusion in wintering and breading storks, respectively, both mediated by Noah effects.

Chapter 5

Temporal unfolding of sharp turns and crawls cause the non-stationary spreading of *C. elegans* search

Abstract

C. elegans movement involves the combination of two elementary motor behaviors: sharp turns and crawls. Sharp turns are characterized by a rapid change in the direction of motion while crawls represent sinuous and continuous forward movement between turns. In absence of food, the interplay of these motor behaviors gives rise to two long-lasting locomotion states named local search and global search, each persisting over scales of minutes up to one hour. Typically, the local searching state occurs in proximity to food sources, manifesting area-restricted search patterns, while the global search takes place away from food sources, accompanied by exploratory and long-range dispersal patterns. Despite being easily discernible by the human eye, detecting sharp turns and crawls in trajectories poses a considerable challenge. In response to this, we present a data-driven method that characterizes *Caenorhabditis elegans* trajectories into sharp turns and crawls based on speed. Our method recovers well-known features of these two motor behaviors, such as their distribution of duration or its temporal unfolding, and also facilitates the characterization of other relevant features, such as their associated turning angle distribution. A detailed description of *C. elegans* locomotion, grounded in the analysis of sharp turns and crawls, is crucial for understanding the origin of the anomalous diffusion dynamics it unfolds in absence of food and environmental gradients. This study validates the high accuracy of our method in detecting the two motor behaviors and offers a versatile tool for annotating animal trajectories based on straightforward movement metrics.

5.1 Introduction

C. elegans locomotion is a subject of interest in a great variety of fields, including genetics, neurobiology, and ecology (Zheng et al., 1999; Spanier et al., 2005; Roberts et al., 2016; Pradhan et al., 2019; Cermak et al., 2020). On mid 70s Croll (1975) laid a solid groundwork for the topic by defining the elementary turning behaviors of *C. elegans*: the Ω turns and the reversals, which consist of a deep body bend resembling the Greek letter Ω and a backward displacement of about 3 seconds, respectively. Building upon this groundwork, Pierce-Shimomura et al. (1999) expanded the understanding of *C. elegans* locomotion by introducing the term "sharp turn". This encompassed Ω turns and reversals, alongside a newly identified behavior termed "pirouette", characterized by paired Ω turns and reversals. Additionally, they introduced the term "runs" or "crawls" (Salvador et al., 2014), referring to the forward displacements occurring between sharp turns. Since then, this categorization provided a valuable framework for dissecting the intricacies of *C. elegans* locomotion across various behaviors, and has laid the foundations for subsequent studies in the field.

Sharp turns and crawls, are the building blocks of complex locomotion states that *C. elegans* unfolds to explore its surrounding environment. For example, when removed from food, *C. elegans* engages a stereotyped sequence of two locomotion states, named local search and global search (Hills et al., 2004; Gray et al., 2005; Flavell et al., 2020), that mainly differ in their proportion of sharp turns to crawls (Calhoun et al., 2014; Salvador et al., 2014; Klein et al., 2017). Right after food removal the animals execute a high frequency of sharp turns that enables them to actively reduce spreading and carry out a thorough local search (Hills et al., 2004; Wakabayashi et al., 2004; Gray et al., 2005; López-Cruz et al., 2019). During the subsequent 15-30 minutes, if no food is encountered, the frequency of sharp turns decreases and the duration of crawls increases, promoting the transition towards the so-called global search, or dispersal state (Hills et al., 2004; Gray et al., 2005; López-Cruz et al., 2019). Notably, while sharp turns are less frequent during the dispersal state, they still occur intermittently and contribute to the emergence of a fat-tailed distribution of inter-reorientation times (Salvador et al., 2014). Such distributions are anticipated by random search theory and are considered a key property of efficient exploratory patterns in non-informed searches (Bartumeus et al., 2005; Bartumeus and Levin, 2008; Méndez et al., 2014; Bartumeus et al., 2016).

The acquisition of high-resolution movement data through worm-trackers has revolutionized the study of *C. elegans* locomotion, turning it into a compelling quantitative subject for laboratory analysis. The tracking of worms is generally

conducted within the context of relocation experiments, where animals are transferred from plates under standard conditions with unlimited food (Brenner, 1974), to experimental plates with specific treatment conditions, such as variations in the presence of food or chemical gradients. This approach has yielded valuable insights into how *C. elegans* modulates sharp turns and crawls when foraging or searching, and how these two motor behaviors influence its overall spreading patterns. Noteworthy contributions to these topics have been made by Hills et al. (2004); Gray et al. (2005); Stephens et al. (2008); Brown et al. (2013); Calhoun et al. (2014); Salvador et al. (2014); Klein et al. (2017); López-Cruz et al. (2019); Pradhan et al. (2019), and Helms et al. (2019). However, and despite these advancements, accurately detecting sharp turns and crawls in trajectory data remains a significant challenge, making it difficult to comprehensively characterize their locomotive properties. Several methods, including supervised annotation, hidden Markov models and posture analysis have been used to address this problem, each of them offering varying advantages and limitations (Pierce-Shimomura et al., 1999, 2005; Peliti et al., 2013; Stephens et al., 2008, 2010, 2011; Cermak et al., 2020).

Here, we propose a straightforward data-driven approach that leverages the analysis of speed as a foundation for annotating *C. elegans* trajectories into sharp turns and crawls. Speed serves as a crucial observable and is easily computed from high-resolution trajectory data. Importantly, our approach steers clear of restrictive mathematical assumptions, making it potentially applicable for annotating high-resolution movement trajectories across different animal species and in various classification contexts. Following the resulting annotation, we characterized the locomotive properties of sharp turns and crawls by modeling their associated distributions of speed and angular velocity and their temporal unfolding.

5.2 Materials and Methods

All data used in this chapter were extracted from relocation experiments described in Experimental data (Chapter 3). These experiments consisted on placing well-fed *C. elegans* individuals in a bare arena of 24.5×24.5 cm, and record its movement during the following 90 minutes, or until the animal reached the edge of the plate. In total 126 trajectories of 4 different strains were recorded: N2 ($n = 48$), CB4856 ($n = 30$), DA609 ($n = 24$); and VC125 ($n = 24$).

Trajectories were reconstructed from the center-of-mass of the worms and then smoothed following Stephens et al. (2010). Speed was calculated as the time derivative of the smoothed positions and the angular velocity was calculated as the time derivative of the turning angle, computed at the trajectory frame-rate (33Hz). Extreme speed values and consecutive observations with the exact same coordinates,

speed or turning angle were assumed to be errors of the tracking system, as the tracker always generated some noise when repositioning the worm. The latter type of errors included zero exact values of speed ($v = 0$), which should not be considered locomotion pauses as defined in (Costa et al., 2019). Indeed, locomotion pauses occurred occasionally in some individual tracks, but they were not specifically characterized in our analysis. In total, less than 0.17% of data were filtered out.

Probability density functions of speed, angular velocity and duration associated with sharp turns and crawls were fitted using maximum likelihood estimation (R package *bmle* (Bolker and R Development Core Team, 2023)). Model selection was performed using the Akaike Information Criteria AIC (R packages *bmle*, *stats* (Bolker and R Development Core Team, 2023; Akaike, 1974; Burnham and Anderson, 2002)). In all cases, statistical tests were done with a significance level $\alpha = 0.05$. Comparison analyses were made using t-tests, or pairwise t-tests (R package *rstatix*). For multiple group comparisons p-values were adjusted using the Holm-Bonferroni correction (Holm, 1979). When needed to meet tests' assumptions we followed the Inter Quartile Range criteria (Hoaglin, 2003) and treated as outliers all data points falling below $Q_1(x) - (1.5 \cdot \mathrm{IQR}(x))$ or above $Q_3(x) + (1.5 \cdot \mathrm{IQR}(x))$. All analyses that follow were performed at the individual level, unless otherwise stated, and were programmed using custom scripts in R (R Core Team, 2023).

5.3 Results

5.3.1 Data exploratory analysis

Similar to Pierce-Shimomura et al. (1999), the distributions of speed and angular velocity of the trajectories were non-Gaussian and heavy tailed (Figure 5.1A, B). Speed was right-skewed with a population average value of $\langle v \rangle = 0.15\,\mathrm{mm} \cdot \mathrm{sec}^{-1}$ (SE = 0.002, N = 126) and a prominent fat tail of low values. On the contrary, angular velocity was symmetric, with an absolute average value of $\langle \omega \rangle = |0.04|\,\mathrm{rad} \cdot \mathrm{sec}^{-1}$ (SE = 0.002, N = 126) and two long tails that were strongly associated with low speeds (Figure 5.1C). These low speeds and high angular velocities occurred in short bursts of about 10 seconds, instead of being long-lasting events. They appeared throughout the trajectories but more frequently at the first 15-30 minutes, echoing the patterning of Ω turns, reversals and pirouettes typically observed in *C. elegans* relocation experiments (Gray et al., 2005; Calhoun et al., 2014; Klein et al., 2017; López-Cruz et al., 2019)(Figure 5.1D-H).

Based on this fact and existing literature on *C. elegans* locomotion (Croll, 1975; Pierce-Shimomura et al., 1999; Stephens et al., 2008, 2010, 2011; Salvador et al.,

2014), we inferred the observed distributions of speed and angular velocity were reflecting the twofold nature of *C. elegans* motor behavior, comprised of sharp turns and crawls. In particular, we hypothesized that sharp turns would be linked to low speed and high angular velocity (in absolute value), and crawls would be associated with high speed and low angular velocity (in absolute value).

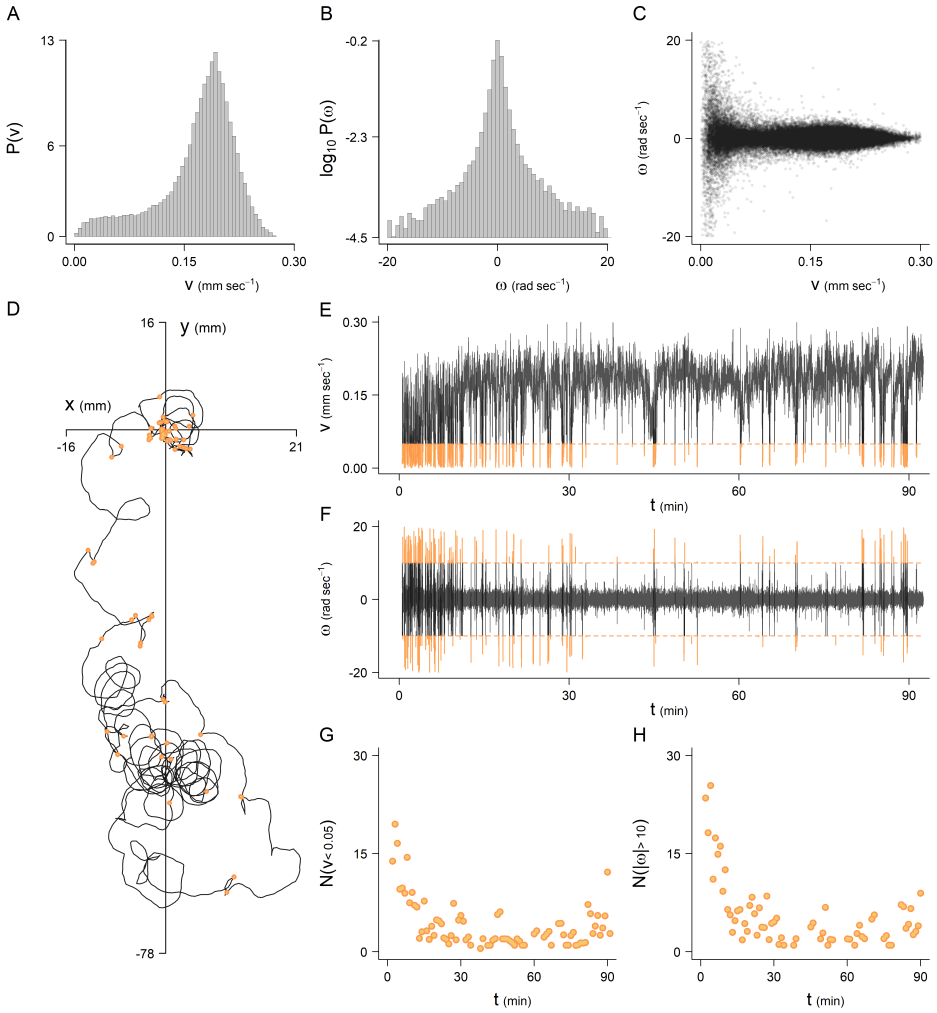


Figure 5.1: Illustrative example of the data exploratory analysis applied to each individual trajectory. (A, B) Histograms of speed and angular velocity. (C) Scatter plot of speed against angular velocity. (D) Trajectory of the analyzed individual (strain N2) with low speeds: $v < 0.05$, and high angular velocities: $|\omega| > 10$, depicted in orange. (E, F) Temporal series of speed and angular velocity. (G, H) Temporal unfolding of low speeds and high angular velocities.

5.3.2 Speed distribution

Based on the idea that speed reflected the key features of both sharp turns and crawls we characterized its probability density function (PDF) as a mixture of two distributions. One right-skewed distribution accounting for the low speeds associated with sharp turns, and one normal distribution accounting for high-cruising speeds associated with crawls (Figure 5.2A). We tested various mixtures of a right-skewed distribution, including Weibull, beta, gamma, and exponential, with a normal distribution (see analytical expressions in Appendix B.1), fitted the mixed probabilistic models using maximum likelihood estimation, and performed model selection based on AIC. The Weibull-normal distribution showed the best fit with data in 90.5% of the individuals ($n = 114$). The beta-normal distribution was best in 6.4% ($n = 8$), and the gamma-normal was best in the remaining 3.1% ($n = 4$; Figure 5.2B).

Given that model selection was not 100% consistent across individuals, we aimed to assess the difference between the first four moments of the fitted models. Specifically, we compared their mean, variance, skewness, and kurtosis. Pair-wise t-tests indicated only the variance and kurtosis of the exponential-normal fits were significantly different from the other fitted models ($\alpha = 0.05$; Figure 5.2C-F). Consequently, we opted to model the PDF of speed for all individuals under study as a Weibull-normal distribution, defined as:

$$f(v) = p_w \underbrace{\left(\frac{\varphi}{\theta} \left(\frac{v}{\theta} \right)^{(\varphi-1)} e^{-(\frac{v}{\theta})^{\varphi}} \right)}_{\text{Weibull } (\mathcal{W})} + p_n \underbrace{\left(\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{v-\mu}{\sigma})^2} \right)}_{\text{Normal } (\mathcal{N})}, \qquad (5.1)$$

where p_w, p_n are the probability weights of the Weibull ($\mathcal{W}$) and the normal ($\mathcal{N}$) component, respectively; thereby, $p_w + p_n = 1$; φ, θ are the shape and the scale of the Weibull; and μ, σ are the mean and the standard deviation of the normal.

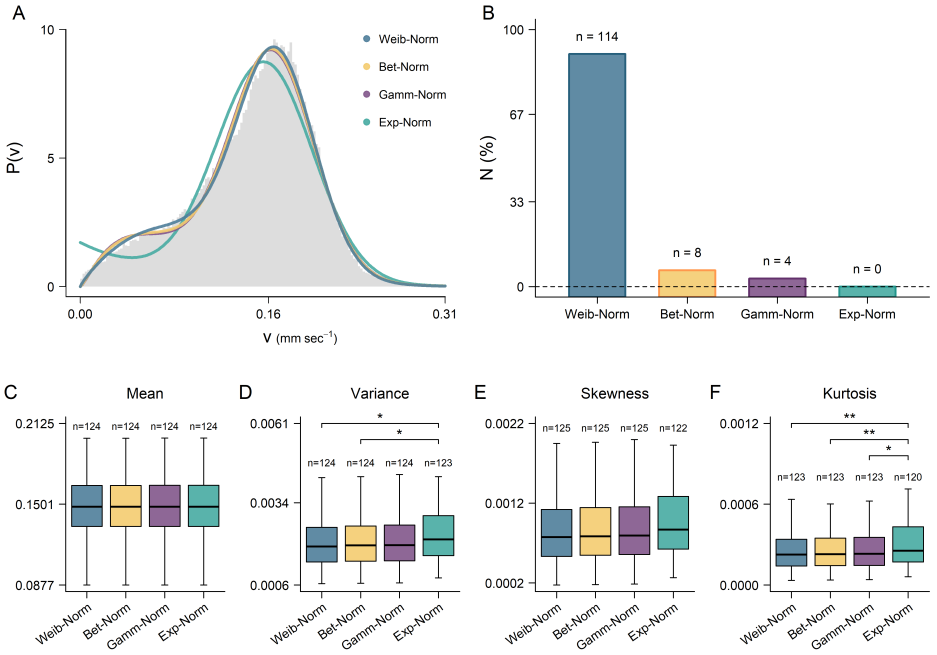


Figure 5.2: Characterization of the speed distribution. (A) Representative example of the fitted models for an N2 individual. Legend names indicate Weibull-normal, beta-normal, gamma-normal and exponential-normal. (B) Model selection for the whole population (N = 126). (C-F) Boxplots comparing the moments of the fitted probabilistic models. Significant differences are denoted with starred braces ($p < 0.01^{**}$, $p < 0.05^{*}$).

Table 5.1: Strain-averaged parameters of the Weibull-normal distribution of speed, measured in mm sec^{-1}. The columns labeled p_w, θ, and φ report values for the probability weight, scale, and shape parameters of the Weibull component, respectively. Similarly, columns p_n, μ, and σ report the values for the probability weight, mean, and standard deviation of the normal component. Finally, N indicates the number of individuals used to compute these average values.

	Weibull (sharp turns)			Normal (crawls)			
Strain	p_w	θ	φ	p_n	μ	σ	N
N2	0.164	0.103	1.78	0.836	0.15	0.255	48
CB4856	0.185	0.118	1.69	0.815	0.18	0.265	30
DA609	0.220	0.104	1.73	0.780	0.17	0.029	24
VC125	0.251	0.099	1.67	0.749	0.15	0.031	24

5.3.3 Behavioral annotation algorithm

We developed a classification algorithm based on the speed probability density function (Eq. (5.1)) to characterize *C. elegans* trajectories into sharp turns and crawls. Our aim was twofold: first, distinguishing whether observed speed values originated from the Weibull ($\mathcal{W}$) or normal ($\mathcal{N}$) component of the mixed distribution, and second, obtaining the temporal sequence of sharp turns and crawls for each individual, referred to as the *ethogram*.

To address this problem we considered the speed as a continuous random variable that has a class function Ω, comprising two PDFs $\Omega = \{\mathcal{W}, \mathcal{N}\}$ with marginal probabilities $P(\Omega{=}\mathcal{W}) = p_w$ and $P(\Omega{=}\mathcal{N}) = p_n$. Therefore, the likelihood that a speed value originates from either PDF is determined by:

$$p(\Omega|v) = \frac{p(v|\Omega)\, p(\Omega)}{\sum_{\Omega} p(v|\Omega) p(\Omega)} \ , \tag{5.2}$$

which by substitution gives:

$$p(\mathcal{W}|v) = \frac{p(v|\mathcal{W}) p_n}{p(v)} \qquad \text{and,} \qquad p(\mathcal{N}|v) = \frac{p(v|\mathcal{N})\, p_w}{p(v)} \ .$$

Then, to initialize the classification algorithm one need to: (i) supply the speed values of a particular individual, (ii) define a bin width to compute empirical probability densities, (iii) define a window size to compute $p(w|v)$ and $p(\mathcal{N}|v)$, and finally (iv) input a summary table with the following columns and at least one annotated row as sharp turn or crawl:

Table 5.2: Summary table of a trajectory used to annotate sharp turns and crawls. Columns begin and end are the initial and final indexes of the windows; t_0 and t_1 are the corresponding initial and final times measured in seconds; and $p(w|v)$, $p(\mathcal{N}|v)$ are the probabilities that the speed values of the windows came from $\mathcal{W}$ or $\mathcal{N}$, respectively. Finally, the column Behavior indicates if windows are annotated as sharp turn or crawl.

begin	end	t_0	t_1	$p(w\|v)$	$p(\mathcal{N}\|v)$	Behavior
1	106	32.0	35.5	0.90	0.10	-
106	211	35.5	38.5	0.99	0.01	sharp turn
...	...	...	...	...	...	...
...	...	...	...	...	...	...
164859	164897	5548.0	5551.5	0.05	0.95	-
164897	164935	5551.5	5555.0	0.03	0.97	-

To streamline the classification process, we pre-annotated as sharp turns all rows of the input summary table showing $p(\mathcal{W}|v) > 0.999999$ (Figure 5.3A). As initial step, the algorithm sorted the table from high to low probabilities based on one of the two behaviors. In our case, we first annotated sharp turns and then crawls, but this order could be reversed without significant differences. Subsequently, the algorithm attempted to assign the first row of the table, the one with the highest $p(\mathcal{W}|v)$, as a member of $\mathcal{W}$. This decision relied on the mismatch between the observed speed distribution, deduced from the annotated states so far, and the modeled speed distribution (Eq. (5.1) fitted to data). The particular cost function we used for that purpose was the sum of squares between empirical and theoretical probability densities (Figure 5.3A-C). If the cost function decreased, the algorithm annotated the row as a sharp turn; otherwise, it was not annotated and marked as checked. This process, including the sorting step, iterated until all rows were annotated or checked. Next, the algorithm proceeded to annotate crawls. Like for the sharp turns, we pre-annotated as crawls all rows showing $p(\mathcal{N}|v) > 0.999999$ (Figure 5.3D) and then, the process continued as described above but this time the table was sorted based on $p(\mathcal{N}|v)$ instead of $p(\mathcal{W}|v)$. After this step, if there were still some rows to be annotated, the algorithm assigned them to $\mathcal{W}$ or $\mathcal{N}$ based on where they caused the smaller increase in the cost function. Finally, the table was time-sorted, and consecutive rows with the same label were merged into a single row, representing a complete behavior. This compacted version of the Table 5.2 constituted our ethogram.

In order to get the best approximation of the real ethograms we ran the algorithm for various windows sizes, and found 3.5 seconds yielded the best classification in terms of goodness-of-fit between observed and estimated speed distributions. We computed two metrics for assessing goodness of fit, the classification error itself, as the squared sum of errors (Figure 5.3G), and the outputted number of behavioral transitions (Figure 5.3H). This latter metric was used to seek for an equilibrium between highly dynamic and highly static annotations. We found that the different strains converged around 3.5 seconds, which represented a good compromise value. Therefore, we decided to rely on this window size to annotate all the trajectories. The resulting ethograms are summarized in Table 5.3 and presented in Figure 5.4. Significantly, we also assessed the accuracy of the algorithm using synthetic trajectories and identified a labeling error rate of approximately $6\% \pm 2\%$ (see Appendix D for a detailed description of this analysis).

Table 5.3: Statistical summary of the ethograms obtained with the behavioral annotation algorithm. All reported values except the number of tracks are strain based averages ± one standard deviation. The duration of the behaviors is expressed in seconds.

Strain	Tracks	Sharp turns duration	Crawls duration	N^{o} of transitions
N2	48	8.5 ± 11.8	41.0 ± 61.8	207 ± 65
CB4856	30	7.7 ± 20.3	36.1 ± 55.4	176 ± 97
DA609	24	8.3 ± 9.8	28.4 ± 38.0	282 ± 75
VC125	24	8.9 ± 11.2	24.0 ± 33.2	302 ± 91

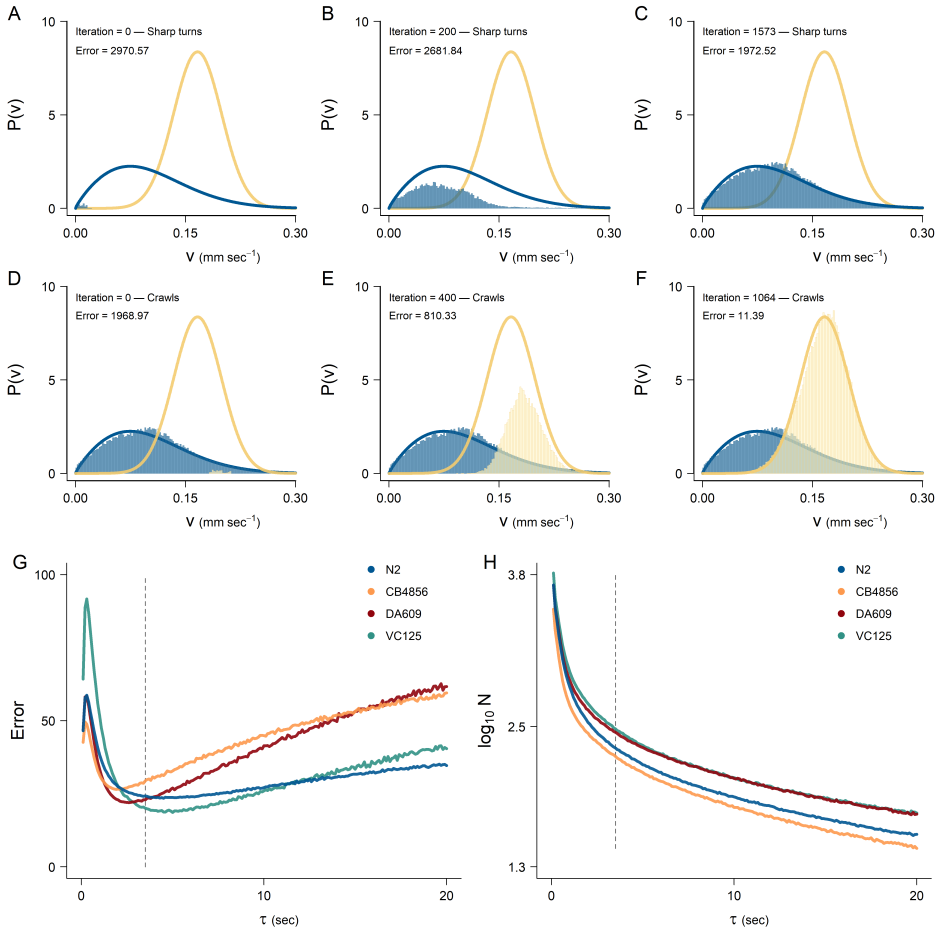


Figure 5.3: Behavioral annotation scheme and accuracy assessment. (A-F) Different stages of the annotation process in an N2 individual. The bin width to construct the histograms was $0.002\,\mathrm{mm\,sec^{-1}}$. (G, H) Strain-averaged profiles of the classification error, and number of behavioral transitions as a function of the classification time window τ.

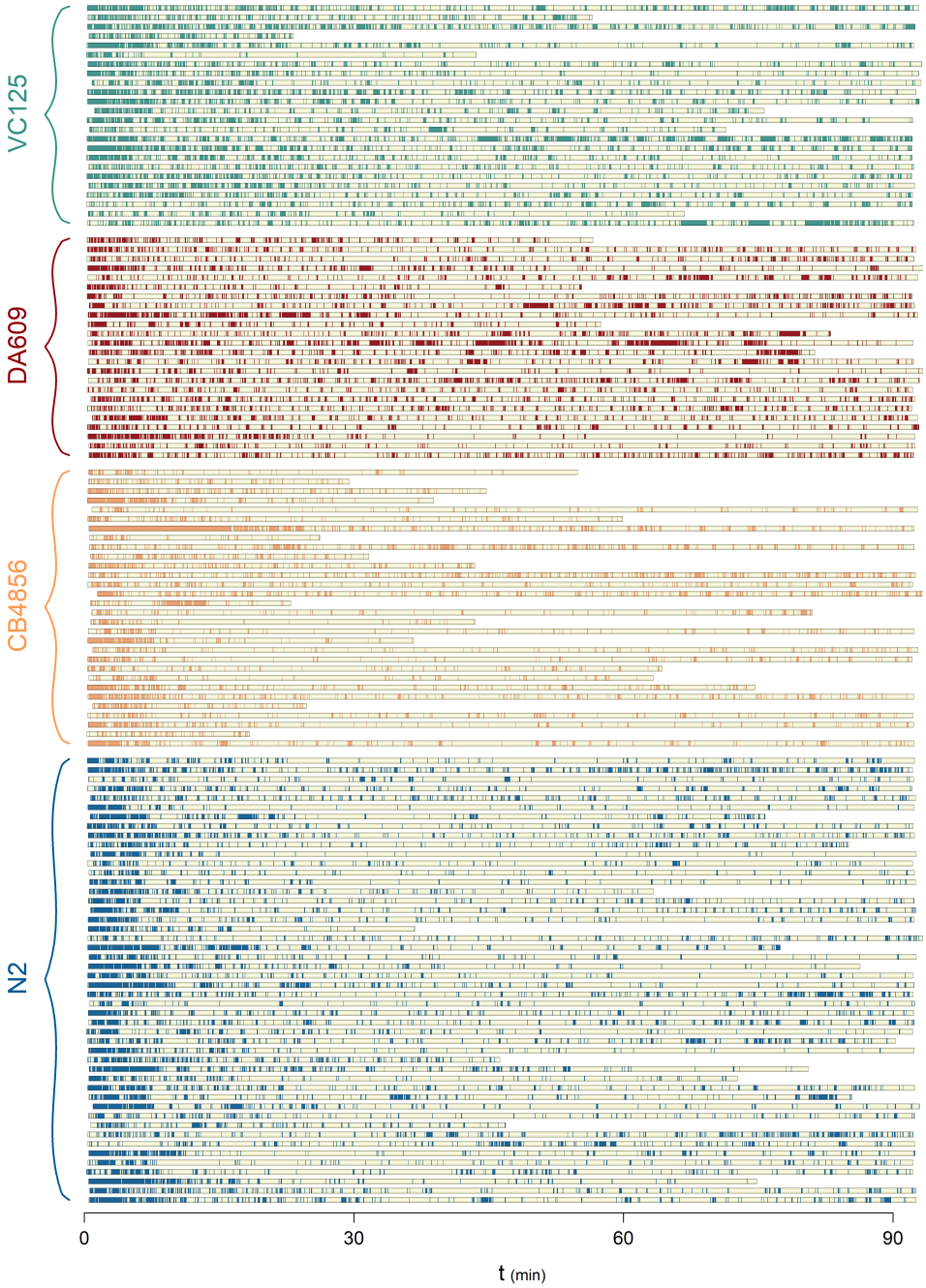


Figure 5.4: Individual ethograms obtained using a time window of $\tau = 3.5$ seconds, and a bin width of $0.002\,\mathrm{mm\,sec^{-1}}$ to estimate empirical probability densities. In all cases crawls are depicted in yellow and sharp turns in blue (N2), orange (CB4856), red (DA609) or green (VC125).

5.3.4 Angular velocity distribution

We characterized the distribution of angular velocity by testing the Cauchy and the normal probabilistic models (Figures 5.5A, B). The fitting process involved separate analyses for sharp turns and crawls, employing maximum likelihood estimation and model selection based on the Akaike Information Criteria (AIC). Out of a comprehensive set of 504 fits (126×4), the Cauchy distribution outperformed the normal distribution in all cases except for the crawling behavior of a singular VC125 individual. For the sake of simplicity and to maintain uniformity in our analysis, we opted to adopt the Cauchy distribution as the unique model of the angular velocity, whose PDF is defined as:

$$f(\omega) = \frac{1}{\pi}\left(\frac{\eta}{(\omega - \bar{\omega})^2 + \eta^2}\right) , \tag{5.3}$$

where $\bar{\omega}$ is the location parameter and η is the scale parameter of the distribution.

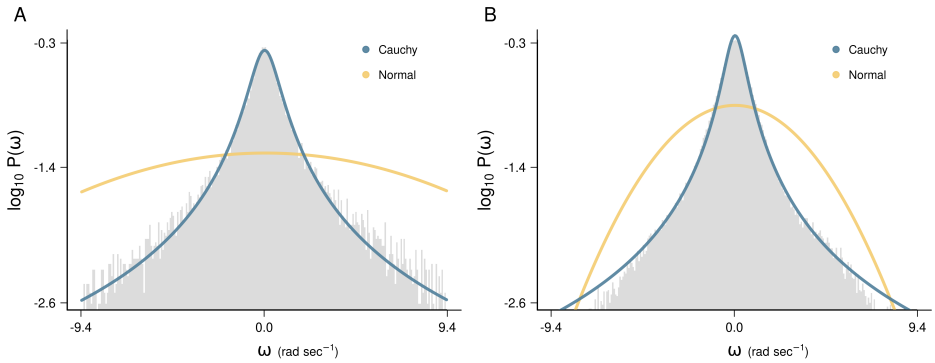


Figure 5.5: Fits of the angular velocity distribution of an N2 individual. (A) Sharp turns. (B) Crawls.

Table 5.4: Strain-averaged parameters of the two Cauchy distributions of angular velocity, measured in $\mathrm{rad\,sec^{-1}}$. Parameters associated with sharp turns are denoted with the subscript s, while those associated with crawls are denoted with the subscript c. Columns $\bar{\omega}_s$ and $\bar{\omega}_c$ represent the location (mean) of the distributions, while η_s and η_c denote their corresponding scales. Since approximately half of the individuals exhibited a negative location (anticlockwise turns) and the other half a positive location (clockwise turns), these values are reported in absolute terms. Finally, the column N reports the number of individuals used to compute these average values.

	Cauchy (sharp turns)		Cauchy (crawls)		
Strain	$\lvert\bar{\omega}_s\rvert$	η_s	$\lvert\bar{\omega}_c\rvert$	η_c	N
N2	0.028	0.507	0.048	0.375	48
CB4856	0.027	0.548	0.034	0.392	30
DA609	0.022	0.533	0.038	0.393	24
VC125	0.033	0.629	0.041	0.437	24

Box 1: Alternative version of the annotation algorithm

An additional version of the algorithm worth commenting involves incorporating both speed and angular velocity to classify sharp turns and crawls. This implementation required: (i) adding angular velocity data into Table 5.2, including observed values and probabilities, and (ii) evaluating the cost function based on both the Weibull-normal distribution of speed (Figure 5.2), and the Cauchy-Cauchy distribution of angular velocity (Figure 5.5).

Although incorporating angular velocity might seem convenient to better detect sharp turns and crawls, we found the error in the resulting classification was significantly higher in all individuals compared to the version considering only speed. This limitation stemmed from our data, and not from the algorithm itself, as the Cauchy distributions of angular velocity associated with sharp turns and crawls were completely overlapped. Indeed, our algorithm is suitable for working with any movement variable other than speed if their generative distributions are distinct enough, potentially accommodating n-dimensional distributions and classifying more than two behaviors.

Interestingly, the challenge of accurate classification due to overlapping distributions is not exclusive to machine learning algorithms but is also observed in biological learning. Stephens et al. (2007) provides a compelling example of a similar problem in the context of foraging behavior, demonstrating that the ability of birds to distinguish between tasty and noxious beetles is reduced when the color distribution of the two prey species overlaps.

5.3.5 Duration of sharp turns and crawls

To characterize the duration of sharp turns and crawls we fitted the probabilistic models of the Weibull, gamma, exponential and double exponential distributions (Figure 5.6). After some previous analyses, power laws were excluded from the analysis as the range of observed durations were too small (less than 1.5 order of magnitude in sharp turns), and long-lasting behaviors were not frequent enough for the maximum likelihood estimation to yield reliable goodness fit. Even when considering truncated power law distributions. In all cases, fits were done using maximum likelihood estimation and model selection was based on AIC (analytical expressions of all models are found in Appendix B.1).

For sharp turns, the gamma distribution was best in 78.6% of the individuals ($n = 99$) and the double exponential was best in 19% ($n = 25$). The remaining two individuals could be classified with any of the two models as $\Delta\text{AIC} < 2$ (Figure 5.6B). Based on these results, and to maintain uniformity and comparability in our analysis we modelled the duration of sharp turns according to the gamma distribution, defined as:

$$f(\tau) = \frac{1}{\Gamma(\kappa)\,\vartheta^{\kappa}}\;\tau^{\,\kappa-1}\,e^{-\tau/\vartheta}\;, \tag{5.4}$$

where τ is the duration; κ is the shape parameter; ϑ is the scale, in seconds; and Γ is the gamma function.

For crawls, the double exponential distribution was best in 65.9% of the individuals ($n = 83$) and the gamma distribution was best in 3.2% ($n = 4$). The remaining 39 individuals could not be assigned to any particular model as $\Delta\text{AIC} < 2$ (Figure 5.6D). Aiming to maintain uniformity and comparability in our analysis we modelled the duration of crawls according to a double exponential distribution, whose PDF is defined as:

$$f(\tau) = p\;\frac{1}{t_1}\,e^{-\tau/t_1} \;+\; (1-p)\;\frac{1}{t_2}\,e^{-\tau/t_2}\;, \tag{5.5}$$

where τ is the duration, p is a probability weight, and t_1, t_2, are the scales of the two exponential functions.

Strain-averaged parameters of the gamma and the double exponential distribution fitted to sharp turns and crawls, respectively, are reported in Table 5.5.

Table 5.5: Strain-averaged parameters of the gamma and double exponential distributions fitted to sharp turns and crawls durations. Columns κ and ϑ represent the shape and scale parameters of the gamma distribution. Columns p, t_1, and t_2 indicate the probability weight and two distinct scales of the double exponential. N denotes the number of individuals used to compute the averages.

	Gamma (sharp turns)		Double exponential (crawls)			
Strain	κ	ϑ (sec)	p	t_1 (sec)	t_2 (sec)	N
N2	1.90	5.46	0.54	16.16	84.40	48
CB4856	2.46	4.61	0.51	13.66	72.00	30
DA609	2.18	4.45	0.49	14.60	56.33	24
VC125	1.92	5.06	0.53	11.75	44.46	24

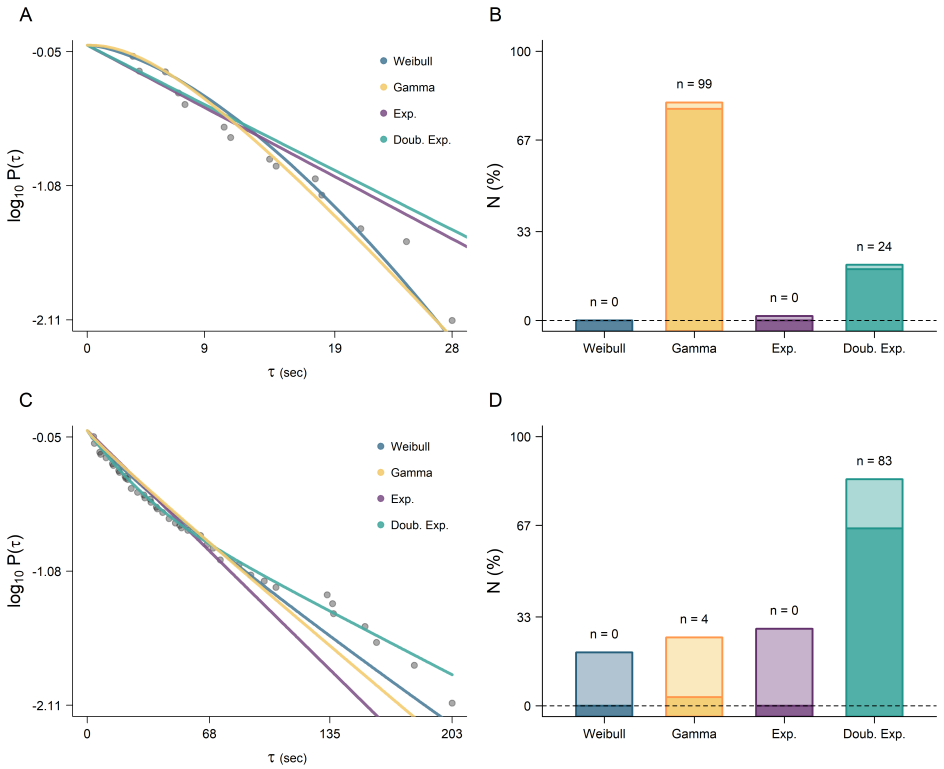


Figure 5.6: Probability distributions and model selection of the duration of sharp turns and crawls. (A) Exceedance probability of sharp turn durations of an N2 individual. (B) Model selection for the whole population (N = 126). (C, D) Show the same information as A, B but for crawls. In panels B, D, counts above bars indicate the number of times were model selection was significant (ΔAIC > 2) and correspond to dark colors, while lightened colors indicate ΔAIC < 2 per model, which are not significantly different.

5.3.6 Temporal unfolding of sharp turns and crawls

The ethograms depicted in Figure 5.4 immediately drew attention to the existence of non-stationary trends. Specifically, they showed the frequency of sharp turns decreased within the initial 15-30 minutes of the experiment, stabilizing thereafter at a roughly constant rate. While the duration of crawls increased over time, transitioning from very short to very long crawls. These trends were more pronounced in the wild type strains (N2, CB4856), than in the mutant strains (DA609, VC125), but appeared as a general pattern at both the population and the individual levels.

The frequency of sharp turns in a given time window depends on both sharp turn and crawl durations. If these durations were strongly correlated, a single time-dependent rate, either of sharp turns or crawls, could explain the patterns observed in our ethograms. However, Figure 5.7 shows no linear correlation between consecutive sharp turn and crawl durations. Long sharp turns never co-occur with long crawls, and short sharp turns are associated with both short and long crawls (Figure 5.7A). Similarly, our results reveal a time-dependent increase in sharp turn rate and a decrease in crawl rate (Figure 5.8A), highlighting the need to consider both dynamics to understand why the frequency of sharp turns decreased over time.

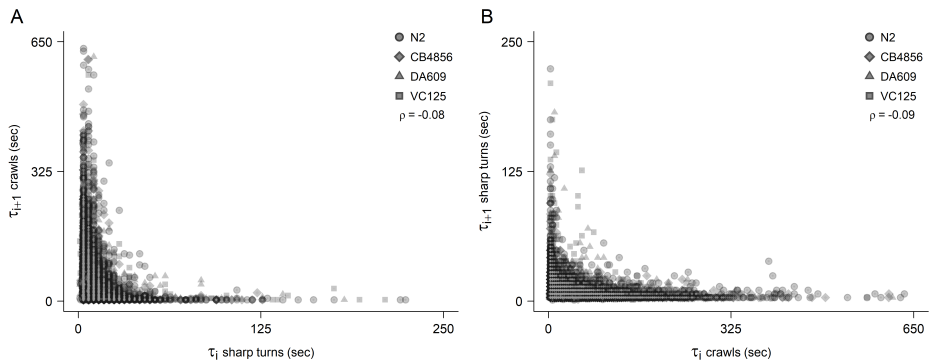


Figure 5.7: Correlation between the durations of consecutive sharp turns and crawls. (A) Sharp turns against subsequent crawls. Pearson correlation $\rho = -0.08$, N = 14666. (B) Crawls against subsequent sharp turns. Pearson correlation $\rho = -0.09$, N = 14618.

Based on these facts and work done by Klein et al. (2017), we modelled the rate of sharp turns as an exponential function of the form:

$$\lambda_s(t) = a_s \left(1 - e^{k_s t}\right) + b_s \ , \tag{5.6}$$

where a_s is the limiting rate, k_s is the memory rate of sharp turns, and b_s is the baseline rate. Although this function is properly defined only when k_s is negative

(positive k_s yield undefined negative rates as time increases), we included positive k_s to account for the fact that some individuals, especially in mutant strains DA609 and VC125, exhibited and increasing duration of sharp turns (Figure 5.4). The underlying idea was to capture the memory dynamics associated with the preconditioning experiences of each individual under study. Thus, negative k_s indicated this memory degraded over time while positive k_s indicated it was maintained.

In the same spirit, we modelled the rate of crawls as an exponential function of the form:

$$\lambda_c(t) = a_c e^{k_c t} + b_c \ , \tag{5.7}$$

where a_c is the excess initial rate; k_c is the memory rate of crawls; and b_c is the baseline rate. Like for sharp turns, we included both positives and negative k_c given that some individuals exhibited a decreasing duration of crawls (Figure 5.4). However, here we did not have to control for the value of k_c provided that Eq. (5.7) never yield negative rates.

In order to fit these trends to data, we drew 20K random points occurring throughout the temporal axis of the individual ethograms, and centered a time window of size τ to each point. We then computed the durations of completed sharp turns and crawls occurring inside the windows and assigned them to the centers (class mark) of non-overlapping time windows of the same size τ, starting at $t = 0$. Finally, we averaged the durations of sharp turns and crawls by each class mark, and computed the rates just as the inverse average durations. To make sure we obtained unbiased results, we fitted the trends for 60 different non-overlapping window sizes, from $\tau = 60$ to $\tau = 650$ sec; and chose the size that better reproduced the observed proportion of sharp turn to crawls: $\tau = 220$ sec (Figure 5.8B).

Table 5.6: Strain-averaged parameters of the exponential functions describing the rate of sharp turns and crawls, expressed in sec^{-1}. Column a_s is the limiting rate of sharp turns, a_c is the excess initial rate of crawls, k_s and k_c are the memory rates, and b_s and b_c are the baseline rates. N reports the number of individuals used to compute the averages.

	Sharp turns			Crawls			
Strain	a_s	k_s	b_s	a_c	k_c	b_c	N
N2	0.175	-0.0017	0.035	0.139	-0.0021	0.017	48
CB4856	0.182	-0.0024	0.056	0.103	-0.0021	0.015	30
DA609	0.160	-0.0005	0.095	0.110	-0.0029	0.026	24
VC125	0.153	-0.0013	0.061	0.159	-0.0017	0.027	24

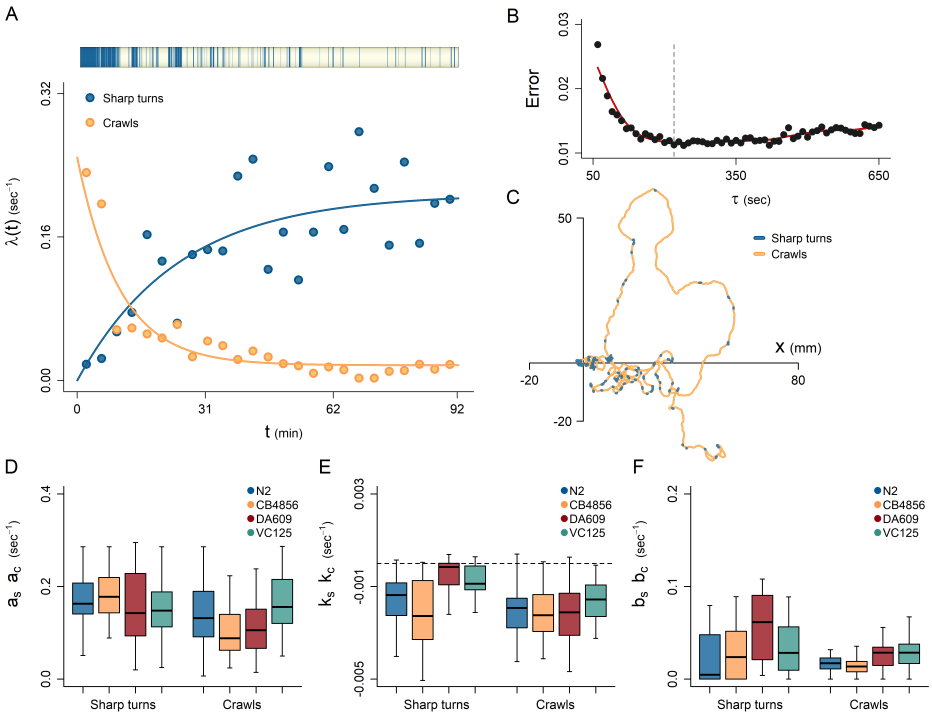


Figure 5.8: Temporal unfolding of sharp turns and crawls. (A) Sharp turns (blue) and crawls (orange) rates fitted to the ethogram of an N2 individual. (B) Strain-pooled error profile of the proportion of sharp turns to crawls as a function of window size τ. (C) Trajectory of the analyzed individual. (D-F) Boxplots comparing rate parameters of sharp turns and crawls (see Eq. 5.6 and Eq. 5.7). (D) Limiting rate a_s and excess initial rate a_c. (E) Memory rates k_s and k_c. (F) Baseline rates b_s and b_c.

5.3.7 Summary results

As anticipated in data exploratory analysis and in line with the design of the classification algorithm, our analysis unveiled noteworthy distinctions in the speed of sharp turns and crawls. The differentiation between them is clearly depicted in the clustering of the parameter space in Figure 5.9D, where blue dots map the Weibull distribution of sharp turns and orange dots map the normal distribution of crawls. The primary contrast lies in the average speed (first moments: θ and μ), which is significantly lower for sharp turns, as confirmed by pair-wise t-test results ($\alpha < 0.05$). Notably, the speed of sharp turns exhibits considerable fluctuations spanning from almost zero to exceptionally high values, as depicted in Figure 5.9B, and further emphasized by the significantly higher variance compared to crawls (pair-wise t-test, $\alpha < 0.05$).

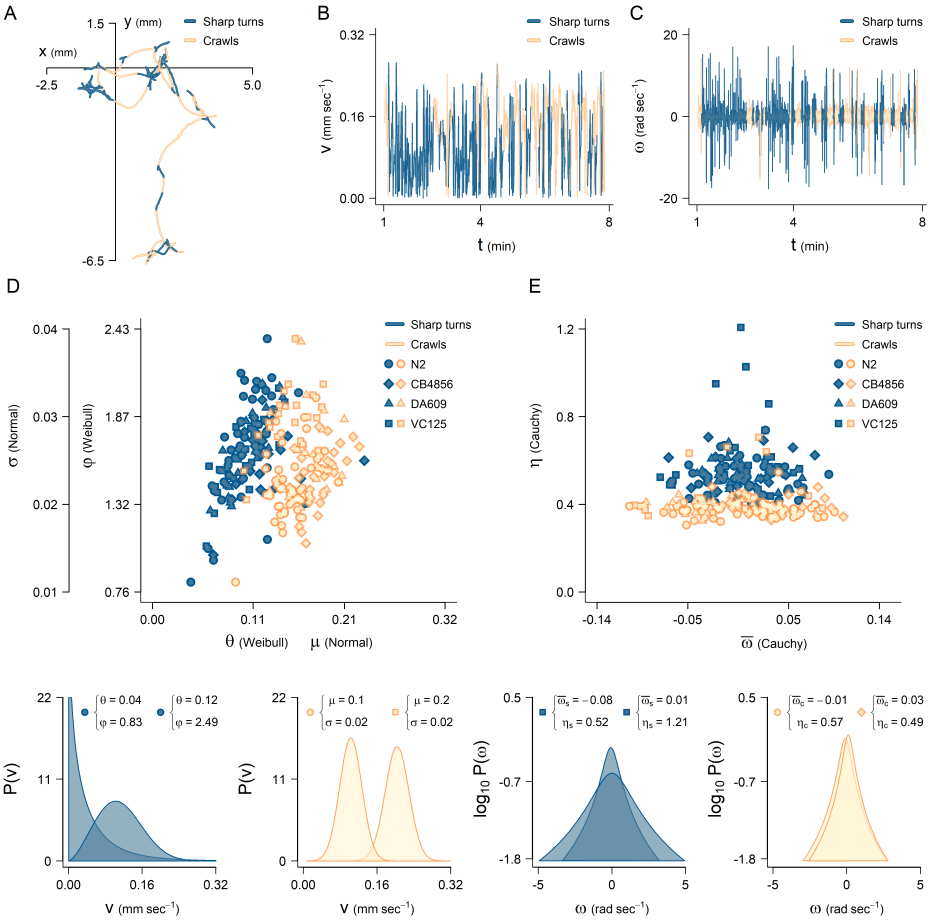


Figure 5.9: Microscopic movement features of sharp turns and crawls. (A-C) Trajectory chunk with annotated behaviors and their relative temporal series of speed and angular velocity. (D) Parameter space of the Weibull-normal speed distribution fitted to each individual. (E) Parameter space of the Cauchy-Cauchy angular velocity distribution fitted to each individual. Bottom panels: Individual pair-wise comparisons for each movement variable (speed and angular velocity) and for each motor behavior (sharp turns and crawls). Their corresponding parameter values are reported in order to locate them on the parameter space.

The primary distinction in angular velocities between sharp turns and crawls (Figures 5.9C, E) lies in their variability (scale parameter η). Sharp turns exhibit significantly higher variance than crawls, as determined by pair-wise t-test results ($\alpha < 0.05$). However, there is no significant difference in their average values (pair-wise t-test on the first moment, $\alpha < 0.05$). Notably, our results unveil a consistent curvature bias in the direction of motion ($\bar{\omega}, \neq, 0$) across all individuals under study.

This bias manifests as a tendency for individuals to turn either clockwise (46.8%) or anticlockwise (53.17%) during the execution of both behaviors. It is worth noting that this bias, measured at 33Hz, would result in a complete loop in approximately 2-3 minutes.

In addition, we analyzed movement features considering the whole scale of sharp turns and crawls, both pooling all individuals and pooling by strain (Figure 5.10). Our results show sharp turns last for a short time: $\langle\tau\rangle \approx 9$ sec, involve a small net displacement: $\langle\Delta x\rangle \approx 0.5$ mm, and cause a large change in the direction of motion: $\langle|\Delta\phi|\rangle \approx \frac{1}{2}\pi$ rad. On the contrary, crawls last for a longer time: $\langle\tau\rangle \approx 30$ sec, involve a larger net displacement: $\langle\Delta x\rangle \approx 3.2$ mm, and cause a smaller change in the direction of motion: $\langle|\Delta\phi|\rangle \approx \frac{1}{6}\pi$ rad (see Table 5.3, to consult strain-average durations).

5.4 Discussion

5.4.1 On behavioral annotation of *C. elegans*

The intricate movement patterns of *C. elegans* are believed to arise from the interplay of two fundamental motor behaviors: sharp turns (reorientations) and crawls (Pierce-Shimomura et al., 1999; Salvador et al., 2014). Despite this apparent simplicity, mapping these behaviors onto individual trajectories poses a challenging task and prompted the development of various methods. Among these, posture analysis (Stephens et al., 2008, 2010, 2011; Brown et al., 2013) and Hidden Markov Models (Gallagher et al., 2013; Lee and Kang, 2015; Roberts et al., 2016) have emerged as the most widely utilized techniques in the field.

Posture analysis stands out as a highly accurate method for the quantitative assessment of behavior in *C. elegans* and model organisms in general (Gomez-Marin et al., 2012). Notably, it is entirely data-driven, alleviating the need for strict mathematical assumptions. Despite its strengths, there are two limitations that render it unsuitable for addressing our specific goals. First, posture analysis necessitates high-throughput imaging of the animals under study, which may not always be feasible, particularly in long-term laboratory experiments or in field studies. While we recognize its significance in advancing behavioral ecology, specially in *C. elegans*, it has yet to become a universally applicable tool in movement ecology. Second, in its simplified form, posture analysis primarily identifies omega turn-like behaviors, making the detection of other sharp turns, such as reversals and pirouettes, difficult (Stephens et al., 2008). Whereas in its more sophisticated form, one can find that *C. elegans* adopts up to 90 different postures in just 15 minutes (Gomez-Marin et al., 2016). As Gomez-Marin et al. (2016) showed in their analysis, such

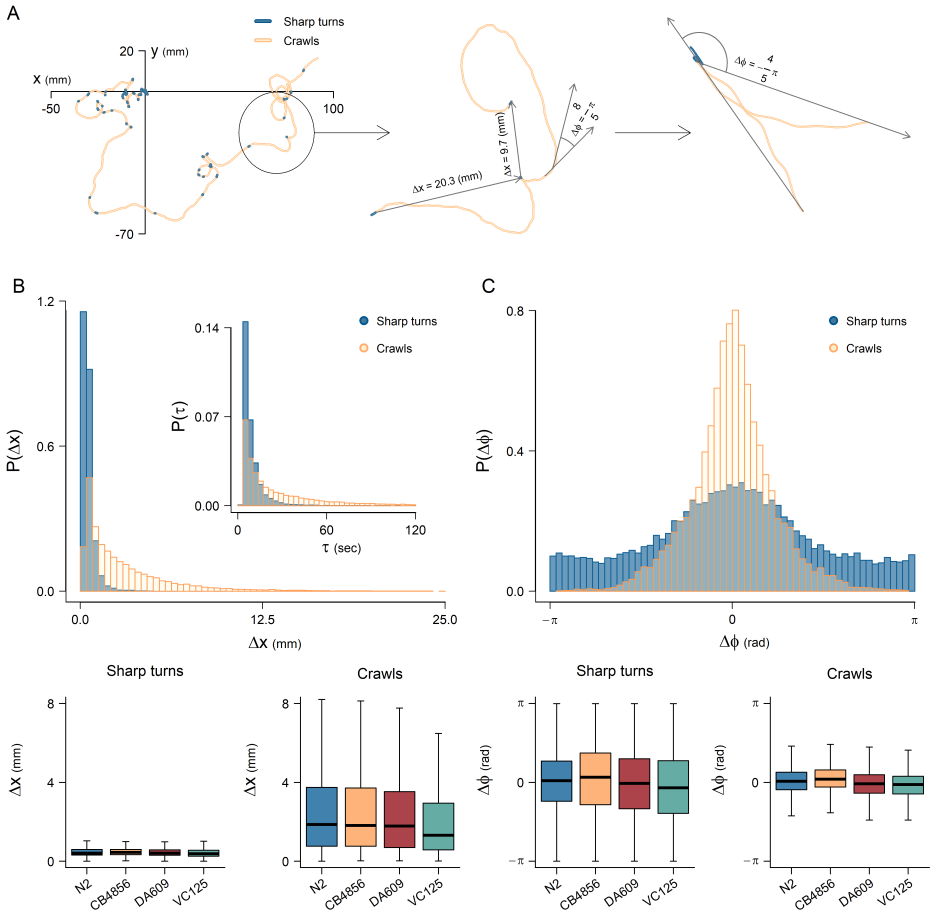


Figure 5.10: Net displacement and turning angle of sharp turns and crawls. (A) Representative trajectory with annotated behaviors. Central zoom-in illustrates the computation of net displacements Δx and turning angles $\Delta\phi$ associated with crawls. Right zoom-in does the same for sharp turns, without showing Δx. For both sharp turns and crawls, turning angles were computed using two vectors representing the starting and end positions 10 second before and after the turn position. (B) Strain-pooled distribution of net displacement and duration (inset) associated with the behaviors. (C) Equivalent distribution for the turning angle. Bottom panels: illustrate that we can pool data from different strains as significant differences occur between motor behaviors and not between strains (pairwise t-test, $\alpha < 0.05$).

an extensive postural repertoire can be simplified to different levels of detail, and likely some of them are close to the description of sharp turns and crawls provided in this thesis. Therefore, further work is needed to better integrate and compare centroid-based and posture-based annotation algorithms.

On the contrary, Hidden Markov Models (HMM) rely on centroid positions of moving animals (rather than images), making them highly applicable across various species in both field and laboratory settings (Patterson et al., 2008; Gallagher et al., 2013; Conners et al., 2021). However, it is essential to note that HMM hinge on certain mathematical assumptions that movement data often fails to meet, leading to potentially inaccurate classifications. In our specific case, HMM struggled to capture the Weibull distribution of speed associated with sharp turns (utilizing the R package *msm* described in Jackson (2011)). This shortcoming pointed out HMM are not the ideal framework for annotating *C. elegans* trajectories mainly because: (i) the non-Gaussian nature of the distributions involved; and (ii) the non-stationary characteristics of the transition rates between sharp turns and crawls (Figures 5.1G, H). While there are methods to address these challenges, they typically involve intricate mathematical and programming techniques beyond the scope of this thesis.

The preceding discussion underscores the relevance of our approach for annotating *C. elegans* trajectories into sharp turns and crawls. Importantly, our proposed method is centroid-based, data-driven, and holds potential for generalization to other moving animals and movement variables. The only foundational assumption is that the variable of interest (in our case, speed) originates from distinct probability distributions. These distributions are assumed to persist throughout the entire observational period, and characterized by significant differences in both scale and shape, making them discernible and distinguishable.

5.4.2 The role of sharp turns and crawls on *C. elegans* spreading

We have modelled key movement features of sharp turns and crawls in terms of duration and temporal unfolding. Our results show the duration of sharp turns follows a gamma distribution while the duration of crawls follows a double exponential (Table 5.5, Figure 5.6). In spite of apparent contradiction with other findings describing the duration of sharp turns as a simple exponential (Helms et al., 2019), the statistical properties using the gamma or the exponential distribution were practically identical. Regarding crawls, our findings coincide with the majority of studies, which describe their duration as a double exponential despite using different techniques and experimental settings (Pierce-Shimomura et al., 1999; Shingai, 2000; Wakabayashi et al., 2004; Srivastava et al., 2009; Stephens et al., 2010; Salvador et al., 2014).

Noteworthy, our analysis revealed the duration of sharp turns and crawls exhibit non-stationary trends over time. As established in literature, *C. elegans* adapts the frequency of sharp turns in response to internal states (Salvador et al., 2014; Bar-

tumeus et al., 2016) or environmental cues (Pierce-Shimomura et al., 1999), resulting in a diverse array of trajectory patterns, including multiscale loops, starred-like local searches, and nearly ballistic relocations (Peliti et al., 2013; Salvador et al., 2014; Bartumeus et al., 2016). Notably, our results suggest *C. elegans* does so by modulating both the duration of sharp turns and the duration crawls through time. Regardless the intra- and inter-strain variability, we found all individuals under study unfolded a highly ordered behavioral template explaining non-stationary spreading dynamics (Moses effect) found in Chapter 4.

Having resolved that, a pivotal question we still to answer is about the origin of long-range correlations of spreading velocity (Joseph effect), and by extension, of the overall superdiffusion also found in Chapter 4. Like non-stationary dynamics, long-range correlations may arise from various sources affecting directional persistence, such as speed correlations (Dieterich et al., 2008), angular velocity correlations (Méndez et al., 2014), or multiscale reorientation patterns (Cole, 1995; Bartumeus and Levin, 2008; Campos et al., 2014; Salvador et al., 2014), to put some examples. Resolving this puzzling question is crucial for a comprehensive understanding of *C. elegans* spreading dynamics, and in the following chapter, we will address it by developing a mechanistic model of *C. elegans* movement.

Chapter 6

Memory decay drives superdiffusion in *C. elegans* search behavior

Abstract

The underlying mechanisms driving superdiffusion in *Caenorhabditis elegans* remain poorly understood. Yet superdiffusion is believed to play a pivotal role in key ecological processes such as dispersal, search behavior, and foraging, all of which impact the fitness of both individuals and species. To address this gap, we developed a mechanistic model of *C. elegans* movement based on the Langevin formalism. This model allowed us to identify the generative mechanisms producing *C. elegans* spreading patterns observed in classical food patch relocation experiments. Our findings unveiled a memory decay reflected in the frequency decay of sharp turns over time and associated to the food patch precondition. This memory decay is modulated by non-stationary crawl and sharp turn rates and manifests as an accelerated spreading pattern, or superdiffusion, for more than one hour. Notably, this non-stationary modulation overcomes the strong influence of curvature bias and looping behavior, phenomena that would otherwise lead to diffusive or subdiffusive spreading. Moreover, this investigation revealed that various mutant and wild-type *C. elegans* strains exhibit an optimal, or marginally optimal, memory decay that maximizes superdiffusion. These findings underscore the importance of assessing non-stationary dynamics when modeling animal movement and search behavior. Additionally, they highlight the relevance of conducting experiments in almost unbounded conditions, and covering a wide range of scales to accurately compute diffusive properties and assess potential anomalies.

6.1 Introduction

In previous chapters, we have learned that *C. elegans* searches for food by modulating the duration of sharp turns and crawls, and that this modulation, together with a systematic curvature bias should explain the complex spreading patterns observed in our experiments. Yet, we have not comprehensively identified the physical mechanisms explaining the emergence of anomalous diffusion, including both non-stationary dynamics (Moses effect) on increment displacements and long range correlations (Joseph effect). In order to address this question we developed a mechanistic model of *C. elegans* movement that enabled us to link microscopic movement parameters to macroscopic diffusive properties. Such model incorporates an explicit representation of sharp turns and crawls that reproduces its most relevant features (i.e., durations, net displacements, turning angles and temporal unfolding), while at the macroscopic scale, it reproduces the overall MSD and its diffusive regimes.

Our model, was built upon the classical Langevin equation (Langevin, 1908) with a minimal set of adaptations required to reproduce *C. elegans* search trajectories. The Langevin formalism is ideal to mechanistically link microscopic dynamics of speed and turn with macroscopic diffusive properties (Méndez et al., 2014). In its original form, it is assumed that both instantaneous speed and turn are fueled by Gaussian white noise (Langevin, 1908; Méndez et al., 2014; Metzler et al., 2014). However, animal movement is usually inconsistent with these assumptions. For example, angular velocities often follow a Cauchy distribution instead of a normal distribution (Wu et al., 2000; Bartumeus et al., 2005). Similarly, their speed correlations tend to decay slower than it happens in white noise. In addition, animal and cell tracks often involve skewed distributions (Upadhyaya et al., 2001; Moore et al., 2017; Podestá et al., 2017), as is the case for the speed of sharp turns in *C. elegans* (Chapter 5, Sections 5.3.2, 5.3.3).

Therefore, to accurately model movement of living organisms using a Langevin approach, it is necessary to incorporate specific modifications. For instance, typical Langevin models introduce elements such as a correlation timescale or fractional noise to account for velocity correlations (Kubo, 1966; Lenz et al., 2013; Metzler et al., 2014; Slezak et al., 2018). Importantly, these adjustments maintain the symmetry of the underlying distributions, as accommodating skewed (non-symmetric) distributions proves exceedingly challenging (Cáceres and Budini, 1997; Sabino and Petroni, 2021). While some approaches have been developed in that regard, with varying degrees of success, they often entail intricate mathematical methods that are difficult to implement, comprehend, or justify from a biological perspective.

Last but not least, the modeling of animal movement needs to address the interplay between various behaviors (Patterson et al., 2008; Bartumeus, 2009). Movement is the result of different behaviors that are modulated by different biological and physical processes. Therefore, to get a comprehensive understanding of how animals move it is convenient to incorporate explicit representations of these behavioral modes and their potential transitions. This specific problem has been addressed in various ways in Hidden-Markov modelling frameworks and similar (Fryxell et al., 2008; Bartumeus et al., 2016; Conners et al., 2021). All of them involving a hierarchical or nested approach with multiple layers. At the most inner layer one typically find the mathematical representation of each behavior, in the form of different parameters, statistical distributions or even different models; and, on top of it, one usually find the mathematical representation of their occurrence and durations, in the form of instantaneous rates, transition matrices, or independent random variables (Bartumeus et al., 2005; Salvador et al., 2014; Metzner et al., 2015; Gurarie et al., 2017; Conners et al., 2021).

In our approach, we have sought a balance between model complexity and the ability to offer meaningful physical or biological interpretations. The model we present for *C. elegans* incorporates two distinct Langevin equations, one for crawls and another for sharp turns. By segregating the modeling of sharp turns and crawls and introducing time-dependent rates and specific duration distributions for each behavior, we effectively addressed the observed transitions between them, as well as their non-stationary patterns. Drawing inspiration from Metzler et al. (2014), we characterized the speed of sharp turns as a random variable that can follow an arbitrary distribution (in this case Weibull distribution), while also ensuring a negative correlation between speed and angular velocity, consistent with empirical data. Additionally, we incorporated a constant curvature in the motion to account for the emergence looping patterns, one of the most salient features of *C. elegans* trajectories. With these refinements, our model successfully reproduces many trajectory outputs, enabling us to establish a mechanistic link between microscopic and macroscopic properties of movement and gain deeper insights into how *C. elegans* exhibits superdiffusion.

6.2 Materials and Methods

Experimental data

All data used in this chapter were extracted from relocation experiments described in Chapter 3. These experiments consisted on placing well-fed *C. elegans* individuals in a bare arena of 24.5×24.5 cm, and record its movement during the following 90 minutes, or until the animal reached the edge of the plate. In total 126 trajec-

tories of 4 different strains were recorded: N2 ($n = 48$), CB4856 ($n = 30$), DA609 ($n = 24$); and VC125 ($n = 24$).

Trajectories were reconstructed from the center-of-mass of the worms and then smoothed following Stephens et al. (2010). Speed was calculated as the time derivative of the smoothed positions and the angular velocity was calculated as the time derivative of the turning angle, computed at the trajectory frame-rate (33Hz). Extreme speed values and consecutive observations with the exact same coordinates, speed or turning angle were assumed to be errors of the tracking system, as the tracker always generated some noise when repositioning the worm. The latter type of errors included zero exact values of speed ($v = 0$), which should not be considered locomotion pauses as defined in (Costa et al., 2019). Indeed, locomotion pauses occurred in some individual tracks, but they were not specifically characterized in our analysis. In total, less than 0.17% of data were filtered out.

Mean square displacement (MSD)

Following methods outlined in Chapter 4, we computed the mean square displacement as $\mathrm{MSD}(t) = \frac{1}{\mathrm{N}} \sum_{i=1}^{\mathrm{N}} |\mathbf{r}_i(t) - \mathbf{r}_i(0)|^2$. For *C. elegans* trajectories, we used strain-based ensembles while for the models we employed ensembles of 1000 trajectories, simulated at the same frame-rate as the empirical tracks (33Hz), and lasting the same duration (90 minutes on average). Since empirical tracks did not start exactly at the origin of coordinates ($t = 0$, $x = 0$, $y = 0$), we adjusted temporal and spatial offsets of model trajectories. The temporal offset was rectified by initializing the models at the same time as the empirical tracks, which started on average at $t = 18$ seconds. The spatial offset was rectified by initializing the models at $x = 0$, $y = 0$, computing their MSD curves, and then adding the MSD of the very first empirical coordinates, which on average was 0.44 mm^2. Importantly, this correction had negligible impact on the long-term MSD while it enabled us to reproduce the early stages of the empirical curves (Figure 6.1A).

In all cases, both empirical and synthetic, the parameters of the MSD were estimated by fitting a log-log piece-wise linear model with 4 regimes, similar to Riahi et al. (2019). The fitting procedure was based on least-squares and was implemented using the R package *segmented* (Muggeo, 2003). Although a greater number of regimes could be easily found, 4 was a good compromise between simplicity and accuracy to obtain reliable estimates of the long-term anomalous diffusion exponent α (Figure 6.1B) and reproduce the complexity of the MSD patterns in *C. elegans*. Smaller number of regimes would yield overestimates of α and oversimplified MSD patterns. The effective diffusion coefficient D_α of each regime was estimated from the intercepts of the piece-wise model, in exponential form and divided by 4.

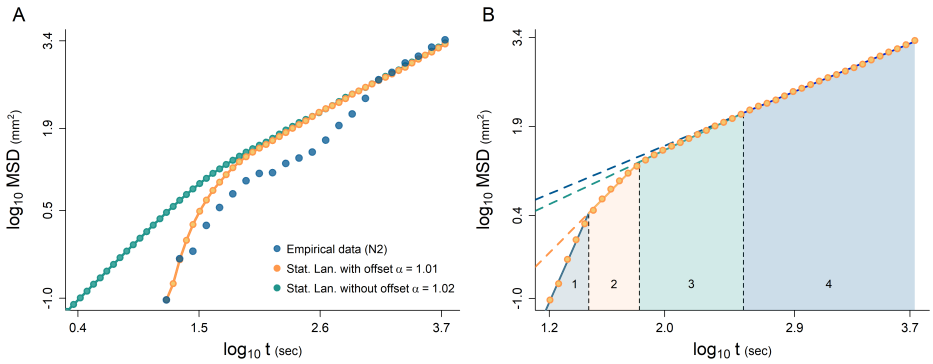


Figure 6.1: Illustrative example of the mean square displacement (MSD) computation. (A) Comparison of the MSD curves with and without offset for the stationary model developed in Section 6.3. In the legend, reported α values correspond to the anomalous diffusive exponent of the last (fourth) fitted regime. (B) Piece-wise model with four regimes fitted over the MSD curve of the stationary model (with offset) shown in panel A. This very same technique was used throughout the whole chapter to obtain the estimates of α, and the effective diffusion coefficient D_α.

Decomposition of the anomalous diffusion exponent was carried out through the computation of the Moses (M), Noah (L), and Joseph (J) exponents. This analysis allows to quantitatively assess the contribution of: (i) non-stationary dynamics ($M \neq 0.5$), (ii) extreme events ($L > 0.5$), and (iii) long-range correlations ($J > 0.5$), on the anomalous diffusion exponent α, given the following summation:

$$\alpha = 2\,(M + L + J - 1)\ , \tag{6.1}$$

usually expressed in terms of the Hurst exponent $H = \alpha/2$ (Chen et al., 2017; Aghion et al., 2021; Vilk et al., 2022; Meyer et al., 2022). The Moses M and Noah L exponents were estimated from the cumulative time series of the increment displacements, as in Chen et al. (2017), while the Joseph exponent J was derived from the ensemble time-averaged mean square displacement (TAMSD), following Meyer et al. (2022). In practice, we fitted a log-log piece-wise linear model with two (M, L) or three (J) regimes, and got the estimates from the slope of the very last regime. Detailed methods on how to estimate these exponents can be found in Chapter 4.

Statistical analyses

Comparison analyses were made using t-tests, or pairwise t-tests with a significance level $\alpha = 0.05$ (R package *rstatix*). For multiple group comparisons p-values were

adjusted using the Holm-Bonferroni correction (Holm, 1979). When needed to meet tests' assumptions we followed the Inter Quartile Range criteria (Hoaglin, 2003) and treated as outliers all data points falling below $Q_1(x) - (1.5 \cdot \text{IQR}(x))$ or above $Q_3(x) + (1.5 \cdot \text{IQR}(x))$. All analyses that follow were performed at the individual level, unless otherwise stated, and were programmed using custom scripts in R (R Core Team, 2023) and C++.

6.3 The model of movement

We modelled *C. elegans* locomotion using a hierarchical approach with two nested layers. The innermost layer of the model contains two Langevin equations describing the dynamics of speed and turn for (i) crawls and (ii) sharp turns, whereas the upper layer contains the equations describing their corresponding duration. Therefore, our modelled trajectories are chains of consecutive sharp turns and crawls that depending on the implementation used can appear according to a stationary (time-independent) or a non-stationary (time-dependent) process. This double approximation allowed us to understand if anomalous diffusion in *C. elegans* comes from the high variability in the duration of these motor behaviors (stationary version), or instead it comes from their temporal unfolding (non-stationary version).

6.3.1 Langevin equation of crawls

Based on Schienbein and Gruler (1993), we modelled movement during crawls by considering the velocity dynamics in the direction of motion $\mathbf{e}_h$ and in the normal direction $\mathbf{e}_\phi$ (Figure 6.2). Specifically, we assumed there is a propulsive force γ acting in the direction of motion, which pulls the speed towards a constant mean speed $\bar{v}_c$, and a curvature force β_c acting on the normal direction, which induces a constant curvature in the motion (similar to Gurarie et al. (2017)). Finally, and based on the results obtained in Chapter 5, we assumed there is Gaussian white noise acting in the direction of motion and Cauchy white noise acting in the normal direction (this latter type of noise was implemented as in Leccardi (2005)). In consequence, the model reads:

$$\frac{d\mathbf{v}}{dt} = -\gamma(v - \bar{v}_c)\mathbf{e}_h - \beta_c \mathbf{e}_\phi + \xi_c \mathbf{e}_h + \zeta_c \mathbf{e}_\phi \ , \tag{6.2}$$

where γ is the propulsive force, β_c the curvature force, ξ_c is the intensity of the Gaussian white noise, and ζ_c is the intensity of the Cauchy white noise.

Alternatively, the very same model can be expressed in two equations describing the dynamics in the direction of motion and in the normal direction by separate (see the derivation in Box 2):

$$\frac{dv}{dt} = -\gamma(v - \bar{v}_c) + \xi_c \ , \tag{6.3}$$

$$\frac{d\phi}{dt} = \frac{\beta_c + \zeta_c}{v} \ . \tag{6.4}$$

Here, it is worth to mention Eq. (6.4) shows us the physical dependency between speed and turn. So that high turns occur when speed is low and low turns occur when speed is high. Importantly in Figure 5.1C from Chapter 5, we saw a clear evidence of this phenomenon happening in *C. elegans*.

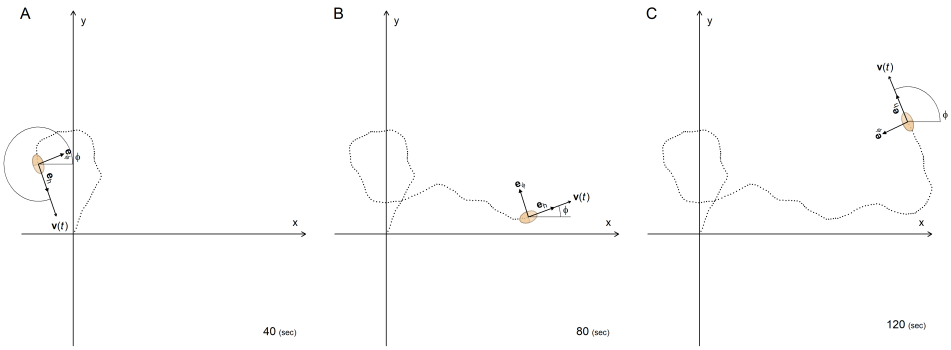


Figure 6.2: Schematic representation of a trajectory generated by the Langevin equation, measured at three different times (A-C). In all panels $\mathbf{e}_h$ and $\mathbf{e}_\phi$, denote the unit vectors in the direction of motion and in the normal direction, respectively, ϕ denotes the direction of motion (sometimes referred as bearing) and $\mathbf{v}(t)$ is the velocity vector. Note that the length of $\mathbf{v}(t)$ is not constant but proportional to speed. The figure was adapted from Schienbein and Gruler (1993).

Box 2: Deriving the Langevin equation of crawls

In order to express the Langevin model in two equations describing the dynamics of speed and turn by separate, one need to recall $\mathbf{v} = v\mathbf{e}_h$, and take its time derivative. So that:

$$\frac{d\mathbf{v}}{dt} = \frac{dv}{dt}\mathbf{e}_h + v\frac{d\mathbf{e}_h}{dt} = \frac{dv}{dt}\mathbf{e}_h + v\frac{d\phi}{dt}\mathbf{e}_\phi \ , \tag{6.5}$$

where in the first step, we used the derivation of the product, and in the second step we took into account that the time derivative of $\mathbf{e}_h$ equals the time derivative of the direction ϕ in the normal direction $\mathbf{e}_\phi$. Further intuition on this last step can be found by explicitly writing the form of $\mathbf{e}_h$, and taking its time derivative:

$$\mathbf{e}_h = (\cos\phi,\, \sin\phi) \ ,$$

$$\frac{d\mathbf{e}_h}{dt} = (-\sin\phi,\, \cos\phi)\,\frac{d\phi}{dt} = \mathbf{e}_\phi\frac{d\phi}{dt} \ ,$$

where, in the last line, we first applied the chain rule and then the definition of $\mathbf{e}_\phi$. Finally, if we substitute the result obtained in Eq. (6.5) in the left-hand side of Eq. (6.2) and multiply the whole expression by $\mathbf{e}_h$ we obtain Eq. (6.3). If instead, we multiply the same expression by $\mathbf{e}_\phi$ we obtain Eq. (6.4).

6.3.2 Langevin equation of sharp turns

Following our approach for crawls, we modelled sharp turns using a Langevin equation. However, given their brief duration, we reduced the model to its bare essentials, akin to the simplicity found in the Langevin equation outlined by Metzler et al. (2014). Our aim was to replicate the fundamental features of sharp turns for establishing connections between microscopic movement dynamics and overarching spreading scales. In particular, we assumed:

- A Weibull distribution of speeds, to reproduce the bias towards low speed values observed in sharp turns.
- A Cauchy distribution of angular velocities, accounting for the large probability of observing extreme turns.
- A negative coupling between speed and angular velocity, to ensure large speed values were associated with small turns and vice versa.

Based on that, we modelled sharp turns using two separate Langevin equations. One for the speed and one for the angular velocity. The speed is modelled differently than we did for crawls. Specifically, we modelled speed as a random variable that follows a Weibull distribution. In consequence, the model reads:

$$\frac{d\mathbf{r}}{dt} = v\mathbf{e}_h, \tag{6.6}$$

where $\mathbf{r}$ is the position vector and v an independent and identically distributed random variable (i.i.d.) representing the speed, that follows a Weibull distribution with scale θ and shape φ.

On the contrary, we modelled angular velocity in the same way as for crawls Eq. (6.4):

$$\frac{d\phi}{dt} = \frac{\beta_s + \zeta_s}{v} \ , \tag{6.7}$$

where β_s is a curvature force inducing a curvature in the motion, and ζ_s is Cauchy white noise affecting the motion in the normal direction.

6.3.3 Coupling crawls and sharp turns

Finally, we integrated the Langevin equations of crawls and sharp turns into a complete model of movement using two distinct approaches. First, we assumed a stationary random process based on the distribution of durations of sharp turns and crawls (fitted in Chapter 5, Section 5.3.5). Second, we assumed a non-stationary process based on the time-dependent rates of the two movement modes (fitted in Chapter 5, Section 5.3.6). Importantly, this double approximation allowed us to understand if anomalous diffusion in *C. elegans* arose from the high variability in the duration of sharp turns and crawls (stationary model), or instead it arose from their temporal unfolding (non-stationary model).

Following results obtained in Chapter 5, we modelled the duration of crawls for the stationary approach according to a double exponential distribution, defined as:

$$f(\tau) = p \ \frac{1}{t_1} e^{-\tau/t_1} \ + \ (1-p) \ \frac{1}{t_2} e^{-\tau/t_2} \ , \tag{6.8}$$

where τ is the duration, p is a probability weight, and t_1, t_2, are the scales of the two exponential functions.

Similarly, we modelled the duration of sharp turns according to a gamma distribution, defined as:

$$f(\tau) = \frac{1}{\Gamma(\kappa)\,\vartheta^{\kappa}}\;\tau^{\,\kappa-1}\,e^{-\tau/\vartheta}\;, \tag{6.9}$$

where τ is the duration, κ is the shape parameter, ϑ is the scale, and Γ is the gamma function. To generate stationary synthetic trajectories we sequentially sampled these distributions and chained the corresponding sharp turns and crawls accordingly, assuming continuity in both speed and direction.

For the non-stationary approach, we modelled the rate of crawls as an exponential function of the form:

$$\lambda_c(t) = a_c e^{k_c t} + b_c\;, \tag{6.10}$$

where a_c is the excess initial rate, k_c is the memory rate, and b_c is the baseline rate. Here is important to note that $k_c < 0$ yields a decaying trend in the rate of crawls, while $k_c > 0$ yields and increasing trend (Figure 6.3A). Meaning that when k_c is negative the duration of crawls increases over time, and when k_c is positive the duration of crawls decreases. We allowed k_c to be either negative or positive to uncover the whole variability observed in empirical data. In the context of our experiments, $k_c < 0$ indicated a degradation of the memory associated with the preconditioning of the worms, which in turn set the initial conditions of the experiment (worms grew feeding *ad libitum*). Otherwise, $k_c > 0$ indicated a reinforcement of this initial condition.

Following a similar approach, we modeled the rate of sharp turns employing an exponential function that could reflect the decrease in durations (rise in the rate) of sharp turns over time:

$$\lambda_s(t) = a_s\left(1 - e^{k_s t}\right) + b_s\;, \tag{6.11}$$

where a_s is the limiting or saturation rate, k_s is the memory rate, and b_s is the baseline rate. Although Eq. (6.11) is not properly defined when k_s is positive (it yields negative rates as time increases), we included the possibility to model positive k_s to account for the fact that some individuals, especially in mutant strains DA609 and VC125, exhibited and increasing duration of sharp turns (Figure 6.3B). Whenever we needed to simulate the trajectories of this special cases we assumed the synthetic worm entering a permanent sharp turn before reaching $\lambda_s(t) = 0$. Since the inverse of a rate (i.e., a duration) tends to infinite as it approaches 0.

To implement this version of the model we relied on the fact that inverse rates are durations. Thus, assuming a walk starting in crawling mode, we got the first duration from an exponential distribution with rate $\lambda_c(0) = a_c + b_c$. This, led the walk to time t, where we obtained the duration of the subsequent sharp turn from an exponential distribution with rate $\lambda_s(t) = a_s \left(1 - e^{k_s t}\right) + b_s$, so on and so forth.

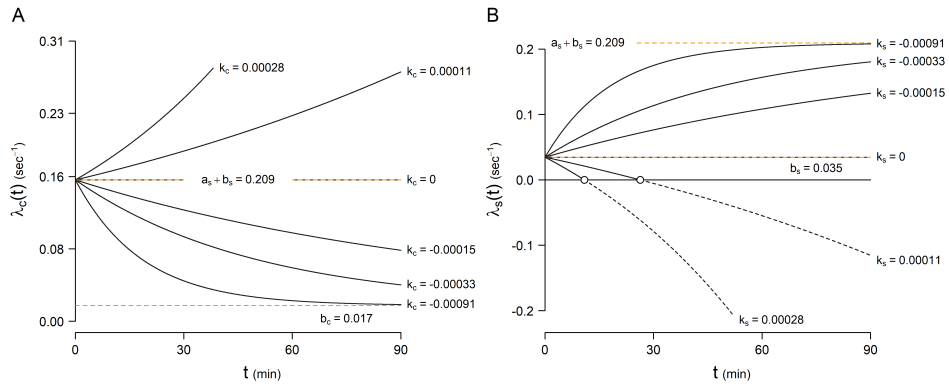


Figure 6.3: Graphical representation of Eq.(6.10) and Eq.(6.11) for various values of the memory rates k_c (crawls) and k_s (sharp turns). (A) Behavior of the time-dependent rate of crawls, $\lambda_c(t) = a_c e^{k_c t} + b_c$. (B) Behavior of the time-dependent rate of sharp turns, $\lambda_s(t) = a_s \left(1 - e^{k_s t}\right) + b_s$. In both panels, the first orange line along the vertical axis corresponds to the baseline rate, b_c and b_s, while the second one represent the excess initial rate $a_c + b_c$ (in Panel A) and the carrying capacity $a_s + b_s$ (in Panel B). Note that b_c sets the minimal value of $\lambda_c(t)$, reached as t tends to infinity, while b_s establishes the initial rate of sharp turns, offering insight into the behavior of $\lambda_s(t)$ for short time intervals when $k_s > 0$.

6.4 Model calibration

6.4.1 Parameter estimation

Parameter estimates were carried at the level of individual and derived from speed or angular velocity of the empirical trajectories. When possible, we estimated them directly from data, otherwise we employed custom numerical fitting procedures (R package *stats*). In all cases, we segmented trajectories into sharp turns and crawls based on the behavioral annotation developed in Chapter 5 Section 5.3.3. Thus, we fit or estimate model parameters conditioned with these behavioral labels.

Summary tables

Here, we present strain-averaged tables with all the parameters used in the stationary and the non-stationary models. Table 6.1 details values for the Langevin

equation of crawls, while Table 6.2 does it for the sharp turns. Table 6.3 reports the parameters required to reproduce the distribution of duration associated with sharp turns and crawls, employed in the stationary model. Finally, Table 6.4 details their time-dependent rates, utilized in the non-stationary model.

Table 6.1: Strain-averaged parameters of the Langevin equation of crawls. The propulsive force γ is expressed in $\sec^{-1}$, the mean speed $\bar{v}_c$ in $\mathrm{mm}\,\sec^{-1}$, and the curvature force β_c in $\mathrm{mm}\,\mathrm{rad}\,\sec^{-2}$, which is reported in absolute terms. Both Gaussian (ξ_c) and Cauchy (ζ_c) white noises are dimensionless. N indicates the number of individuals.

Strain	γ	$\bar{v}_c$	ξ_c	$\lvert\beta_c\rvert$	ζ_c	N
N2	0.724	0.150	0.030	0.0074	0.00059	48
CB4856	0.796	0.183	0.033	0.0063	0.00078	30
DA609	0.645	0.171	0.033	0.0660	0.00072	24
VC125	0.683	0.151	0.037	0.0636	0.00070	24

Table 6.2: Strain-averaged parameters of the Langevin equation of sharp turns. The scale of the speed θ is expressed in $\mathrm{mm}\,\sec^{-1}$ and the curvature force β_s in $\mathrm{mm}\,\mathrm{rad}\,\sec^{-2}$, which is reported in absolute terms. Both the shape φ and the Cauchy white noise ζ_c are dimensionless quantities. N indicates the number of individuals.

Strain	θ	φ	$\lvert\beta_s\rvert$	ζ_s	N
N2	0.103	1.778	0.0015	0.00026	48
CB4856	0.118	1.688	0.0015	0.00031	30
DA609	0.104	1.728	0.0011	0.00027	24
VC125	0.099	1.671	0.0016	0.00029	24

Table 6.3: Strain-averaged parameters describing the duration of sharp turns and crawls. Columns κ and ϑ denote the shape and scale of the gamma distribution (sharp turns), while p, t_1, and t_2 denote the probability weight and the two distinct scales of the double exponential distribution (crawls). N indicates the number of individuals.

	Gamma (sharp turns)		Double exponential (crawls)			
Strain	κ	ϑ (sec)	p	t_1 (sec)	t_2 (sec)	N
N2	1.90	5.46	0.54	16.16	84.40	48
CB4856	2.46	4.61	0.51	13.66	72.00	30
DA609	2.18	4.45	0.49	14.60	56.33	24
VC125	1.92	5.06	0.53	11.75	44.46	24

Table 6.4: Strain-averaged parameters of the exponential functions describing the rate of sharp turns and crawls. Column a_s indicate the value of the limiting rate of sharp turns, a_c is the excess initial rate of crawls, k_s and k_c represent the memory rates, and b_s and b_c are the baseline rates. All of them are expressed in sec^{-1}. N reports the number of individuals used to compute the averages.

	Sharp turns			Crawls			
Strain	a_s	k_s	b_s	a_c	k_c	b_c	N
N2	0.175	-0.0017	0.035	0.139	-0.0021	0.017	48
CB4856	0.182	-0.0024	0.056	0.103	-0.0021	0.015	30
DA609	0.160	-0.0005	0.095	0.110	-0.0029	0.026	24
VC125	0.153	-0.0013	0.061	0.159	-0.0017	0.027	24

6.4.2 Calibration of crawls

Mean speed $\bar{v}_c$, and standard deviation σ_c

Parameters $\bar{v}_c$ and σ_c are simply the mean and standard deviation of the speed distribution associated with crawls. Thus, we obtained them directly from fits of Eq. (5.1), in Chapter 5.

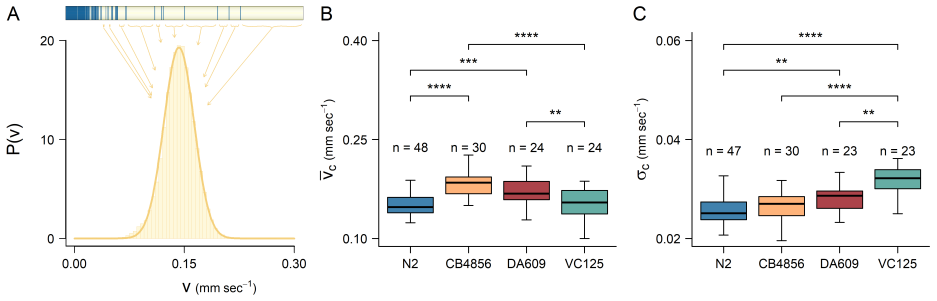


Figure 6.4: Estimation of $\bar{v}_c$ and σ_c. (A) Distribution of speed associated with crawls of an N2 individual. In the ethogram, sharp turns are depicted in blue and crawls in yellow. (B-C) Box-plots of $\bar{v}_c$ and σ_c for the strains under study (N2, CB4856, DA609, VC125). In both panels significant differences between strains are denoted with a starred brace (t-test with Holm-Bonferroni correction $\alpha = 0.05$).

Propulsive force γ and the Gaussian white noise ξ_c

According to Schienbein and Gruler (1993), the propulsive force in the direction of motion of the Langevin equation is the decaying rate of the speed correlations, which can be defined as:

$$\langle (v(t) - \bar{v})(v(t+\tau) - \bar{v}) \rangle = \frac{\xi}{2\gamma} e^{-\gamma\tau} \ , \tag{6.12}$$

where $v(t)$ is the speed at time t, $\bar{v}$ is the mean speed, τ is the time span, and $\langle \cdot \rangle$ denotes averaging by τ. At the limit $\tau \to 0$ we obtain the following expression:

$$\langle (v(t) - \bar{v})(v(t+0) - \bar{v}) \rangle \equiv \frac{1}{n} \sum_{i=1}^{n} (v_i - \bar{v})^2 = \frac{\xi}{2\gamma} \ ,$$

since averaging correlations of a stochastic process when $\tau = 0$ (left-hand side of the equation), is equivalent to averaging squared deviations (the right-hand side of the equation turns out to be the expression for the variance), which in turn, can be expressed as a function of γ and ξ. Therefore, $\sigma^2 = \frac{\xi}{2\gamma}$. Indeed, Schienbein and Gruler (1993) showed that if the propulsive force γ and the Gaussian white noise ξ are driven by the same mechanism, then $\xi = \sqrt{2\gamma}\sigma$, where σ is the standard deviation. In fact, we tested that this relationship holds, so that we computed ξ_c (Gaussian noise of crawls) in terms of γ (propulsive force) by using Eq. 6.12 and σ_c (standard deviation of the speed of crawls), already computed in the previous section.

To obtain estimates of ξ_c we first needed to compute γ, and this required a series of steps as follows. Given that speed increased with time, we detrended its signal before computing the correlations. To do so, we roll averaged speed values using a point-wise running window of 600 seconds width, and subtracted the obtained time-dependent average from the raw signal (Figure 6.5A). Then, we computed the correlations of crawls as indicated in Eq. (6.12), but assuming $\bar{v}_c = 0$, as we had already detrended the signal. Importantly, a t-test comparing the standard deviation of speed before and after detrending indicated there was no significant difference between the two (p-value $= 0.77$, $\alpha = 0.05$). In the same line of reasoning, the values of the fluctuations with and without applying the detrending were strongly correlated (Pearson correlation coefficient, $\rho = 0.84$) and the magnitudes did not change (Figure 6.5C).

On crawling mode, speed correlations showed three relevant scales, that we also observed when computing speed correlations without factorizing trajectories into crawls and sharp turns (Appendix C). This means that the overall speed correlation

pattern of *C. elegans* is driven by crawls. The largest correlations occurred within 1 second, decaying fast. However, the decaying pattern is interrupted between 1 and 3 seconds where correlations showed a peak associated to the sinusoidal and cyclic body movement of the crawling behavior. From 3 to 125 seconds (2 minutes) correlations are of smaller magnitude and gradually weaken until reaching the value zero (Figure C.2D).

We estimated the parameter γ by fitting only the first velocity correlation regime. Hence, for the sake of simplicity, and based on other analysis, we assumed the initial decay (correlations covering less than 1 second) to involve the most prevalent correlation, as subsequent scales involved much smaller correlations (Appendix C). In particular, we fitted the model $f(\tau) = ae^{-k\tau} + c$ using non-linear least squares (R packages *nls.multstart* and *minpack.lm*, (Padfield et al., 2021)), where the rate k gave us the estimate of γ (Figure 6.5E). To made sure the fitting process was consistent with the theoretical formalism explained above, we tested that $a + c = \sigma_c^2$. The span of the first exponential regime (around 1 second) was determined using the R package *segmented* (Muggeo, 2003). Finally, we estimated ξ_c in terms of the propulsive force γ and the standard deviation of the speed σ_c associated with crawls (Figure 6.5F)

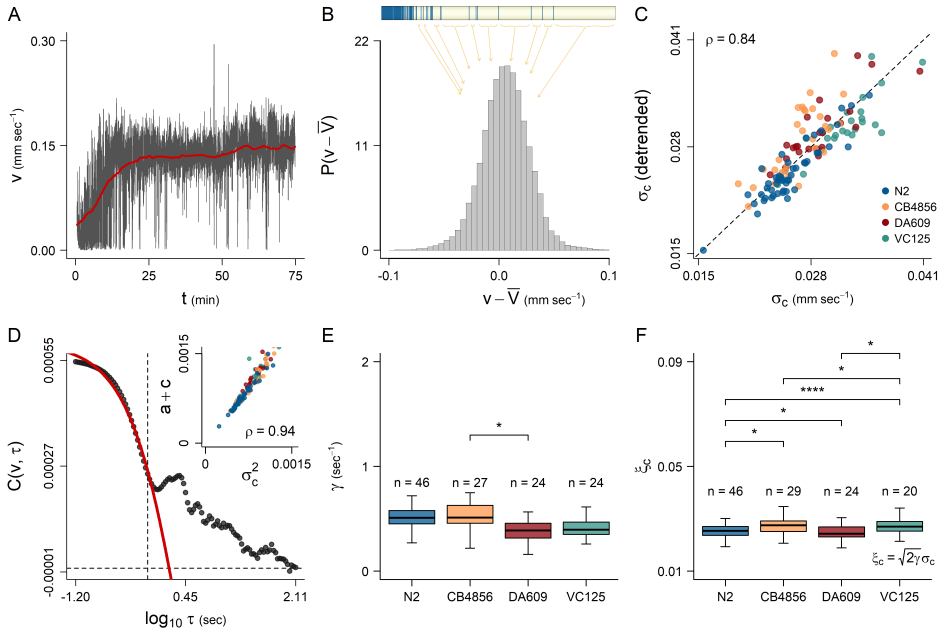


Figure 6.5: Estimation of the propulsive force γ and the intensity of the Gaussian white noise in the direction of motion ξ_c. (A) Temporal series of speed of an N2 individual with the rolling average $\bar{V}$ depicted in red. (B) Histogram of the speed fluctuations (or detrended speed values) of crawls. In the ethogram, sharp turns are depicted in blue and crawls in yellow. (C) Pearson correlation of the standard deviation of speed (of crawls) with and without detrending. The dashed line indicates exact equality. (D) Speed correlations of crawls with the first exponential regime fitted in red. The dashed lines indicate the zero coordinates. The inset shows that, as expected from theory, the correlation at $\tau = 0$ and the variance of the speed can be assumed to be equal. (E, F) Box-plots comparing the estimated γ and ξ_c of the strains under study. ξ_c was estimated assuming $\xi_c = \sqrt{2\gamma}\sigma_c$, so that computed its value by using both the propulsive force γ and the standard deviation of the speed associated with crawls σ_c. In both panels significant differences between strains are denoted with a starred brace (t-test with Holm-Bonferroni correction $\alpha = 0.05$).

Curvature force β_c

We estimated the curvature force of crawls as the product of their mean speed and mean angular velocity. Thus, we assumed:

$$\beta_c = \langle v \frac{d\phi}{dt} \rangle \approx \bar{v}_c \langle \frac{d\phi}{dt} \rangle \ ,$$

where $\langle \frac{d\phi}{dt} \rangle$ is the mean, or location $\bar{\omega}_c$, of the angular velocity distribution fitted to crawls in Chapter 5, Section 5.3.4. In our data the value of β_c was clearly determined by $\bar{\omega}_c$ and not by $\bar{v}_c$ (Figure 6.6B).

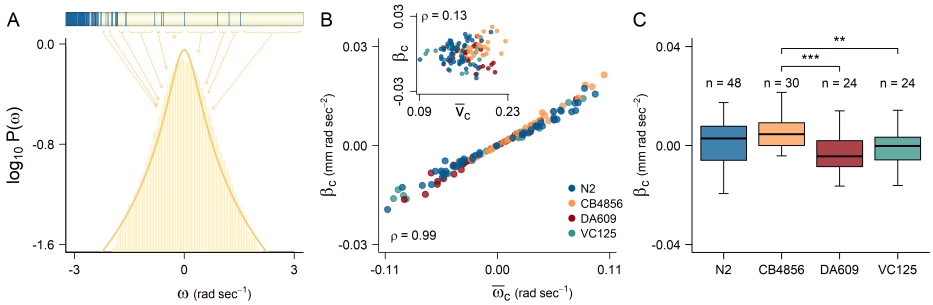


Figure 6.6: Estimation of the curvature force β_c. (A) Distribution of angular velocity associated with crawls of an N2 individual. In the ethogram, sharp turns are depicted in blue and crawls in yellow. (B) Pearson correlation of the location parameter $\bar{\omega}_c$ of the angular velocity distribution and the curvature force β_c. The inset shows the same relation but for the mean speed $\bar{v}_c$. (C) Box-plots comparing β_c among strains. Significant differences are denoted with a starred brace (t-test with Holm-Bonferroni correction $\alpha = 0.05$).

Cauchy white noise ζ_c

The intensity of the noise acting in the normal direction of motion could not be directly estimated from data, so that we estimated it numerically. Such magnitude represents the width of the fluctuations around the mean angular velocity of the crawls. Therefore, since we had a Cauchy distribution of angular velocities, we assumed $\zeta_c = f(\eta_c)$, where η_c is the scale parameter or variance of the angular velocity distribution, fitted for crawls in Eq. (5.3) from Chapter 5 (see Figure 5.5B). The crux of this approach was that a direct measure of η_c could be obtained from trajectory data, either empirical or synthetic. Based on this fact, we built an optimization algorithm to find the value of ζ_c minimizing the difference between the η_c of the model trajectories and the η_c of the empirical trajectories. To keep things as simple as possible while maintaining a certain mathematical formalism we assumed $\zeta_c = C\eta_c$, and optimized the constant C. However, one can assume almost any arbitrary function and obtain similar results (Figure 6.7).

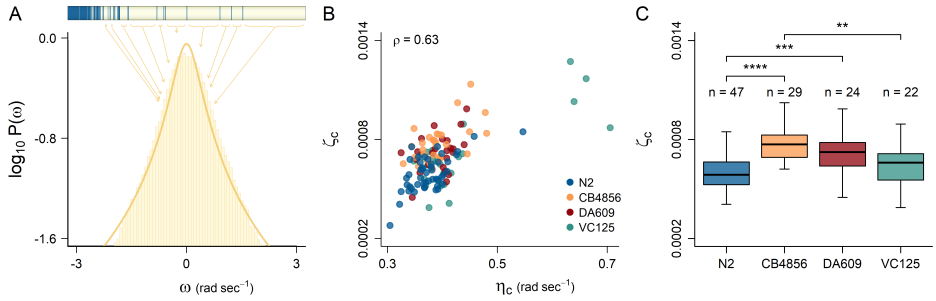


Figure 6.7: Estimation of the Cauchy white noise ζ_c associated with crawls. (A) Distribution of angular velocity of an N2 individual. In the ethogram, sharp turns are depicted in blue and crawls in yellow. (B) Pearson correlation of the scale parameter η_c of the angular velocity distribution and ζ_c. (C) Box-plots comparing ζ_c among strains. Significant differences between strains are denoted with a starred brace (t-test with Holm-Bonferroni correction $\alpha = 0.05$).

6.4.3 Calibration of sharp turns

Scale θ and shape φ of the Weibull random speed

We modelled the speed of the sharp turns as an i.i.d. Weibull random variable with scale θ and shape φ parameters (Figure 6.8). Their values were obtained from the fits of the speed distribution of sharp turns i.e., Eq. (5.1) (Chapter 5) and the inverse Weibull PDF were used as random number generators to obtain speed values sequences in sharp turn modes.

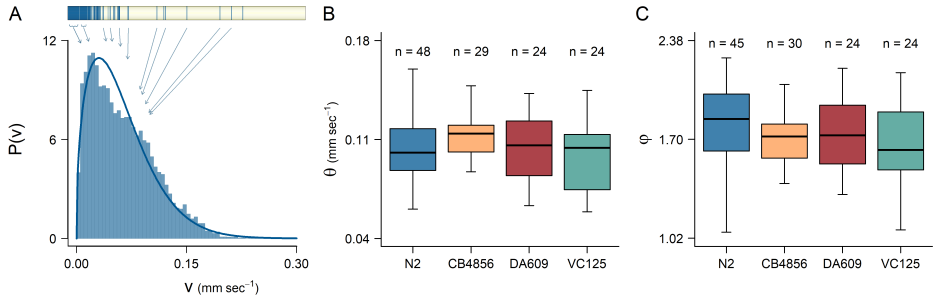


Figure 6.8: Estimation of the scale θ and shape φ parameters of the Weibull speed distribution associated with sharp turns. (A) Speed distribution of the sharp turns of an N2 individual. In the ethogram, sharp turns are depicted in blue and crawls in yellow. (B-C) Box-plots comparing θ and φ among strains. No significant differences were found (t-test with Holm-Bonferroni correction $\alpha = 0.05$).

Curvature force β_s

The curvature force of sharp turns could not be directly estimated from data, so we approximated it numerically. Indeed, the relation we assumed was the same as for crawls, namely, that β_s equals the mean speed times the mean angular velocity, but here we needed to add a scaling factor to ensure the model output distributions were statistically equivalent to empirical ones. In practice, we measured β_s as:

$$\beta_s = \langle v \rangle \langle \frac{d\phi}{dt} \rangle \, C \ ,$$

where $\langle v \rangle$ is the mean value of the Weibull speed distribution associated with sharp turns, $\langle \frac{d\phi}{dt} \rangle$ is the mean value of the corresponding angular velocity distribution, and C is an arbitrary constant ensuring there is no offset between model input and model output parameters (re-measured from the resulting trajectories). The two mean values were obtained from the fits of Eq. (5.1) and Eq.(5.3) in Chapter 5. In particular, $\langle v \rangle$ was computed using the mean formula of the Weibull distribution: $\theta \, \Gamma(1 + \frac{1}{\varphi})$, where θ is the scale parameter, φ is the shape, and Γ is the Gamma function, while $\langle \frac{d\phi}{dt} \rangle$ was directly set as the location parameter $\bar{\omega}_s$ of the angular velocity distribution fitted to sharp turns. Finally, the constant C was obtained by minimizing the difference between model input and model output parameters like we did for the intensity of the Cauchy white noise.

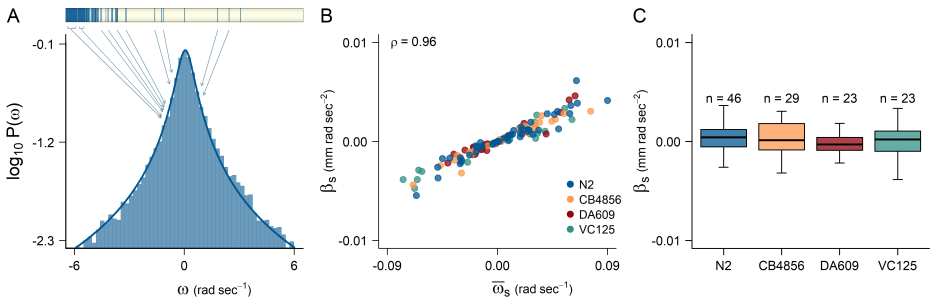


Figure 6.9: Estimation of the curvature force β_s associated with sharp turns. (A) Distribution of angular velocity of an N2 individual. In the ethogram, sharp turns are depicted in blue and crawls in yellow. (B) Pearson correlation of the location parameter $\bar{\omega}_s$ of the angular velocity distribution and the curvature force β_s. (C) Box-plots comparing β_s among strains. No significant differences were found (t-test with Holm-Bonferroni correction $\alpha = 0.05$).

Cauchy white noise ζ_s

To estimate the intensity of the Cauchy white noise on sharp turns we performed the same fitting procedure as we did for crawls. That is, we assumed $\zeta_s = C\eta_s$ and optimized the constant C to find the value of ζ_c minimizing the difference between the η_s of the model trajectories and the η_s of the empirical trajectories (Figure 6.10).

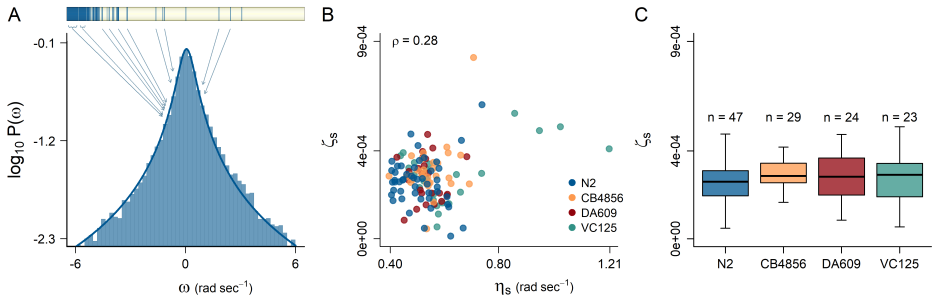


Figure 6.10: Estimation of the Cauchy white noise ζ_s associated with sharp turns. (A) Distribution of angular velocity of an N2 individual. In the ethogram, sharp turns are depicted in blue and crawls in yellow. (B) Pearson correlation of the scale parameter η_s of the angular velocity distribution and ζ_s. (C) Box-plots comparing ζ_s among strains. Significant differences between strains are denoted with a starred brace (t-test with Holm-Bonferroni correction $\alpha = 0.05$).

6.5 Model validation

6.5.1 Microscopic movement properties

We validated our implementation of the Langevin equations of crawls and sharp turns (separately), along with the methods used for parameter estimation, by assessing three key properties. First, we made sure all model input parameters could be re-estimated from model output trajectories without significant offset. Second, we verified the difference between empirical parameter estimates and synthetic parameter estimates was statistically zero. Third, we confirmed the variability around synthetic parameter estimates followed a normal distribution. These properties were checked for each of the 9 parameters (see Table 6.1 and Table 6.2 in Section 6.4.1) associated with each of the 126 individuals under study.

In particular, we ran the following individual-based protocol. We simulated 1000 crawl trajectories and 1000 sharp turn trajectories of 20 minutes length each, at the empirical frame-rate (33Hz). We then re-estimated the corresponding parameters from these trajectories and calculated the error distribution as

$Z_i = (S_i - E_i)/\sigma_i$; where i is one of the 9 parameters, E_i is the empirical estimate of i (model input), S_i is the corresponding synthetic estimate of i (obtained from the model output), and σ_i is the standard deviation of S_i. Through this procedure, we confirmed that both the implementations of the Langevin equations and the methods used for parameter estimation were effectively unbiased, as evidenced by the fact that, in all cases, estimate errors followed a normal distribution with $\mu \approx 0$ and $\sigma \approx 1$ (Figure 6.11).

6.5.2 Mesoscopic movement properties

After confirming the accurate implementation of sharp turns and crawls at the microscopic scale, we proceeded to validate the model at the mesoscopic scale, covering time spans from seconds to hours. At this stage, we transitioned to testing the full models, namely the stationary and non-stationary models. This allowed us to assess how effectively we integrated the Langevin equations governing crawls with those governing sharp turns. To do so, we generated a total of 100 synthetic trajectories for each *C. elegans* individual and type of model (stationary and non-stationary), matching both the empirical frame rate (33Hz) and corresponding track length, which was 90 minutes in average. This resulted in two populations of N = 12600 trajectories, distributed across four strains: N2 ($n = 4800$), CB4856 ($n = 3000$), DA609 ($n = 2400$), and VC125 ($n = 2400$).

Following the same methodology used for empirical data (see Chapter 5, Sections 5.3.5 and 5.3.6 for detailed methods), we characterized the temporal unfolding of sharp turns and crawls, and their associated distributions of net displacement, duration, and turning angle (Figure 6.12). This analysis confirmed two key findings: (i) the non-stationary model accurately reproduced the temporal unfolding of sharp turns and crawls, while the stationary version failed in this regard (Figures 6.12A, D), and (ii) both versions of the model effectively replicated the distributions of net displacement, duration, and turning angle associated with sharp turns and crawls (Figure 6.12B, C, E-G). The results obtained for strains CB4856, DA609, and VC125 can be found in the Appendix D.2.

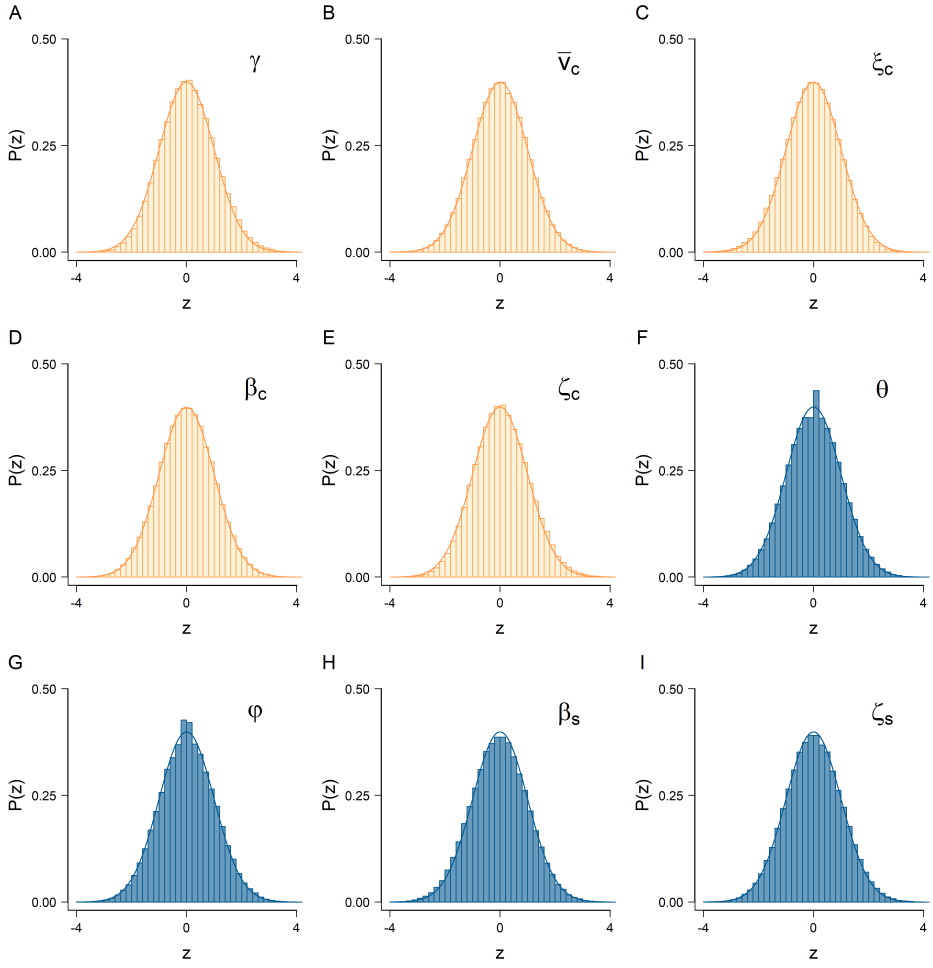


Figure 6.11: Strain-pooled z-scores of the model parameters. In all cases, z-scores were computed at the level of individual. Once standardized, they were grouped by parameter into a single distribution that included all individuals. Panels A-E, show the results of the parameters associated with crawls. Panels F-I, show the results of the parameters associated with sharp turns. For parameters that could be confused between crawls and sharp turns we used a subscript. (A) Propulsive force γ. (B) Mean speed $\bar{v}_c$. (C) Gaussian white noise in the direction of motion ξ_c. (D) Curvature force β_c. (E) Cauchy white noise in the normal direction ζ_c. (F) Scale parameter of the Weibull distribution of speed of sharp turns θ. (G) Shape parameter of the Weibull distribution of speed of sharp turns φ. (H) Curvature force β_s. (I) Cauchy white noise in the normal direction ζ_s.

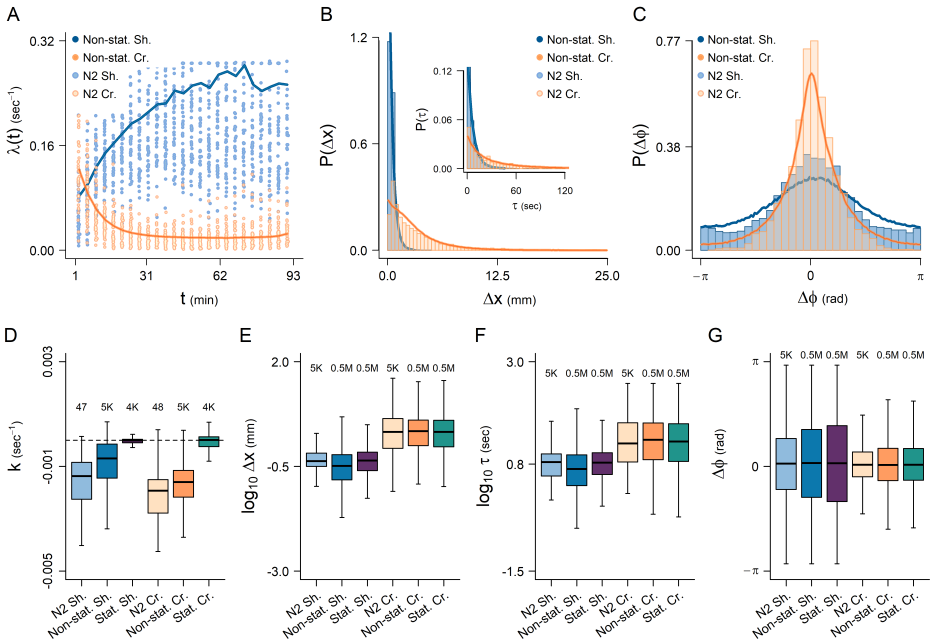


Figure 6.12: Validation of the stationary and the non-stationary models for the strain N2 (48 individuals) at mesoscopic scale. Legend names indicate sharp turns (sh) and crawls (cr) of empirical data (N2), non-stationary, and stationary models (in this case also parameterized based on N2 individuals). (A) Temporal unfolding of sharp turns and crawls. (B) Distributions of net displacement and duration (inset). (C) Distribution of turning angle. (D) Boxplots comparing the memory rate of sharp turns (k_s) and crawls (k_c) in empirical and synthetic tracks. (E-G) Boxplots of the variables shown in panels B and C.

6.5.3 Macroscopic movement properties

Assessing the accuracy of the models

Finally, we validated the macroscopic movement properties of both the stationary and the non-stationary models by focusing on the mean square displacement (MSD). Our first objective was determining if the models could reproduce the different diffusive regimes observed in *C. elegans* (see Chapter 4 for a detailed analysis of empirical MSD curves and Section 6.2, for a summary description of the methods used to estimate MSD parameters). To address that, we created a synthetic replicate of the empirical tracks for each type of model (stationary and non-stationary). These replicates were generated by simulating 1000 trajectories per *C. elegans* individual, from which we selected the one minimizing the sum of squares between synthetic and empirical point-wise distances to the origin. Once all empirical trajectories were adequately replicated, we grouped them by strain and type of model,

and computed the MSD. As shown in Figure 6.13 the non-stationary model could reproduce the MSD of the empirical trajectories much better than the stationary model. Interestingly, the difference between the two models occurred mainly at the early stages of the spreading ($\log_{10}(t) < 3$, or $t < 1000$ sec), when the bulk of non-stationary trends took place. Afterwards they converged in almost identical trends, suggesting the complex patterns of *C. elegans* MSD curves came from the unfolding of sharp turns and crawls and not just as a consequence of the small ensemble sizes. Tables reporting differences between empirical and model based MSD curves can be found in the Appendix D.3.

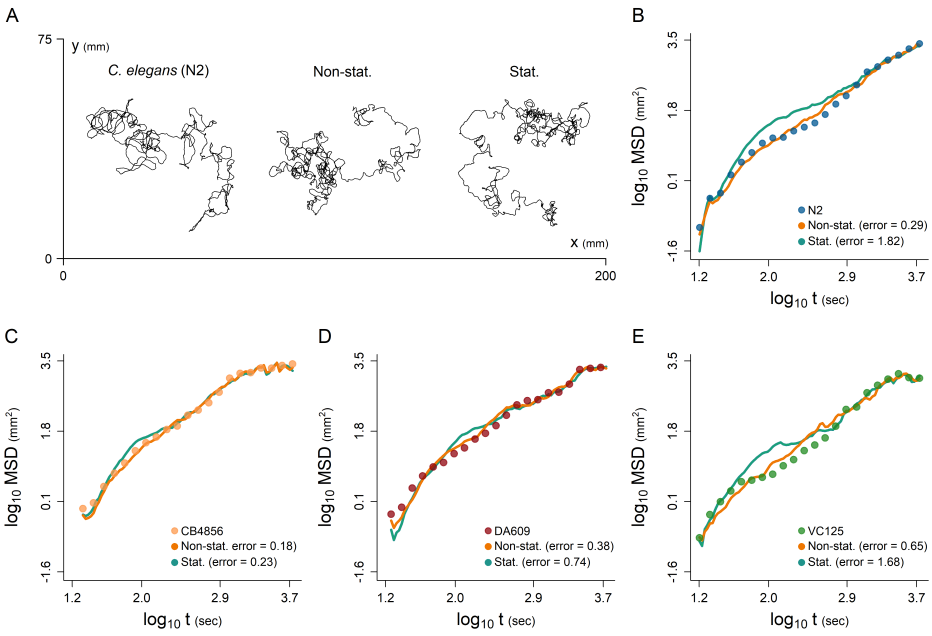


Figure 6.13: Replication of the empirical mean square displacement (MSD) curves. (A) Example trajectory of an N2 individual with its corresponding synthetic copies, generated with the stationary and the non-stationary models. In both cases, these copies were obtained from a population of 1000 trajectories by minimizing the sum of squares between the synthetic and the empirical point-wise distances to the origin. (B-E) MSD curves for the different strains i.e., N2, CB4856, DA609 and VC125, and their replicates. The error value reported in panels B-E equals the sum of squares between empirical MSD and synthetic MSD curves.

Assessing the average individual of the strains

Additionally, we sought to determine whether the MSD variability at level of strain could be approximated considering the "average" individual of the strains, with all model parameters being averaged as reported in Tables 6.1-6.4. To investigate

this, we first parameterized the MSD curves of both the particular individuals and the average individuals. Next, we pooled all parameter estimates by strain, and finally assessed the differences in spreading properties of the particular individuals against the average individuals. As shown in Figure 6.14, the parameters obtained with the average N2 where statistically the same as those encompassing all N2 individuals, especially in the long-term limit, represented by third and fourth MSD regimes (Figure 6.14A-H). On the contrary, in the short scale limit (first and second regimes) (Figure 6.14I-P) pooling individual variability at the strain level, hence assuming an "average" strain behavior, produces discrepancies in the spreading patterns. Noteworthy, the above results can be applied for both the non-stationary and the stationary models, and for all strains under study (results shown in Appendix D.3). All in all, this analysis suggests that we can rely on the concept of an average strain individual to study the mechanisms by which *C. elegans* generates long-term superdiffusion.

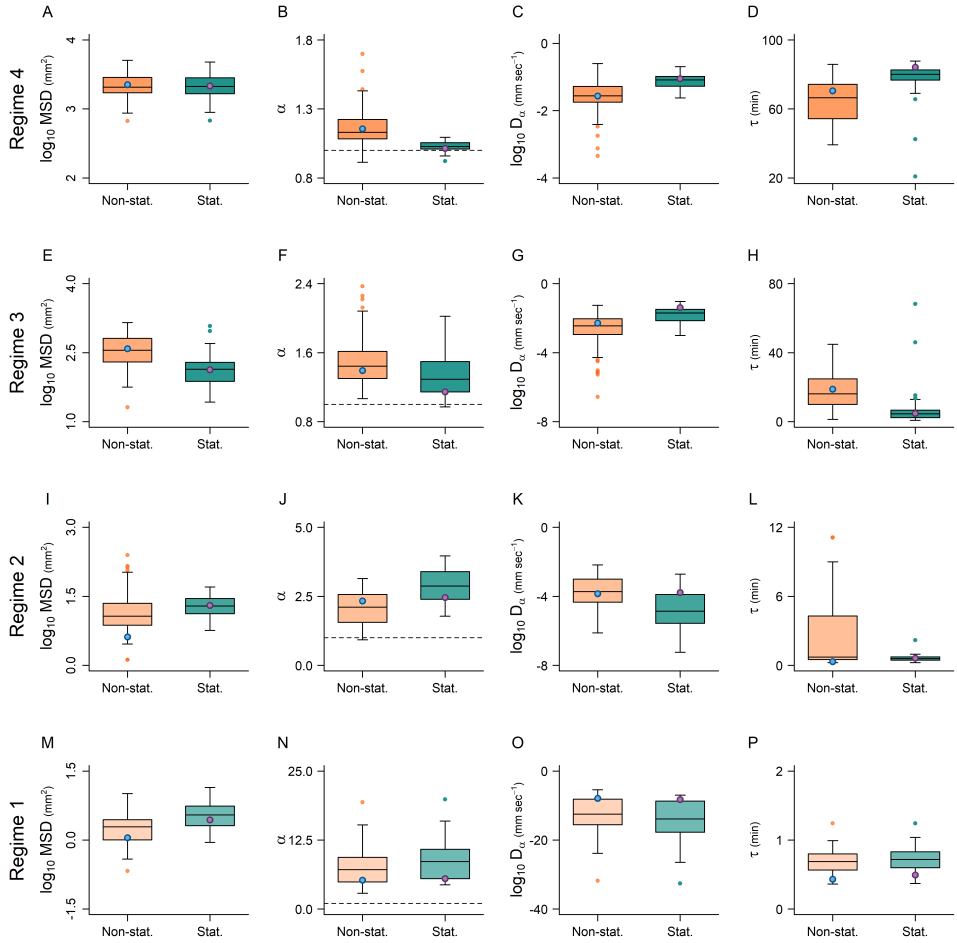


Figure 6.14: Mean square displacement (MSD) comparison between the average individual and the particular individuals of the strain N2, modelled with the non-stationary and the stationary models. All boxplots were constructed by pooling individual-based estimates, obtained from ensembles of 1000 trajectories (total ensemble size N = 48000 per model). Similarly, the big blue and purple dots where obtained from the average individual (total ensemble size N = 1000 per model). From left to right, each row of panels show: MSD measured at the end of the regime, anomalous diffusion exponent with $\alpha = 1$ denoted as a dashed line, effective diffusion coefficient D_α, and duration of the regime. (A-D) Characterization of the fourth (long-term) regime. (E-H) Characterization of the third (midterm) regime. (I-L) Characterization of second (early transient) regime. (M-P) Characterization of the first (initial momentum) regime.

6.6 Results

Next, we explored the stationary and the non-stationary models for a comprehensive analysis of *C. elegans* spreading patterns at the level of strain. More specifically, we wanted to know whether and how *C. elegans* strains could generate superdiffusion during a relevant temporal scale. Following the macroscopic validation of the models (Section 6.5.3), we utilized strain-averaged parameters for this analysis. The values we used for each parameter can be consulted in Tables 6.1, 6.2 (shared parameters), Table 6.3 (stationary model) and Table 6.4 (non-stationary model).

6.6.1 Does *C. elegans* superdiffuse?

To know whether and how *C. elegans* superdiffuses, we made a Hurst exponent decomposition of the trajectories produced by strain-average models, considering both the stationary and the non-stationary formulations (see Section 6.2 for a short summary of this analysis or refer to Chapter 4 for a thorough description). Our findings (Figure 6.15, Table 6.5) indicated that the temporal unfolding of sharp turns and crawls (non-stationary model) is responsible for the non-stationary dynamics on increment displacement, and the main source of superdiffusion at the long-term limit (around hours). Suggesting the same mechanism could be operating in *C. elegans*. This finding is supported by consistent values of the Moses exponent larger than 0.5 ($M \approx 0.56$), and of the anomalous diffusion exponent larger than 1 ($\alpha \approx 1.15$) in all simulations involving the non-stationary model. Surprisingly, the strain VC125 did not exhibit as much superdiffusion as the other strains, despite showing $M = 0.57$ and $J = 0.52$. Further investigation is required to explore various hypotheses explaining this abnormal result but, notably, the VC125 strain displayed a less time-dependent unfolding of sharp turns and crawls, which may contribute to a faster convergence towards the diffusive regime compared to other strains (see k_c on Table 6.4).

Additionally, our results clearly indicated the stationary model was unable to generate superdiffusion, as it consistently exhibited $\alpha \approx 1$ and $M \approx L \approx J \approx 0.5$ during almost the entire simulation (Figure 6.15, Table 6.5). This lack of distinct signatures of anomalous diffusion in comparison to the non-stationary model, further supports the notion that the primary driver of superdiffusion in *C. elegans* is the Moses effect. Consequently, neither the double exponential distribution (representing the duration of crawls) nor the gamma distribution (representing the duration of sharp turns) could account for the emergence of superdiffusion in *C. elegans*. Furthermore, alternative distributions, including the truncated power law and the Weibull distribution, yielded comparable results provided all the rest of parameters were the same.

Apart from anomalous diffusion exponents, Table 6.5 and Figure 6.15D illustrate an important result. The final MSD (i.e., the overall spatial coverage in 90 minutes) of the non-stationary models was only marginally larger than that of the stationary models. Consequently, the anomalous diffusion exponent α seemed not to determine the size of the explored area, but rather reflected differences in the spatial sampling effort across time. To gain a better understanding of this interpretation, it is valuable to review the sequence of events occurring in the synthetic trajectories. Synthetic trajectories coming from the non-stationary model, which exhibited superdiffusion, initially spent a substantial amount of time moving near the origin of coordinates due to the early high frequency of sharp turns. This high frequency of sharp turn came from the combination of relatively long sharp turns with relatively short crawls. Over time, the number of sharp turns decreased (as they became shorter and crawls longer), leading to accelerated spreading. In contrast, the stationary model allocated its sampling effort differently. In this case, sharp turns and crawls occurred according to a stationary (no time dependence) random process. Hence, the duration of sharp turns and crawls, as well as their rates, where in average constant over time, keeping a regular sampling effort. As illustrated in Figure 6.15D, both the stationary and the non-stationary models converged towards the same spatial coverage (MSD), but the diffusivity they produced to reach this coverage were different. In the non-stationary model, remaining near the origin for a while caused an imbalance on the overall spreading that was then compensated by a relatively long superdiffusive regime, which, in the long run ($>$ 90 minutes), should converge to normal diffusion. However, in the stationary model, such an imbalance did not occur, as the tension between staying and leaving reached equilibrium much faster (see column Duration on Table 6.5).

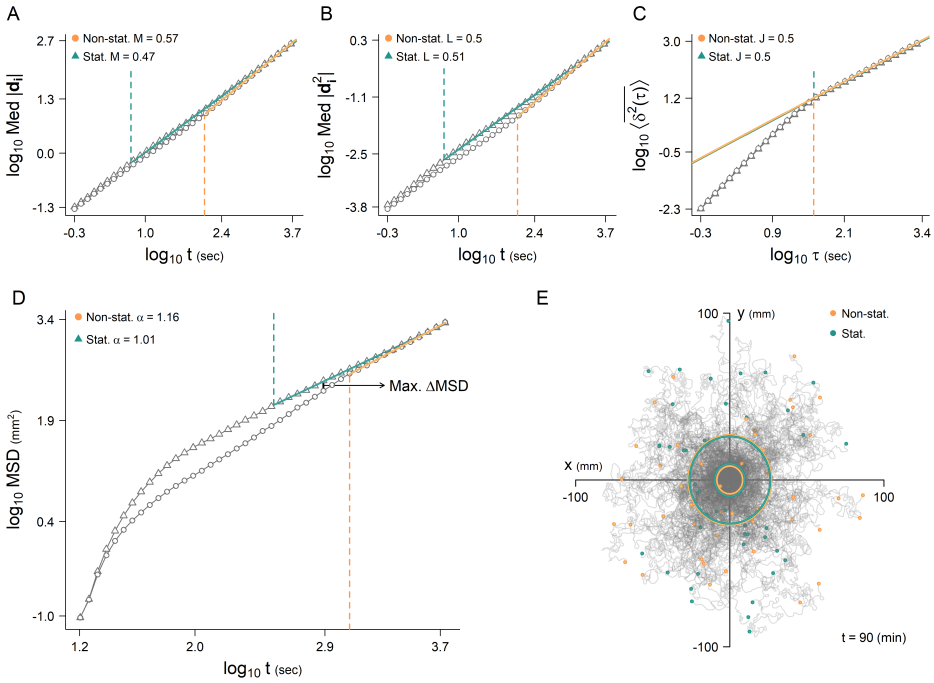


Figure 6.15: Mean square displacement (MSD) of the stationary and the non-stationary models of the average N2 individual. Input parameter values of the models can be consulted in Tables 6.1-6.4. (A-C) Decomposition of the long-term anomalous diffusion exponent α. (A) Moses exponent M (non-stationary dynamics) (B) Noah exponent L (extreme events), and (C) Joseph exponent J (long-range correlations). Estimates of M and L were obtained by fitting a piece-wise linear model with two regimes (on absolute increment displacements $|\mathbf{d}_i|$). For J three regimes were fitted on the ensemble time-averaged mean square displacement (TAMSD) $\langle\overline{\delta^2(\tau)}\rangle$, and J was computed on the third regime. (D) MSD curves of the models with the corresponding estimates of α, obtained by fitting a piece-wise linear model with four regimes. The time at which the two models exhibit the greatest difference in spatial coverage (in linear scale) is denoted with an arrow. (E) Random sample of 96 (i.e., 48×2) trajectories simulated with the stationary and the non-stationary models. Colored dots indicate the end of the trajectories. The small circles show the coverage area of the models at the time of maximum difference, also denoted in panel D. Big circles depict the final coverage area, which correspond to the MSD at the end of the simulation (90 minutes). To compute the radius of the circles we used $r(t) = \sqrt{\mathrm{MSD}(t)/\pi}$, which equals $\langle|x(t)|\rangle$ or $\langle|y(t)|\rangle$. In all panels spatial units are expressed in mm and dashed lines indicate the initiation of the long-term regimes, used to estimate the corresponding exponents.

Table 6.5: Mean square displacement (MSD) and anomalous diffusion exponent decomposition of the stationary and the non-stationary models. All estimates were obtained at the long term limit, (last observed regime), from ensembles of 1000 trajectories simulated using strain-average individuals, as reported in Tables 6.1-6.4. The final MSD, measured after 90 minutes, is expressed in mm^2. The anomalous diffusion exponent α, reported with the standard error, is dimensionless, like the Moses M, Noah L and Joseph J exponents. The effective diffusion coefficient D_α is expressed in $mm^2\,s^{-1}$. The duration of the diffusive regime is reported in hours, minutes and seconds, and can be subtracted from 90 minutes to determine the starting time of the regime.

Model	Strain	MSD	$\alpha \pm \sigma_\alpha$	M	L	J	D_α	Duration
Non Stat.	N2	2247	1.16 ± 0.018	0.57	0.50	0.50	0.027	01:10:20
	CB4856	4887	1.15 ± 0.025	0.55	0.51	0.51	0.064	01:05:34
	DA609	3645	1.16 ± 0.020	0.55	0.50	0.51	0.042	01:12:11
	VC125	2243	$1.06 \pm 0.14^*$	0.57	0.50	0.52	0.063	00:33:52
Stat.	N2	2156	1.01 ± 0.011	0.47	0.51	0.50	0.088	01:24:14
	CB4856	4040	0.96 ± 0.056	0.47	0.51	0.51	0.319	00:58:44
	DA609	3047	0.92 ± 0.030	0.46	0.51	0.51	0.286	01:09:53
	VC125	2153	1.05 ± 0.007	0.47	0.51	0.51	0.066	01:27:16

Note*: we interpret that this abnormally low value of α reflects a faster transition from superdiffusion towards diffusion in the VC125 strain. So fast, that the last regime detected, the one used to compute spreading properties, was already the diffusive regime.

6.6.2 Loops and sinuosity prevent superdiffusion

In order to get a more comprehensive understanding of why the stationary model did not exhibit superdiffusion, we examined the influence of smooth turning dynamics on the mean square displacement (MSD). In our model, these type of turns were encapsulated in the parameters β_c (curvature bias) and ζ_c (sinuosity) of the crawls. Thus, by adjusting their values the model could generate crawls spanning from straight-line trajectories to intricate looping patterns. Specifically, when ($\beta_c = 0$, $\zeta_c = 0$), crawls followed straight lines, while ($\beta_c \neq 0$, $\zeta_c = 0$) resulted in perfectly smooth loops. In turn, higher values of ζ_c increased sinuosity, while higher values of β_c increased the curvature, and consequently shrank the size of the loops.

Based on these facts, we focused on characterizing the effect of β_c and ζ_c, keeping all the rest of parameters fixed at the values of the average N2 individual (Tables 6.1-6.2, 6.3). Our hypothesis was that *C. elegans* exhibited too large values of β_c and ζ_c for it to generate superdiffusion with a simple stationary random process. To test that, we characterized the MSD curve of the stationary model for a range of β_c and ζ_c values (Figure 6.16 and Table 6.6). Like in other MSD

analyses, we simulated ensembles of 1000 synthetic trajectories to estimate the anomalous diffusion exponent α. However, here we executed 5 independent runs of the whole simulation and averaged the results to better characterize the variation of α in function of β_c and ζ_c.

Our results demonstrated that both the curvature force β_c and the sinuosity ζ_c reduced superdiffusion, having β_c a much stronger effect than ζ_c (Figure 6.16, Table 6.6). Given that *C. elegans* exhibited relatively large values of β_c and ζ_c, these results explained why the stationary model converged so fast towards the diffusive regime. Both Figure 6.16A and Table 6.6 show the maximum superdiffusion was produced when $\beta_c = 0$ and $\zeta_c \to 0$. In this limiting parameterization of the model two facts explained the emergence of a long transient superdiffusion. First, the model was as close as possible to the ballistic regime. Provided that sharp turns were kept intact (untouched), anything disturbing the straightness of crawls (i.e., $\beta_c \neq 0$, $\zeta_c \neq 0$) automatically shorted their length, and therefore pull the MSD curve towards the diffusive regime (see tracks depicted in Figure 6.16A). Second, the distinct scales of the double exponential and the gamma distribution governing the duration of crawls and sharp turns, lead to a multiscale walk. Furthermore, by applying the Hurst exponent decomposition we found that anomalous diffusion in this circumstance was driven by long-ranged correlations on increment displacements, or the so-called Joseph effect (Table 6.6).

Finally, to fully integrate our results with fundamental random search theory (Bartumeus et al., 2005; Bartumeus and Levin, 2008; Viswanathan et al., 2011; Méndez et al., 2014), we tried to approximate the paradigmatic Lévy walk with our stationary model. To produce the so-called Lévy-Langevin model (inset in Figure 6.16B) we implemented the following adjustments: (i) sharp turns were assumed to be instantaneous reorientations ($\vartheta = 0$), (ii) velocity was constant ($\xi_c = 0$), and (iii) crawl durations followed a truncated power law with $\mu = 2$, $a = 0.061$ and $b = 10^6$ seconds (instead of a double exponential). In this model, as the curvature and sinuosity approaches zero we find convergence to the Lévy walk (Bartumeus et al., 2005; Viswanathan et al., 2011), and consequently, superdiffusion becomes clearer.

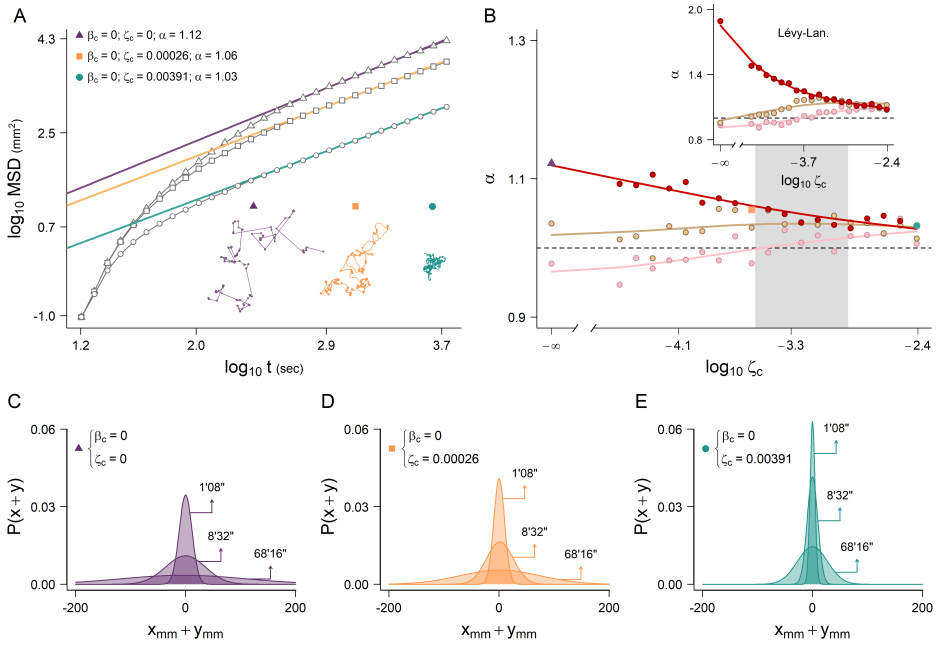


Figure 6.16: Smooth turning analysis of the stationary model. All parameter except β_c and ζ_c were fixed at the values of the average N2 individual, which can be consulted in Tables 6.1-6.3. To reduce spurious variability associated with random number generators, we ran this simulation for 10 different seeds (i.e., initialization values for the random number generators), and averaged the obtained parameter estimates. (A) Mean square displacement (MSD) for increasing values of ζ_c (sinuosity) and $\beta_c = 0$ (unbiased curvature), with representative examples of the corresponding 90 minutes tracks. (B) Anomalous diffusion exponent α as a function of ζ_c for unbiased-turning ($\beta_c = 0$) and biased-turning ($|\beta_c| > 0$) tracks. The Purple, orange and green dots indicate the anomalous diffusion exponent of the MSD curves shown in panel A, while the shaded region indicates the range of ζ_c values observed in *C. elegans* individuals under study, which importantly, all showed $|\beta_c| > 0$. The inset panel illustrates the results of the same analysis for the stationary model parameterized as a Lévy walk. That is, collapsing sharp turns to be instantaneous reorientations, and modelling crawl durations with a truncated power law (scaling exponent: $\mu = 2$, boundaries: $a = 0.061$, $b = 10^6$ seconds). (C-E) Propagators of the models in A evaluated at three equidistant times in logarithmic scale.

Table 6.6: Decomposition of the anomalous diffusion exponent of the stationary model for different β_c and ζ_c values. Other parameters were fixed at the values of the average N2 individual (see Tables 6.1-6.3). The Lévy-Langevin was simulated by collapsing sharp turns to be instantaneous reorientations, and modelling crawl durations as a truncated power law with scaling exponent $\mu = 2$, and boundaries:$a = 0.061$, $b = 10^4$ seconds (further details can be found in the main text). All estimates were obtained at the long-term limit from ensembles of 1000 trajectories. The final MSD, measured at 90 minutes, is expressed in mm^2. The anomalous diffusion exponent α, reported with the standard error, is dimensionless, like the Moses M, Noah L and Joseph J exponents.

Model	β_c	ζ_c	MSD	$\alpha \pm \sigma_\alpha$	M	L	J	Duration
Stat.	0.0000	0.00000	17107	1.12 ± 0.016	0.44	0.51	0.58	01:23:47
	0.0000	0.00026	6694	1.06 ± 0.013	0.45	0.51	0.54	01:24:45
	0.0000	0.00391	899	1.03 ± 0.008	0.49	0.51	0.52	01:24:44
	0.0025	0.00000	4563	1.04 ± 0.121	0.43	0.52	0.46	01:01:08
	0.0025	0.00026	4580	1.03 ± 0.014	0.45	0.51	0.53	01:24:41
	0.0025	0.00391	825	1.01 ± 0.010	0.49	0.51	0.51	01:24:58
	0.0100	0.00000	537	0.98 ± 0.053	0.45	0.52	0.44	01:16:58
	0.0100	0.00026	948	0.98 ± 0.008	0.46	0.51	0.48	01:26:15
	0.0100	0.00391	761	1.01 ± 0.014	0.50	0.51	0.51	01:21:49
Lévy Lan.	0.0000	0.00000	51048	1.89 ± 0.020	0.49	0.50	0.80	01:12:58
	0.0000	0.00026	4876	1.20 ± 0.023	0.49	0.50	0.55	01:17:03
	0.0000	0.00391	604	1.08 ± 0.016	0.49	0.50	0.51	01:16:15
	0.0025	0.00000	1810	0.96 ± 0.019	0.48	0.51	0.39	01:21:00
	0.0025	0.00026	2965	1.16 ± 0.016	0.48	0.51	0.52	01:21:30
	0.0025	0.00391	587	1.12 ± 0.066	0.49	0.50	0.52	01:07:51
	0.0100	0.00000	469	0.93 ± 0.065	0.50	0.50	0.45	01:15:04
	0.0100	0.00026	739	1.00 ± 0.017	0.50	0.50	0.49	01:19:2
	0.0100	0.00391	560	1.09 ± 0.009	0.49	0.50	0.51	01:23:45

6.6.3 Memory decay is the cause of superdiffusion

In previous sections, we learned that *C. elegans* individuals under study exhibited superdiffusion due to non-stationary dynamics of increment displacements, a phenomenon known as the Moses effect (Chen et al., 2017). These non-stationary increment displacements arose because *C. elegans* extended the duration of crawls over time while reducing the duration of sharp turns. A double modulation that effectively diminished the frequency of sharp turns over time, and prompted the transition from area restricted search to global search patterns (Hills et al., 2004;

Gray et al., 2005; López-Cruz et al., 2019). Notably, these dynamics, both in the increment displacements and the frequency of sharp turns, were encapsulated in the memory rates k_c and k_s of the non-stationary model, which therefore constituted the generative mechanism of anomalous diffusion in the model.

To gain deeper insights into how this mechanism operates, it is convenient to examine the impact of k_c and k_s on the frequency of sharp turns. When both k_c and k_s are set to 0, the duration of crawls and sharp turns remain constant over time, resulting in a constant (stationary) number of sharp turns. Negative k_c and k_s, signify an elongation of crawl durations and a shortening of sharp turn durations, both effects causing a frequency decay of sharp turns over time. With sharper declines corresponding to more negative values of k_c and/or k_s. Conversely, positive k_c and k_s indicate a shortening of crawl durations and an elongation of sharp turn durations, leading to an increased frequency of sharp turns over time. With greater values of k_c and/or k_s amplifying the increase in the frequency of sharp turns.

On average, *C. elegans* appears to autonomously regulate both k_c and k_s, displaying prolonged crawl durations (i.e., $k_c < 0$) and shortened sharp turn durations (i.e., $k_s < 0$) over time. These adjustments work synergistically to reduce the frequency of sharp turns over time. However, it is important to note that the average and variance of crawl durations were four times greater compared to sharp turns, indicating that most of the variability observed in this temporal modulation arose from crawling behavior.

Building upon these facts, we hypothesized the existence of optimal values for k_c and k_s that generate a non-stationary process producing the largest possible superdiffusive exponent α. Furthermore, we aimed to determine how closely single worms approach these optimal values of k_c and k_s when searching for food. To address this question, we characterized the mean square displacement (MSD) of the non-stationary model in function of the memory rates k_c and k_s (Figure 6.17), while fixing the rest of parameters at the values of the average N2 individual (see Tables 6.1-6.2, 6.4).

We constructed square matrices comprising $15 \times 15 = 225$ combinations of k_c, and k_s values, and for each combination we generated an ensemble of 1000 synthetic trajectories. Then, we computed the MSD curves, and estimated both α and D_α by fitting a log-log piece-wise linear model with four regimes, as detailed in Section 6.2. Aiming to reduce spurious variability associated with random number generators, we ran this simulation for 10 different seeds (i.e., initialization values for the random number generators), and averaged the obtained parameter estimates. Therefore, each combination of k_c and k_s end up containing an ensemble of 10^4 trajectories (i.e., 10×1000). Through this protocol we ensured the observed variability on (i)

the MSD, (ii) the anomalous diffusion exponent α, and (iii) the effective diffusion coefficient D_α, was function of k_c and k_s. Finally, we used a smoothing spline to increase the resolution of the heatmaps shown in Figures 6.17A-C. Raw versions of the very same heatmaps can be consulted in Appendix D.4.

Our results confirmed the presence of an optimal memory decay that maximized superdiffusion (Figure 6.17B). We observed a large variability among individuals and strains, but a substantial number of individuals across strains were at the optimum, or very close to it. Suggesting that superdiffusion plays a crucial role in the search strategy of *C. elegans*. When considering strains, the N2 and VC125 appeared to be closer to the optimal k_c and k_s values compared to the CB4856 and DA609, as indicated by the colored crosses in Figures 6.17A, C. Importantly, there were individuals from all four strains exhibiting an optimal memory decay (Figure 6.17B). Indeed, all individuals experienced the same preconditioning, which led them to expect finding food nearby, and had to adjust this expectation as time passed without encountering resources. The non-trivial result was that the transition from local to global search patterns occurred at an optimal pace. A pace that allowed them to maximize the value of the diffusive scaling exponent, and hence the spreading properties of their exploratory behavior. In practice, the worms optimally compensated the amount of time searching nearby and their capacity to explore at far-reaching distances.

As shown in Figures 6.17D, E, this phenomenon was clearly apparent by analyzing limiting cases. The trajectory associated with the fast memory decay (bottom left corner) represented a *C. elegans* individual losing its preconditioning experience very fast (i.e., the memory of nearby food), and therefore it departed from the origin of coordinates since the very beginning. The optimal memory decay represented an individual that spent an optimal amount of time ensuring there was no food left nearby, without losing the long-term (i.e., 90 minutes) spreading capacity (see the last observation of the MSD curves in Figure 6.17E). Finally, the no memory decay represented a *C. elegans* individual that assumed food was nearby even though it did not find it over the course of 90 minutes. Of note, the initial condition, defined by the values of b_s and $a_c + b_c$, was the same in all three cases (i.e., fast, optimal, and no-memory). However, in the no memory decay the initial condition was kept constant (stationary) over time given that $k_s = k_c = 0$.

We ran equivalent simulations as the ones shown in Figure 6.17, but fixing the average number of sharp turns occurred in 90 minutes, for all combinations of k_c and k_s (see Appendix D.4). The obtained results illustrated the fact that the existence of an optimal memory decay does not hinge on the number of sharp turns executed throughout the 90 minutes of the simulation. Rather, it is contingent on the pace of the transition from high-to-low frequency of sharp turns. The

reverse order (i.e., transitioning from low-to-high frequency of sharp turns) do not yield superdiffusion but rather subdiffusion, as illustrated in Figure 6.17B (see the positive domain of k_s and k_c, where the frequency of sharp turns increases over time).

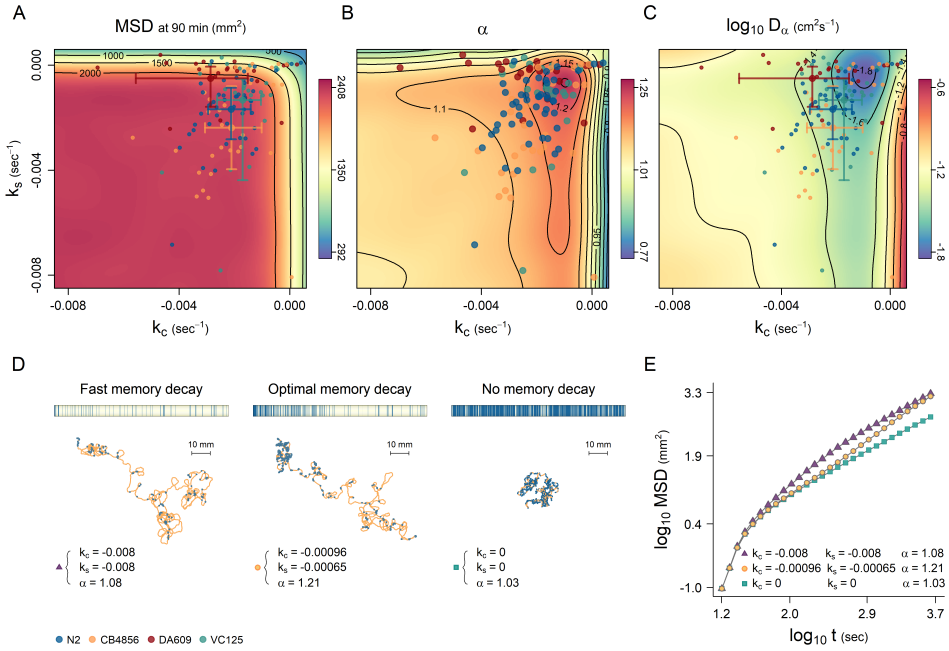


Figure 6.17: Mean square displacement (MSD) of the non-stationary model as a function of the memory rates k_c and k_s (other model parameters were fixed at the values of the average N2 individual, as reported in Tables 6.1-6.2, 6.4). (A) Overall spatial coverage in 90 minutes. Colored crosses represent the variability around average k_c and k_s of the strains under study: N2 (blue), CB4856 (orange), DA609 (red), and VC125 (green). (B) Anomalous diffusion exponent α with colored dots indicating the values of k_c and k_s for each *C. elegans* individual. (C) Effective diffusion coefficient D_α showing the same crosses depicted in A. (D) Examples of annotated trajectories with sharp turns (blue) and crawls (yellow) at relevant points of the parameter space. The ethograms above depict the unfolding of sharp turns and crawls through time. (E) MSD curves of k_c and k_s values of the tracks shown in panel D (ensemble size N = 1000). All heatmaps are smoothed representations of the raw simulations, which comprised $15 \times 15 = 225$ combinations of different k_c and k_s values. In its turn, each pair of k_c and k_s contains a whole MSD curve, equivalent to those depicted in panel E, computed with an ensemble of 1000 trajectories from which estimates of α and D_α were extracted. In all cases, both α and D_α were fitted at the long-term limit (last observed regime) using a log-log piece-wise linear model with four regimes.

6.7 Discussion

Random walks have long been recognized as powerful tools for understanding movement from first principles, ever since their introduction by Pearson (1905) and the subsequent development of the Langevin formalism (Langevin, 1908). These movement models have found wide application in both physical and biological sciences, facilitating the comprehension of diverse phenomena ranging from anomalous diffusion (Méndez et al., 2014) to the mechanistic understanding of animal search behavior (Salvador et al., 2014). Notably, among the various available models, the Langevin model has emerged as particularly influential in movement ecology (Uhlenbeck and Ornstein, 1930; Lenz et al., 2013; Gurarie et al., 2017; Méndez et al., 2014).

Drawing upon the extensive research on the Langevin formalism, we utilized classical Langevin equations to construct a comprehensive model of *C. elegans* search behavior. Our investigation underscored the pivotal role of non-stationary turning dynamics in driving anomalous diffusion in *C. elegans*. Specifically, we found that the accelerated spreading observed in both empirical and synthetic model trajectories, originated from time-dependent fluctuations in angular velocity, primarily influenced by the frequency decay of sharp turns over time. In our non-stationary model, these dynamics were encapsulated in the memory rates k_c and k_s, which govern the rate (or duration) of crawl and sharp turns over time (Figure 6.17). Gaining such thorough understanding of *C. elegans* spreading dynamics not only contributes to broadening the field of theoretical ecology but also provides valuable insights into the underlying biological mechanisms shaping animal search behavior.

Due to computational limitations, our study solely examines the relationship between mean square displacement (MSD) and memory rate parameters k_c and k_s for the average N2 individual. Future research should aim to generate the landscapes depicted in Figure 6.17 at the individual level, providing a more comprehensive understanding of whether *C. elegans* optimizes superdiffusion. Additionally, our findings underscore the importance of conducting experiments aimed at elucidating how the decaying frequency of sharp turns over time manifests at the biological level. Various mechanisms and physiological pathways, including metabolic perception, memory degradation, and learning, has been found to reinforce and modulate this non-stationary temporal trend (Hills et al., 2004; Gray et al., 2005; López-Cruz et al., 2019). Even, genetic and epigenetic inheritance also exert significant influences on shaping the observed behavioral patterns (Pradhan et al., 2019).

As first demonstrated by Hills et al. (2004), when *C. elegans* is relocated from a food source to an empty arena, it typically exhibits an area-restricted search

behavior lasting approximately 15-30 minutes before transitioning to extensive exploratory patterns. By explicitly incorporating these time-dependent dynamics into a Langevin modeling framework encompassing both crawls and sharp turns, we illustrated non-stationarity as the driving mechanism behind a prolonged transient superdiffusive regime. Remarkably, this transient superdiffusion persisted for over an hour, a noteworthy timescale considering the short lifespan of *C. elegans*, which reaches reproductive maturity in just three days (Byerly et al., 1976).

To gain further insights into the emergence of superdiffusion in *C. elegans*, we compared the mean square displacement curves (MSD) of the non-stationary model (exhibiting superdiffusion) with those of an equivalent stationary model (showing diffusive behavior). This comparison yielded valuable insights into the adaptive role of superdiffusion. As illustrated in Figures 6.15D and 6.17E, executing a superdiffusive spreading pattern may allow *C. elegans* to initially explore its local surroundings, thereby reducing the risk of overlooking nearby food resources, while still retaining the potential to cover significant distances later on.

We propose that, in our experimental setup, the initial local search reflects a memory effect resulting from the preconditioning experienced by the worms, which were fed *ad libitum* throughout their lives. Models replicating a memoryless spreading strategy, such as the stationary model or the non-stationary model with knocked-out k_c and k_s, spread out from the origin much earlier. They spent less time in proximity to the initial position and quickly converged towards normal diffusion.

Finally, exploration of the stationary model yielded insightful results regarding crawling behavior, including both curvature bias and trajectory sinuosity. Our findings revealed that these two phenomena prevent superdiffusion from occurring, prompting the question of their biological significance. Are they merely constraints, or do they possess adaptive value? Drawing from *C. elegans* literature, it appears plausible that they play an ecologically relevant role, given that *C. elegans* demonstrates the ability to modulate both the sinuosity and the curvature of its trajectory under appropriate external stimuli (Pierce-Shimomura et al., 1999, 2005; Peliti et al., 2013). Further investigation is required to understand whether the widespread presence of curvature bias in roundworms (observed across all 126 individuals) should be interpreted as a constraint or a selected trait.

Chapter 7

An ordered mind: *C. elegans* search behavior outperforms equivalent random but stationary movement patterns

7.1 Abstract

Search behavior is fundamental for the survival of organisms, driving the evolution of adaptive strategies influenced by biological constraints and selection pressures. While previous research has focused on exploitative-information mechanisms like chemotaxis and path integration, gaps remain in understanding how organisms navigate uncertainty and adapt their search behaviors. Here, we investigate the contribution of superdiffusive spreading in the search performance of *C. elegans*-inspired random walk models. Our results demonstrate that time-dependent superdiffusion and turning dynamics play pivotal roles in enhancing search performance in *C. elegans*, producing an effective balance of encounters with both nearby and distant targets. These findings were supported by both synthetic and empirical evidences, which in conjunction reveal the transition from local to large-scale search patterns in *C. elegans*, constitutes an optimal search strategy to effectively balance the exploitation-exploration tradeoff. Furthermore, our study suggests that *C. elegans* exhibits a rudimentary parental care indicating an adaptive response to varying environmental resource availability. This research contributes to a deeper understanding of the fundamental principles governing biological searches and their eco-evolutionary ramifications, offering insights into optimal search strategies across diverse organisms and environments.

7.2 Introduction

Search performance stands as a pivotal outcome of movement in living organisms, with profound implications across ecological scales—from individual fitness to population and ecosystem dynamics (Bell, 1991; Turchin, 1998). Efficient search strategies enable organisms to effectively locate and exploit resources, thus ultimately shaping their survival and reproductive success.

Evolution has endowed organisms with sensory and cognitive faculties to glean insights from environmental cues, encompassing resource availability and predator behavior, thus mitigating the impact of environmental unpredictability. However, the act of searching inherently embodies uncertainty, an unavoidable aspect owing to the finite nature of sensory and cognitive abilities. Organisms frequently find themselves navigating environments in search of vital resources (be it resources, mates, or shelter) armed with only partial and occasionally biased environmental information.

Despite the imperative of comprehending how organisms contend with such uncertainty, research within controlled laboratory settings, particularly focusing on cue-deprived environments, remains scant (Salvador et al., 2014; Sims et al., 2019). The existing body of literature predominantly accentuates exploitative-information mechanisms like chemotaxis (Saeki et al., 2001; Berg and Berry, 2004; Pierce-Shimomura et al., 2005), and path integration (Kim and Dickinson, 2017), inadvertently leaving a void in our understanding of how organisms navigate uncertainty and adapt their search behaviors in response.

From a theoretical standpoint, biologically inspired random walks offer significant insights into the mechanisms governing search performance, particularly in environments lacking sufficient information. Traditionally, random walk studies in the context of search efficiency have focused on metrics such as the mean first passage time (MFPT) (Viswanathan et al., 2011; Méndez et al., 2014). However, such analyses may inadvertently neglect critical facets of search behavior, including temporal constraints and the spatial arrangement of targets in relation to the searcher's position (Bartumeus et al., 2014; Chupeau et al., 2015; Bartumeus et al., 2016). In the context of diffusion, maximizing the MFPT involves maximizing the spatial coverage, that is, maximizing the mean square displacement (MSD), by reaching far the fastest possible. Nonetheless, in many contexts, optimal search is not only about maximizing the MSD but also about maximizing encounters with nearby targets. In other words, optimal searches aim at minimizing the drop of local target encounters as the MSD (spatial coverage) is increased. Superdiffusion is known to be one of the processes allowing to minimize this effect (Viswanathan et al., 2011; Bartumeus et al., 2014; Méndez et al., 2014).

Explicitly considering all the aforementioned factors is paramount for accurately evaluating search performance. Real-world search processes are intricately influenced by factors such as energy availability and the rate of energy consumption during movement. Moreover, acknowledging spatial asymmetries in target distribution and the potential for detection error or rewarding resource-patch revisits can significantly refine our comprehension of biological search efficiency.

For instance, integrating both proximal and distal targets into search performance evaluations has unveiled the effectiveness of stationary superdiffusive processes, such as Lévy walks with a scaling exponent of $\mu = 2$, in navigating intricate environments. A compelling and lively discourse exists in the literature regarding whether these patterns stem from optimized search processes or arise from cue-driven behavior in complex environments (Benhamou, 2007; Bartumeus and Levin, 2008; Sims et al., 2008; Kölzsch et al., 2015; Benhamou and Collet, 2015). Naturally, resolving this debate necessitates exploring model organisms under controlled conditions, wherein a search context (characterized by partial or scarce information) is imposed and meticulously controlled (e.g. (Kölzsch et al., 2015; Sims et al., 2019)).

In this study, we aim to address these gaps of knowledge by exploring the role of time constraints and spatial target distribution in *C. elegans* search behavior. By employing synthetic simulations and experimental observations in controlled environments, we seek to elucidate adaptive motor templates leading to an effective sampling of the environment, so-called optimal search strategies. Through our example, we aim to provide novel insights into the fundamental principles governing biological searches and its potential adaptive implications.

7.3 Materials and Methods

7.3.1 Search efficiency simulations

We conducted simulations to evaluate the efficiency of *C. elegans* search behavior employing the two movement models developed in Chapter 6. In all cases the so-called non-stationary and stationary models were parameterized using the average N2 individual, as reported in Tables 6.1-6.4, (Section 6.4.1, from Chapter 6), and search efficiency was evaluated as the time spent to find a target, either in terms of the mean first passage time (MFPT), or in terms of the N-passage time (NPT), depending on the context.

Movement models

The non-stationary movement model was a realistic representation of *C. elegans* search behavior, characterized by a frequency decay of sharp turns over time that produced long transient superdiffusion (i.e., $\alpha \approx 1.14$). In turn, the stationary movement model represented an equivalent search strategy but lacking this time-dependent dynamics. Instead, it executed sharp turns and crawls according to a random stationary process producing normal diffusion $\alpha \approx 1$. Therefore, it did not exhibit an ordered transition from intensive (local) to extensive (global) search patterns.

In addition to these two models, we also evaluated search performance of the so-called Lévy-Langevin model (see Section 6.6.2 in Chapter 6), and explored the role of both the curvature bias β_c and sinuosity ζ_c in each type of model.

Landscape geometries

We employed the circular geometry to simulate the symmetric search condition, where all targets are equidistant from the searcher position and directionality does not matter (Figure 7.1A, C). Conversely, we used the cardioid geometry to represent the asymmetric search condition, where both near and far targets are present, and directionality plays a crucial role (Figure 7.1B, D). Remarkably, at the initiation of the simulation (i.e., $t = 0$), the cardioid has an equal proportion of near-to-far targets, a non-trivial property in enclosed two-dimensional topologies. This geometric representation mirrors conditions frequently observed in search and foraging behaviors (Bell, 1991; Stephens et al., 2007). Conceptually, it represents an environment where an animal departs from a patch, expecting to encounter a high density of nearby targets separated by a finite distance from other patches located farther away.

Assessing search efficiency

Mean first passage time (MFPT)

We assessed search efficiency by computing the mean first passage time (MFPT) (Viswanathan et al., 2011; Méndez et al., 2014) of *C. elegans*-like movement models in both circular and cardioid geometries. Individual walkers were released at the origin of coordinates (x_0 in Figure 7.1) and tracked until reaching the edge of either the circle (symmetric search) or the cardioid (asymmetric search). For each realization, we recorded the first passage time and the spatial coordinates of the target encounter. This process was repeated 1000 times for each model type, and the search efficiency was computed as $\rho = \mathrm{MFPT}^{-1}$.

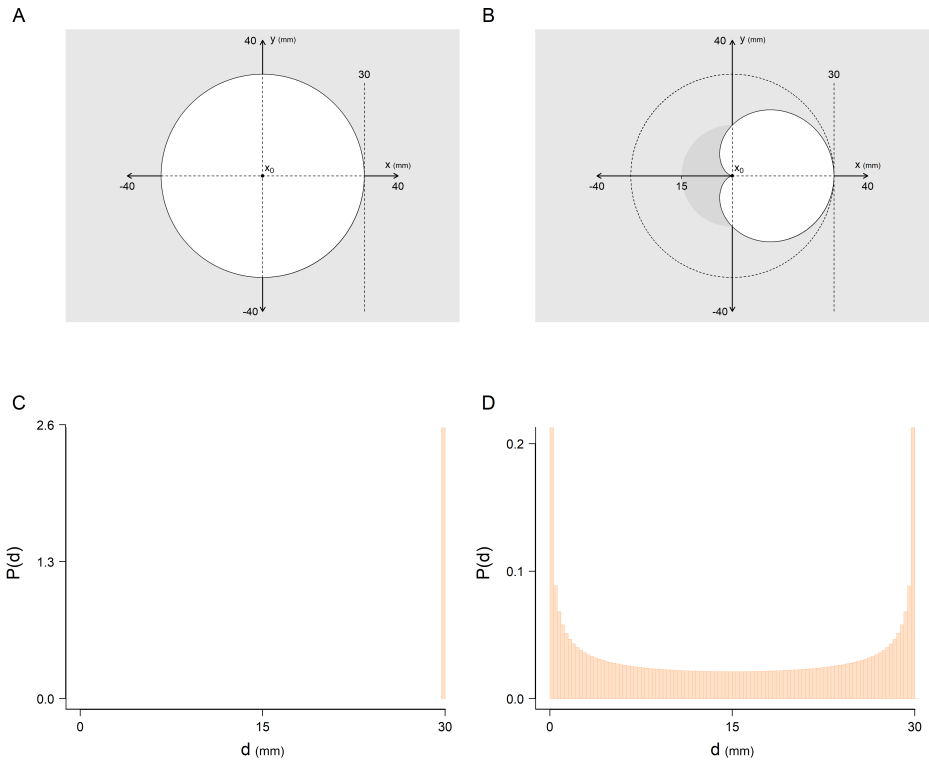


Figure 7.1: Schematic representation of the geometries used in the search efficiency simulations. In upper panels x_0 indicates the initial position of the searcher. (A) Circular geometry representing the symmetric search condition. (B) Cardioid geometry representing the asymmetric search condition. (C) Distribution of distances to target from the initial position in the circular geometry. (D) Distribution of distances to target from the initial position in the cardioid geometry. Note that the density (quantity) of near and far targets is exactly the same.

To streamline the simulation we leveraged the fact that the cardioid geometry fits entirely within a circle, provided the two geometries share the same radius. This implies that before reaching the edge of the circle, a walker must necessarily cross the cardioid. Consequently, we could simultaneously obtain the MFPT for both geometries in a single simulation, stopping the walk the first time it crossed the circle.

Drawing from Bartumeus et al. (2014) we factorized search efficiency in the asymmetric search condition (cardioid) in three complementary quantities: (i) the overall MFPT, (ii) the MFPT_{near} associated with near encounters, and (iii) the MFPT_{far} associated with far encounters. So that, $\text{MFPT} = \text{MFPT}_{near} + \text{MFPT}_{far}$. Since the cardioid geometry exhibits the same number of near-to-far tar-

gets (Figure 7.1), we defined as near targets all positions satisfying $\sqrt{x^2 + y^2} \leq r/2$, where x and y are the cardioid positions, and r is the radius of the cardioid (see Figure 7.1D). Similarly, far targets were defined as all positions satisfying $\sqrt{x^2 + y^2} \geq r/2$. Finally, we computed the probability of finding near or far targets as $P(near) = \langle N_{near} \rangle / \langle N_{total} \rangle$ and $P(far) = \langle N_{far} \rangle / \langle N_{total} \rangle$, where the different $\langle N_x \rangle$ represent the average number of hits near, far and in total.

N-passage time simulations

We conducted N-passage time simulations using the same approach, but fixing the travel time to 90 minutes. This duration was selected for two main reasons. Firstly, it corresponded to the duration of the experimental data used to fit the two movement models (the stationary and the non-stationary one). The non-stationary model was able to reproduce the search behavior and dispersal patterns of *C. elegans* for up to 90 minutes, beyond which its effectiveness was unknown. Secondly, we aimed to directly compare search efficiency and the mean square displacement of the non-stationary and the stationary models. Specifically, we sought to determine whether the observed superdiffusion in *C. elegans* (i.e., $\alpha \approx 1.14$) had adaptive and ecologically relevant implications. Remarkably, within the 90-minute timeframe, the two movement models (the stationary and the non-stationary) covered a comparable spatial area but exhibited different diffusive regimes (i.e., $\alpha \approx 1$ *vs* $\alpha \approx 1.16$, respectively).

7.3.2 From synthetic simulations to real experiments

We conducted a series of foraging experiments with *C. elegans* and *Escherichia coli* OP50 to complement the results of the synthetic search efficiency simulations with real data. These experiments involved placing individual worms of the strain N2 into large foraging plates (24.5 × 24.5 cm) containing 55 patches of bacteria that were spatially distributed to mirror either the symmetric or asymmetric search condition. The symmetric search condition (circle in the synthetic simulations) was approximated using the so-called homogeneous landscape (Figure 7.2A), while the asymmetric search condition (cardioid in the synthetic simulations) was approximated using the so-called heterogeneous landscape (Figure 7.2B). These landscapes were virtually generated using custom scripts in R and C++ (see Box 3) and subsequently reproduced in an agar plate using an automatic micro-pipetting robot, as explained below.

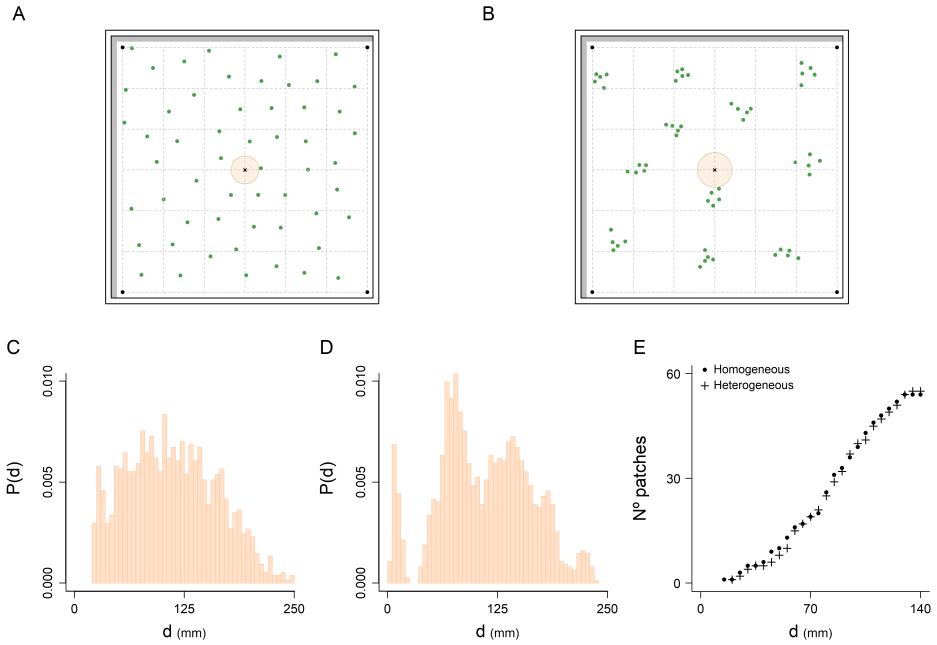


Figure 7.2: Schematic representation of the landscapes utilized in the foraging experiments with *C. elegans* and *E. coli* OP50. (A) Homogeneous landscape representing the symmetric search condition, where in average all patches of bacteria are equidistant from each other, rendering directionality inconsequential. (B) Heterogeneous landscape simulating the asymmetric search condition, where directionality matters due to the presence of both nearby and distant patches from the perspective of the searcher. In both cases, the initial position of the experiment is denoted with an x, and the circle that surrounds it shows the area devoid of patches, typically defined in terms of approximation time (Bartumeus et al., 2016). The black dots on the corners of the landscapes represent reference points to place the patches in its relative location, and the gray shaded region show the offset that the automatic micro-pipetting robot required to move. (C) Pooled histogram displaying distances between all patches in the homogeneous landscape. (D) Pooled histogram showing distances between all patches in the heterogeneous landscape. (E) Radial profile detailing the number of patches in the landscapes depicted in panels A and B. Note that the area devoid of patches is slightly smaller in the homogeneous landscape compared to the heterogeneous landscape.

Box 3: Generating the foraging landscapes

In our code, the positions of food patches were generated using polar coordinates, i.e., ($x = r\cos\phi$, $y = r\sin\phi$), where r represents the radial component of the polar coordinates (i.e., the distance to the origin), ϕ is the absolute angle, and x, y are the coordinates of the patches expressed in Cartesian space. Thus, to determine the position of a food patch, we needed to randomly select values for both the angle ϕ and the radial component r. In the homogeneous landscape, values of r were sampled from a gamma distribution with shape $\alpha = 1.2$ and scale $\theta = 50$ mm, while in the heterogeneous landscape, they were sampled from a uniform distribution with a lower limit $a = 20$ mm and upper limit $b = 148.77$ mm. The choice of these particular distributions was based on visual appearance of the landscapes. In practice, we aimed to fit 55 patches into a square space of 24.5 × 24.5 cm, thus, alternative distributions could potentially yield similar results.

Homogeneous landscape

The homogeneous landscape was generated assuming each patch had a self-centered circular area of exclusion of size $A = \pi r_i^2$, in which no other patches could grow. Therefore, the minimum distance between patches was $d = 2r_i$. In our case we set $A = 1255.794$ mm^2 and $d = 9.99$ mm. The process went as follows:

1) A first patch grew in a random location of the landscape.
2) A random distance to the origin r was sampled assuming a gamma distribution with shape $\alpha = 1.2$ and scale $\theta = 50$ mm.
3) A random direction ϕ (angle) was sampled from the uniform distribution, bounded from $-\pi$ to π.
4) In case the position ($x = r\cos\phi$, $y = \sin\phi$) was not occupied by a previously grown patch we let the current patch grown.
5) The steps 2-4 were repeated until completing $K = 1000$ trials.

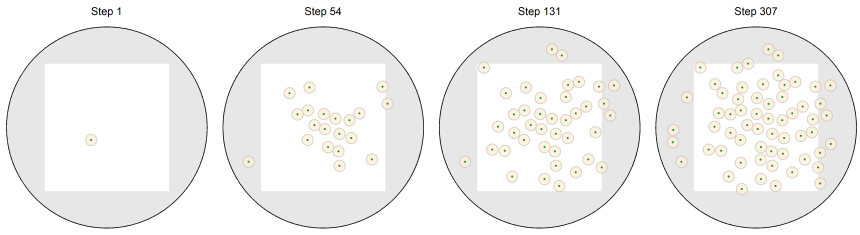


Figure 7.3. Colonization process of homogeneous landscapes. From left to right: steps 1, 54, 131 and 307. Green circles represent food patches and light brown circles represent the area of exclusion of each patch. The white region represents the area of the foraging plates used to construct the actual experimental landscapes, which were squares of 24.5 × 24.5 cm.

Heterogeneous Landscape

In the heterogeneous landscapes we assumed each patch had a self-centered circular area of exclusion of size $A = \pi r_i^2$, in which foreign patches could not grow. When a founder patch grew, it propagated $n = 4$ additional patches around itself given a normal distribution with mean $\mu = 5$ mm and standard deviation $\sigma = 1.67$ mm (defining the radial component of the coordinates centered at the founder patch), and a uniform distribution bounded from $-\pi$ to π. These n patches formed a cluster with variable area of exclusion, depending on how close or sparse they grew. Overall, the process to generate heterogeneous landscapes went as follows:

1) A first patch grew in a random location of the landscape.
2) Right after, $n = 4$ more patches grew around it assuming a normal distribution centered at the founder patch (determining the radial component of the coordinates), and a uniform distribution determining their direction.
3) A random distance to the origin r was sampled from a uniform distribution with bounds $[a = 20, b = 148.77]$ mm.
4) A random direction ϕ (angle) was sampled from the uniform distribution, bounded from $-\pi$ to π.
5) In case the position ($x = r\cos\phi$, $y = \sin\phi$) was not occupied by a previously grown cluster, we let a new founder patch to grow.
6) The steps 2-5 were repeated until completing $K = 1000$ trials.

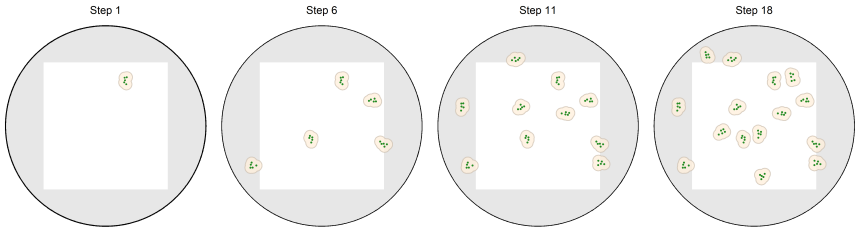


Figure 7.4. Colonization process of heterogeneous landscapes. From left to right: steps 1, 6, 11 and 18. Green circles represent food patches and light brown circles represent the area of exclusion of each cluster. The white region represents the area of the foraging plates used to construct the actual experimental landscapes, which were squares of 24.5 × 24.5 cm.

Once the landscapes were generated, we optimized the similarity of their radial profile of patches (i.e., number of patches found as a function of the distance to the origin) to ensure comparability between them, specially in relation to the initial condition of the search, defined by the distance to the nearest target (see Figure 7.2E). This optimization process involved applying a linear transformation of the landscapes to determine the positioning that minimized the sum of squares in the radial profile of patches.

7.3.3 Landscape preparation

Large plates (24.5 × 24.5 cm, Square BioAssay Dish, Corning Inc.) were prepared one week before the experiments with 230.4 mL of foraging medium (FM: 3 g/L NaCl, 20 g/L agar, 25 mL/L potassium phosphate buffer pH 6, 1 mL/L (1M) $MgSO_4$, 5 mg/L cholesterol, 1 mL/L (1M) $CaCl_2$, 10 mg/L chloramphenicol). This medium was specifically designed by Madirolas et al. (2023) to prevent bacterial growth due to the bacteriostatic effect of the chloramphenicol, while maintaining the health of the bacteria. Notably, we omitted novobiocin from the solid medium whereas it was included in the original recipe.

After pouring, plates were dried under the hood for 2-3 hours, closed, and stored at room temperature (21 °C). The following day, we reversed them down to prevent moisture accumulation on the surface of the agar. One day before the experiment (i.e., 24 hours), the foraging plates were seeded with *E. coli* OP50 using an automatic micro-pipetting robot (Opentrons OT-2) as shown in Figure 7.5. The robot distributed 55 patches of bacteria according to homogeneous and heterogeneous landscape configurations shown in Figures 7.2A, B. Bacterial patches consisted of drops of 0.75 µL of *E. coli* OP50 suspended in foraging buffer at OD = 0.2 (FB: 3 g/L NaCl, 25 mL/L potassium phosphate buffer pH 6, 1 mL/L (1M) $MgSO_4$, 1 mL/L (1M) $CaCl_2$, 10 mg/L chloramphenicol, 100 mg/L novobiocin). Of note, foraging buffer also prevented bacterial growth due to the bacteriostatic effect of antibiotics (i.e., chloramphenicol and novobiocin), in addition to the chloramphenicol found in the solid foraging medium. In all cases, plates were visually inspected to contain 55 regular circular patches after running the robot.

7.3.4 Bacterial cultures and preparation

Cultures

E. coli OP50 was recovered from long-storage glycerol stock (−80 °C), streaked on nematode growth medium plates (NGM: 3 g/L NaCl, 2.5 g/L peptone, 20 g/L agar, 25 mL/L potassium phosphate buffer pH 6, 1 mL/L (1M) $MgSO_4$, 5 mg/L cholesterol, 1 mL/L (1M) $CaCl_2$), and stored at 4 °C. Liquid cultures of bacteria too seed foraging plates were prepared in two steps. First we grew a pre-culture by inoculating one bacterial colony from the NGM plate in 5 mL of LB medium (Luria Broth) and incubating for 24 hours at 22 °C (orbital shaking at 300 rpm, in a closed 50 mL Falcon tube). Second, we prepared a culture by inoculating 1 µL of the pre-culture in a 10 mL fresh LB, and incubating for another 24 hours in the exact same conditions.

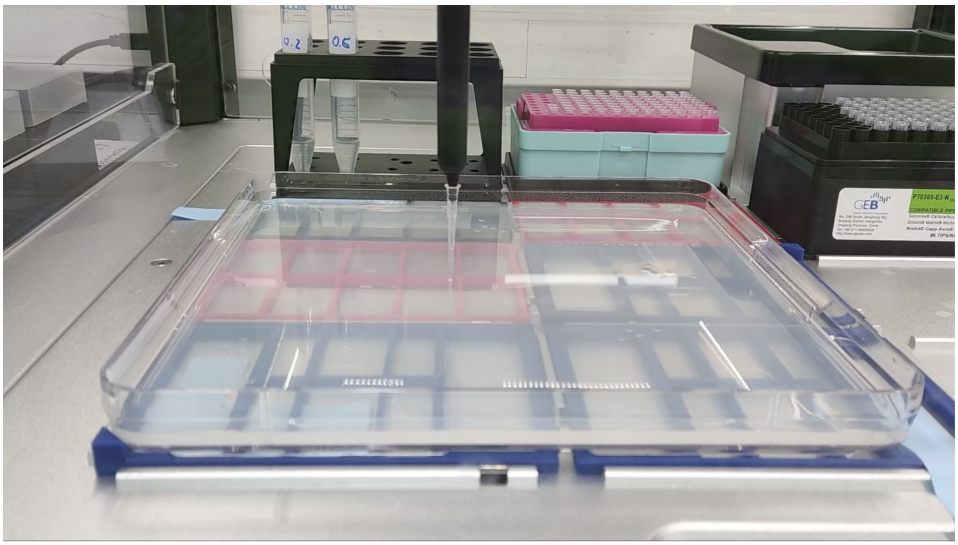


Figure 7.5: Foraging landscape preparation using an automatic micro-pipetting robot. The robot disposed drops of bacteria at specified spatial locations according to treatment condition (e.g., bacteria concentration, species, spatial configuration). In this case, drops of 0.75 µL were disposed following the heterogeneous landscape configuration using the left most falcon (OD = 0.2), which contained *E. coli* OP50 suspended in foraging buffer. The photo was taken at the Center for Integrative Biology (Toulouse), where the experiments were conducted.

Bacterial density

Right after culturing the bacteria, optical density (OD) was measured at 600 nm wavelength using a spectrophotometer (Jenway 7200, Cole-Parmer, Staffordshire, UK). This particular spectrophotometer was most accurate for ODs between 0.1 and 1, thus we diluted the bacterial culture in order to obtain OD measurements within the range of values (pre-wash measurements in Table 7.1). Then we washed the bacteria three times with foraging buffer. To do so, we transferred the culture into a 5 mL Eppendorf and centrifuged it at full speed (5500 rpm) during five minutes using a top-table centrifuge (Benchmark, My Fuge 5). We removed the supernatant, re-filled the Eppendorf with fresh foraging buffer and gently homogenized the content using a micropipette of 1000 µL. This process was repeated until completing three washes. Afterwards we measured the OD of the washed culture (first post-wash measurement of each day in Table 7.1) and performed serial dilutions to bring the culture at OD = 0.2, which was the targeted OD for the foraging experiments.

Additionally, we performed serial dilutions, of the washed culture, in M9 buffer (M9: 5 g/L NaCl, 3 g/L KH_2PO_4, 7.52 g/L $Na_2HPO_4 \cdot 2H_2O$, 1 mL (1M) $MgSO_4$) to estimate the number of colony forming units (CFU) per µL, at OD = 1. These

serial dilutions were carried in 96-well plates, from which drops of 10 µL of diluted bacteria were seeded in a fresh NGM plate (Figure 7.6A). After 1-3 days incubating the colonies at room temperature (21 °C), the number of CFU per µL at OD = 1 were computed to keep track of the amount of bacteria we disposed in each patch in the foraging landscapes (Figure 7.6B and Table 7.2).

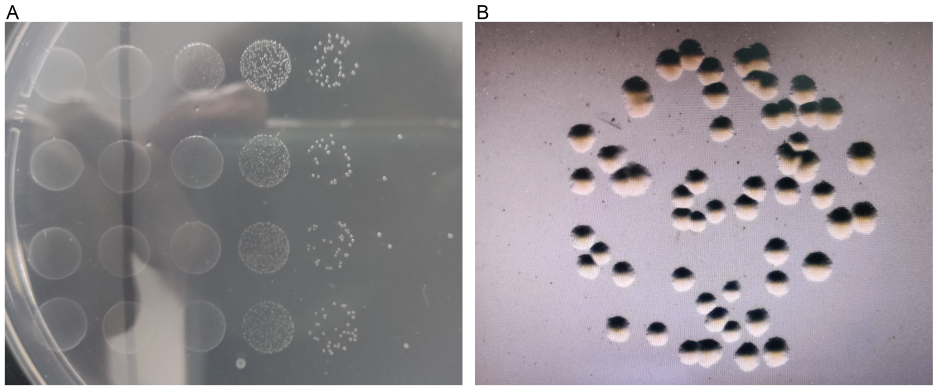


Figure 7.6: Illustrative example of *E. coli* OP50 colonies grown in NGM plates. (A) Serial dilutions of a washed culture of *E. coli* OP50 seeded in an NGM plate. (B) Zoom-in of the colonies grown in the fifth drop (i.e., $C_i = 10^{-6}$) of a serial dilution equivalent to the one depicted in panel A. These colonies were manually counted and used to estimate the number of colony forming units (CFU) per µL at OD = 1.

7.3.5 Strains maintenance and manipulation

The experimental line of *C. elegans* (strain N2) was derived from long-term storage dauer larvae, maintained at 15 °C without food. Prior to the experiments, we allowed the worms to undergo 5 generations of non-dauer life cycle recovery, totaling approximately 15 days, and ensuring worms never exceeded 30 generations from the foundational population obtained from the *Caenorhabditis* Genetics Center, University of Minnesota (https://cgc.umn.edu/). Throughout the experimental period, worms were cultured on NGM, feeding on *E. coli* OP50 lawns, and maintained at 21 °C ± 0.1. Line transfers were carried at regular intervals of 2-3 days to prevent food depletion and starvation.

Assayed individuals were in all cases young adults, aged 48 ± 2 hours, that were synchronized from the main line through bleaching and egg collection. The bleaching procedure (i.e., synchronization) was performed 72 hours prior to the experiment, and rescued eggs were shaken in M9 buffer at room temperature (21 °C) for 20 ± 2 hours, until a significant portion of L1 larvae had hatched. These synchronized L1 larvae were then seeded onto NGM plates with *E. coli* OP50 lawns and allowed to grow for 48 ± 1 hours.

Table 7.1: Record of *E. coli* OP50 cultures utilized for seeding experimental foraging plates. The reported day (Exp. Day) denotes the day on which the actual experiment was conducted, thus, OD measurements were taken the day before, cultures were prepared two days before, and pre-cultures were prepared three days before. Serial dilutions and CFU counting were performed using the first post-wash culture of each day. The last post-wash culture, targeted at OD = 0.2, was employed to seed the experimental foraging plates.

Exp. Day	Status	Dilution	Measured OD	Real OD	Target OD
22-11-2022	Pre-wash	×10	0.274	2.740	-
	Post-wash	×10	0.258	2.580	-
	Post-wash	×10	0.121	1.210	1.25
	Post-wash	×1	0.495	0.495	0.50
	Post-wash	×1	0.205	0.205	0.20
23-11-2022	Pre-wash	×10	0.330	3.300	-
	Post-wash	×20	0.301	6.020	-
	Post-wash	×10	0.121	1.210	1.25
	Post-wash	×1	0.495	0.495	0.50
	Post-wash	×1	0.205	0.205	0.20
26-11-2022	Pre-wash	×10	0.343	3.430	-
	Post-wash	×10	0.247	2.470	-
	Post-wash	×10	0.128	1.280	1.25
	Post-wash	×1	0.494	0.494	0.50
	Post-wash	×1	0.198	0.198	0.20

Table 7.2: Record of *E. coli* OP50 colonies grown on NGM plates to track the amount of bacteria seeded per patch in the foraging experiments. The reported day (Exp. Day) indicates the day on which the actual experiment was conducted, with CFU assessed 2-3 days later. The column Drop indicates the drop used to count the colonies (see Figure 7.6A), C_i reports the concentration of the culture in the drop, Counts and Average denote the number of colonies found, OD indicates the optical density measured at a wavelength of 600 nm, and CFU reports the estimated number of colony forming units per µL at OD = 1. This CFU value was obtained by dividing the Average by the OD, and scaling the result to the appropriate concentration.

Exp. Day	Drop	C_i	Counts	Average	OD	CFU
22-11-2022	6	10^{-7}	2, 6, 1, 0	2.25	2.58	$8.7 \cdot 10^5$
23-11-2022	6	10^{-7}	6, 9, 9, 5	7.25	6.02	$1.2 \cdot 10^6$
26-11-2022	5	10^{-6}	17, 24, 25, 33	22.3	2.47	$9.0 \cdot 10^5$

7.3.6 Behavioral assay and data collection

All experiments were carried with well-fed young adult hermaphrodites of 48 ± 1 hours adulthood, previously synchronized as described above. Before placing the worms in the experimental arenas, we washed them with M9 buffer, utilizing a micropipette of 1000 µL. Subsequently, we collected the worms using a triton X-100 coated tip and transferred them into a 1.8 mL Eppendorf tube. The worms were then centrifuged for approximately 5 seconds (Thermo Fisher Scientific, mySPIN 6), after which the supernatant was removed, and refilled with fresh M9 buffer. This process was repeated for a total of 6 washes. Then we added 2 mL of M9 buffer + 0.1% triton X-100 into a small Petri plate of 6 cm in diameter, and transferred the worms from the Eppendorf to the plate using a triton X-100 coated tip, where they remained for 2 hours $\pm$ 5 min.

This period without food served two main purposes. First, to ensure that the memory of nearby food, experienced by the worms during their lifespan was eradicated, or at least substantially reduced. Second, to ensure the worms were motivated to actively search for food during the foraging experiment. Right after completing two hours without food, we picked individual worms swimming in the Petri plate with M9 buffer + 0.1% employing a micropipette of 2 µL, and transferred them at the center of the foraging plates were they remained exploring and foraging bacteria for a total of 9 hours. This transfer manually executed with the help of the template shown in Figure 7.7A. After 9 hours, we putted the foraging plates 15 ± 1 minutes on the freeze to sleep the worms, and subsequently put them on the fridge until data collection.

Assayed plates were examined under a stereo microscope (Leica S9D) and manually scanned in a systematic zigzag pattern from the bottom left corner to the top right corner. The number of patches found (Figure 7.7C) and the number of eggs within each patch, or outside them, were recorded using standardized experimental sheets. Importantly, we found all eggs in clear association with a patch encounter, or a cluster of patches and were never away from them. Special care was taken to thoroughly inspect all patches on the plates and all space between them.

7.4 Results

In the following sections we delved into the analysis of both stationary and non-stationary models to provide a thorough examination of search performance in *C. elegans*. Our primary objective was twofold: first, to ascertain whether the phenomenon of superdiffusion, observed in both empirical data (Chapter 4) and non-stationary model trajectories (Chapter 6), yielded a positive impact on search

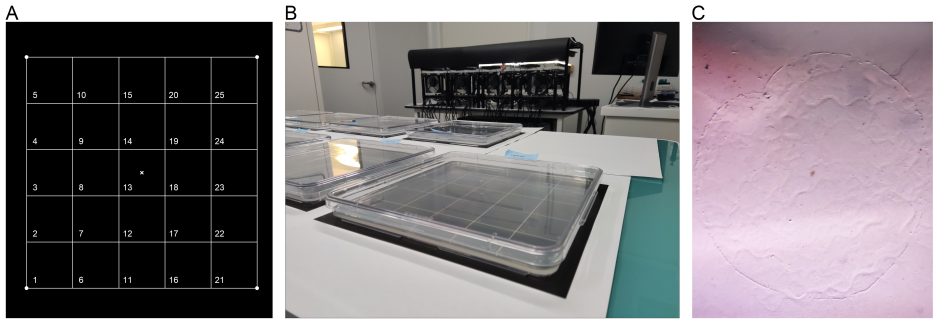


Figure 7.7: Foraging experiments with *C. elegans* and *E. coli* OP50 in virtually unbounded landscapes. (A) Template employed to place individual worms at the center of the foraging landscapes. (B) Picture of the laboratory during the course of the behavioral assay, conducted at the Center for Integrative Biology (Toulouse). (C) Illustrative example of a consumed patch examined under the stereo microscope during data collection.

performance; and second, to elucidate the adaptive significance of curvature bias and sinuosity, two salient features of *C. elegans* trajectories.

In all cases, search performance was assessed in both circular and cardioid geometries. Since a circle have all "targets" at the same distance, a unique overall search efficiency was computed. For the cardioid, instead, we differentiated encounters with near and far targets (see Section 7.3.1). Meaning that we factorized the search efficiency into three elements: (i) near encounters, (ii) far encounters, and (iii) total encounters, which equals the sum of near and far encounters. Lastly, we presented the results of the foraging experiments inspired by synthetic search efficiency simulations, as a real-world case study where the ecological implications of search behavior were explicitly assessed.

7.4.1 First Passage Times: a too simple measure of search efficiency?

We conducted search efficiency simulations to assess the mean first passage time (MFPT) on both the circle and the cardioid geometries using the stationary and the non-stationary models developed in Chapter 6 (briefly described in Section 7.3.1). These simulations were scaled to match the radius where the mean square displacement curves of the two models exhibit maximal difference, as illustrated in Figure 6.15 from Chapter 6. As anticipated, the MFPT of the stationary model was notably lower than that of the non-stationary model. This discrepancy arises from the fact that the stationary model tends to depart more rapidly from the origin of coordinates, compared to the non-stationary model.

Consequently, MFPT simulations appear inadequate for capturing movement properties other than those promoting the tendency to reach distant locations. Specifically, crucial tradeoffs in a search process such as balancing spatial coverage with detection errors, or the presence of time constraints to put some relevant examples, seem to be overlooked in search performance assessments based on MFPT metrics.

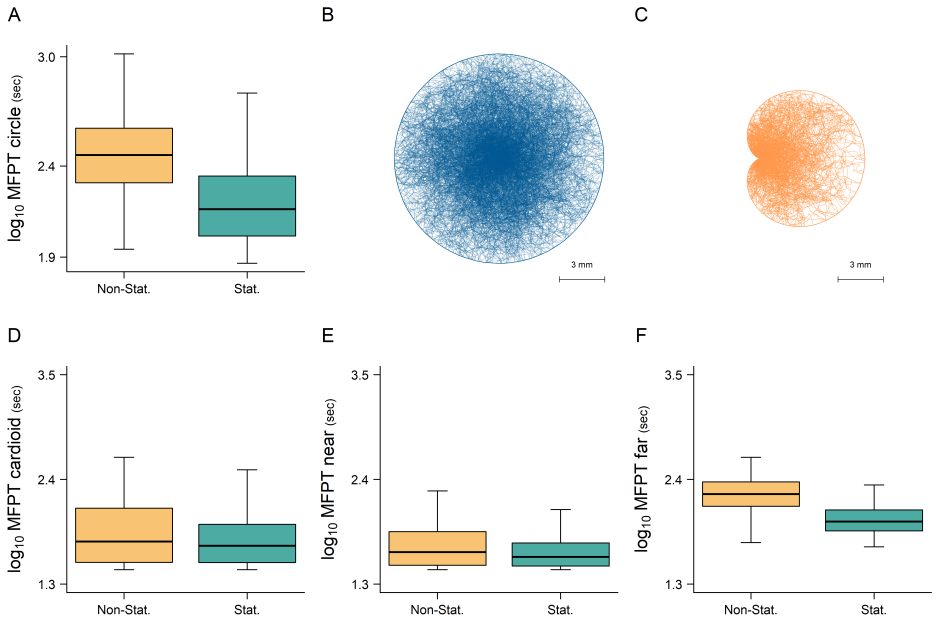


Figure 7.8: Mean first passage time (MFPT) simulations in circular and cardioid geometries for the non-stationary and stationary models. (A) MFPT on the circle. (B, C) Schematic representation of the simulations in the circle and the cardioid geometries, respectively. In both cases depicted tracks were generated with the non-stationary model. (D) Overall MFPT on the cardioid. (E, F) MFPT associated to near and far encounters in the cardioid, respectively. In all cases simulations were scaled to match the radius of maximal difference between the mean square displacement curves of the two models (stationary and non-stationary), which correspond to $r = 6.93$ mm.

7.4.2 The role of superdiffusion in *C. elegans* search behavior

In this section, we present the outcomes of N-passage time simulations conducted with truncated duration (90 minutes) in circular and cardioid geometries (Figure 7.9 and Tables 7.3, 7.4). The search efficiency of various *C. elegans*-inspired movement models was evaluated by computing the average number of hits (encounters) within these geometries.

This analysis allowed us to quantify the role of superdiffusion in *C. elegans* search behavior across a wide range of spatial scales (Figure 7.9). Notably, comparing the stationary model (depicted by the green dotted line) with the non-stationary model (orange dotted line) yielded insightful findings (Figure 7.9). The difference between these two models is a direct measure of the net increase in targets found when increasing the anomalous diffusion exponent α from 1.01 (stationary model) to 1.16 (non-stationary model). Additionally, this comparison can be extended by analyzing the optimal non-stationary model ($\alpha = 1.21$), illustrated by the yellow polygon in Figure 7.9.

Our findings indicate that time-dependent superdiffusion enhanced search performance by producing a faster spreading (larger scaling exponent) while increasing the number of nearby encounters (Figure 7.9 and Table 7.3). This trend is particularly pronounced when comparing the stationary (green) and the non-stationary (orange) models, as the overall increase in targets primarily stems from near encounters (Figure 7.9D-F).

The stationary Lévy-Langevin walk with sinuosity (red) and the non-stationary *C. elegans* walk (orange) (see Figure 7.9) yield similar MSD and scaling exponents, and therefore are comparable in diffusive properties (Table 7.3). In this case, the non-stationary *C. elegans* walk performs only slightly better and most of the difference comes from the fact that the Lévy-Langevin walk with sinuosity can not reach far targets as effectively as the non-stationary model (Figure 7.9). So, in this example, where there is not a significant change in the scaling exponent (i.e., $\alpha = 1.16$ and $\alpha = 1.17$) non-stationary dynamics drives an increase in far-reaching targets.

While our results suggest that enhanced search performance is a general consequence of superdiffusive spreading, further investigation is warranted to ascertain whether all types of superdiffusion are equally advantageous, or if temporally structured sampling (as discussed in the Moses effect in Chapter 4) is more beneficial. Specifically, we wonder if a non-stationary process leading to a sampling strategy that starts locally, and expands non-linearly in time towards larger scales performs equally well as an equivalent but random (or time disordered) sampling strategy. Ideally, such assessment should involve comparing stationary and non-stationary diffusive and superdiffusive models that yield identical mean square displacements within a designated timeframe, a fundamental condition that was not met in our case when comparing the two Lévy-Langevin models or the two stationary models (see Figure 7.9 and Tables 7.3, 7.4).

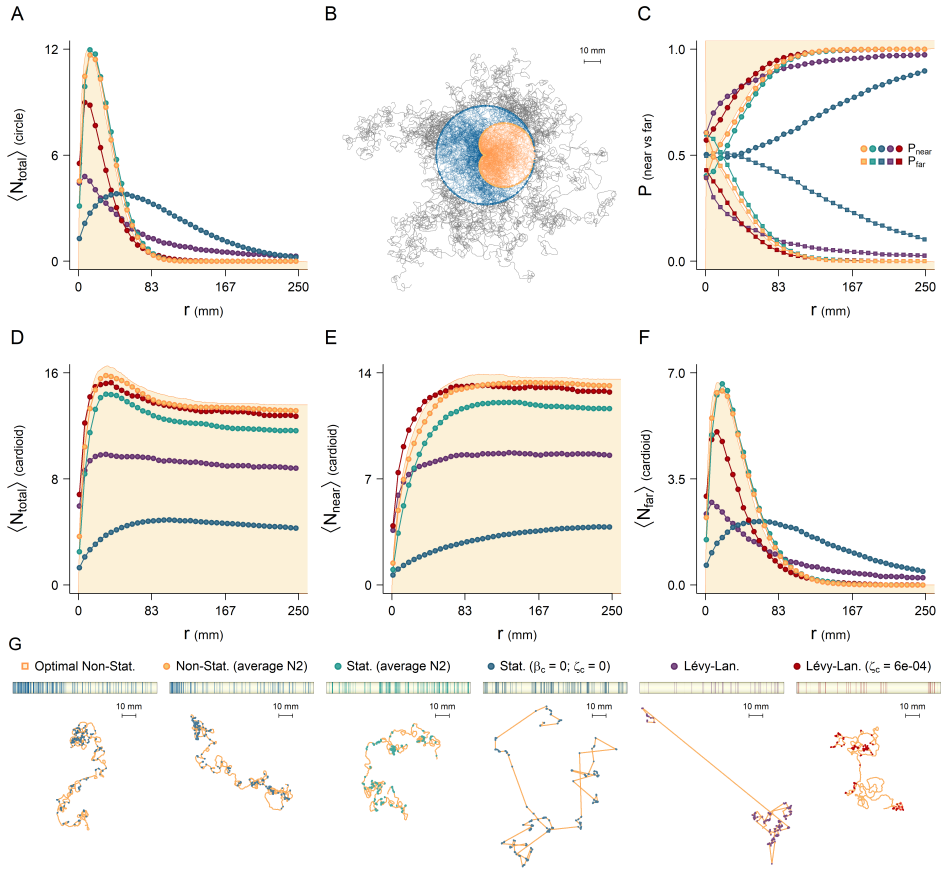


Figure 7.9: Search efficiency simulations in circular and cardioid geometries with truncated time (90 minutes) for different bio-inspired random walks. (A) Average number of hits with the edge of circles of increasing size (i.e., radius). (B) Schematic representation of the simulations. In the example, both the circle and the cardioid geometries have a radius of 30 mm, which is approximately the radius of the mean square displacement (i.e., $r = \sqrt{\text{MSD}/\pi}$) of *C. elegans*, evaluated after 90 minutes. The tracks depicted correspond to the non-stationary model parameterized as the average N2 individual. (C) Probability of near and far hits within cardioids of increasing size. (D) Average total number of hits with the edge of cardioids of increasing size (i.e., radius). (E, F) Average number of near and far hits in the cardioid. (G) Representative examples of the tracks generated by the random walk models under study. In all cases, simulations were carried with $\text{N} = 10^4$ synthetic tracks, parameterized using the average N2 individual as reported in Tables 6.1-6.4 from Chapter 6. The solid yellow polygon depicted in the plots show the results obtained with the non-stationary model with optimal memory decays $k_c = -0.00096$ and $k_s = -0.00065$ (note that these optimal values are close to the average *C. elegans* behavior). These memories maximized the anomalous diffusion exponent α of the MSD (see Figure 6.17B, E), and maximizes near and far hits. Crawl durations in the Lévy-Langevin walk were modelled assuming a truncated power law with scaling exponent: $\mu = 2$, and boundaries: $a = 0.061$, $b = 10^4$ seconds. The parameter β_c controls the curvature bias of the crawls and ζ_c the sinuosity.

Table 7.3: Relevant spatial scales of search, expressed in millimeters, for the different random walk models under study: optimal non-stationary model (Non-Stat.†) with memory decays $k_c = -0.00096$ and $k_s = -0.00065$, non-stationary model (Non-Stat.), stationary model (Stat.), stationary model with knocked out curvature and sinuosity (Stat.‡), Lévy-Langevin walk (Lévy-Lan.), Lévy-Langevin walk with sinuosity (Lévy-Lan.§, $\zeta_c = 0.00059$). Crawl durations in the Lévy-Langevin walks were modelled assuming a truncated power law with scaling exponent: $\mu = 2$, and lower and upper boundaries of $a = 0.061$ and $b = 10^4$ seconds, respectively. In all cases, simulations were carried with $N = 10^4$ synthetic tracks, parameterized using the average N2 individual as reported in Tables 6.1-6.4 from Chapter 6. Column MSD_r reports the radius of the mean square displacement (i.e., $r = \sqrt{MSD/\pi}$) evaluated after 90 minutes, which is the duration of the simulations. $\alpha \pm \sigma_\alpha$ reports the anomalous diffusion exponent of the random walks models with their standard error. Columns Circ. and Card., indicate the optimal radius (radius showing the maximal number of hits) for the circle and the cardioid, respectively. For the latter we also show the optimal radius for the near (Near) and far (Far) hits separately. Finally, Bal. reports the radius for which the probability of finding near and far targets in the cardioid is balanced (i.e., $P(near) = P(far)$).

Model	MSD_r	$\alpha \pm \sigma_\alpha$	Circ.	Bal.	Card.	Near	Far
Non-Stat.†	25.7	1.21 ± 0.187	11.3	9.3	32.5	99.1	15.3
Non-Stat.	26.5	1.16 ± 0.018	15.0	10.1	33.8	148.1	17.8
Stat.	27.0	1.01 ± 0.011	14.3	19.1	34.8	124.9	19.8
Stat.‡	86.3	1.12 ± 0.016	50.8	33.9	105.4	250.0^{*}	59.6
Lévy-Lan.	117.7	1.89 ± 0.020	4.5	-	32.8	130.1	5.0
Lévy-Lan.§	31.0	1.17 ± 0.015	9.8	-	35.3	89.6	12.8

Table 7.4: Average number of encounters at the scales of maximal search efficiency, for the random walk models under study: optimal non-stationary (Non-Stat.†) with memory decays $k_c = -0.00096$ and $k_s = -0.00065$, non-stationary (Non-Stat.), stationary (Stat.), stationary with knocked out curvature and sinuosity (Stat.‡), Lévy-Langevin walk (Lévy-Lan.), and Lévy-Langevin walk with sinuosity (Lévy-Lan.§, $\zeta_c = 0.00059$). Crawl durations in the Lévy-Langevin walks were modeled assuming a truncated power law with scaling exponent: $\mu = 2$, and lower ($a = 0.061$) and upper ($b = 10^4$) boundaries (in seconds). In all cases, simulations were carried out with N = 10^4 synthetic tracks, parameterized using the average N2 individual as reported in Tables 6.1-6.4 from Chapter 6. Column names indicate the number of hits with the edge of the circle and the cardioid, and the number of near and far hits in the cardioid. All average values are reported with their associated standard deviation. The table is a continuation of Table 7.3.

Model	⟨N circ.⟩	⟨N card.⟩	⟨N near⟩	⟨N far⟩
Non-Stat.†	12.0 ± 9.03	16.5 ± 10.11	13.9 ± 10.79	6.7 ± 7.10
Non-Stat.	11.7 ± 8.77	15.8 ± 9.81	13.4 ± 10.66	6.5 ± 7.06
Stat.	12.0 ± 8.99	14.5 ± 9.55	12.1 ± 10.20	6.6 ± 7.35
Stat.‡	3.8 ± 3.01	4.9 ± 3.42	3.9 ± 3.44	2.1 ± 2.48
Lévy-Lan.	4.9 ± 6.17	9.9 ± 8.67	8.7 ± 8.94	2.7 ± 4.51
Lévy-Lan.§	9.2 ± 9.16	15.4 ± 11.47	13.2 ± 11.59	5.1 ± 6.93

7.4.3 Turning dynamics and its impact on search performance

In this section, we investigated the impact of the curvature bias, the sinuosity, and the frequency of sharp turns on the search performance of *C. elegans*. To do so, we altered the non-stationary model developed in Chapter 6 by knocking-out the non-stationary dynamics affecting the rate of sharp turns and crawls over time. Instead, we used fixed rates leading to a constant frequency of sharp turns. By doing this, we could neatly quantify the effect of turning dynamics with three specific movement parameters (i.e., curvature, sinuosity, and sharp turns frequency) without the influence of non-stationary trends. The specifics of how we implemented this cancellation are detailed below.

Impact on search performance of curvature bias and sinuosity

We conducted N-passage time simulations of 90 minutes in length, varying β_c (curvature bias) and ζ_c (sinuosity) parameters across a range of values (Figure 7.10). Search performance was assessed by quantifying the number of hits (encounters) on both circular (Figure 7.10B) and cardioid (Figure 7.10C, E-H) geometries, with radii scaled to $r = 23$ mm. This specific radius was chosen to effectively demonstrate the positive influence of the curvature bias and the sinuosity on search performance, but other radii ranging from $r = 19$ to $r = 35$ mm yielded similar results.

Beyond these spatial scales, the beneficial impact of β_c and ζ_c became less pronounced. Consequently, it should be noted that reported findings are contingent upon both time and spatial scales.

Heatmaps shown in Figure 7.10 were constructed by simulating ensembles of 1000 synthetic trajectories for 10×10 pairs of β_c and ζ_c values. These synthetic trajectories were generated using the non-stationary model, parameterized as the average N2 individual, as reported in Tables 6.1-6.2, except for the parameters β_c, ζ_c that were scanned, and the parameters defining the rate of crawls, which were fixed to be $a_c = a_s = k_c = k_s = 0$, $b_c = 0.021$ and $b_s = 0.035$, to hold the frequency of sharp turns constant at the average frequency of the strain N2. Importantly, this parameterization of the model allowed us to assess the effect of the curvature bias β_c and the sinuosity ζ_c without the influence of non-stationary dynamics in the frequency of sharp turns and crawls.

We show that combinations of increased curvature and sinuosity lead to smaller mean square displacement (MSD) radius (Figure 7.10A). A small amount of curvature and a good amount of sinuosity increased search performance (total number of hits) in the circular geometry (Figure 7.10B). Noteworthy, approaching the ballistic motion (i.e., $\beta_c \to 0$ and $\zeta_c \to 0$) is not as bad as having a strong curvature, but it is not the best strategy at all. In the cardioid geometry (Figure 7.10B) the best search performances require introducing a strong curvature and low sinuosity. By comparing Figure 7.10C and Figure 7.10F, we clearly see that in the cardioid geometry the overall search performance is governed by the nearby hits, as they are 4-fold the number of far away hits. In relative proportions, (Figure 7.10E, G) the near and far hits show the exactly opposite patterns, as the two proportions must sum one.

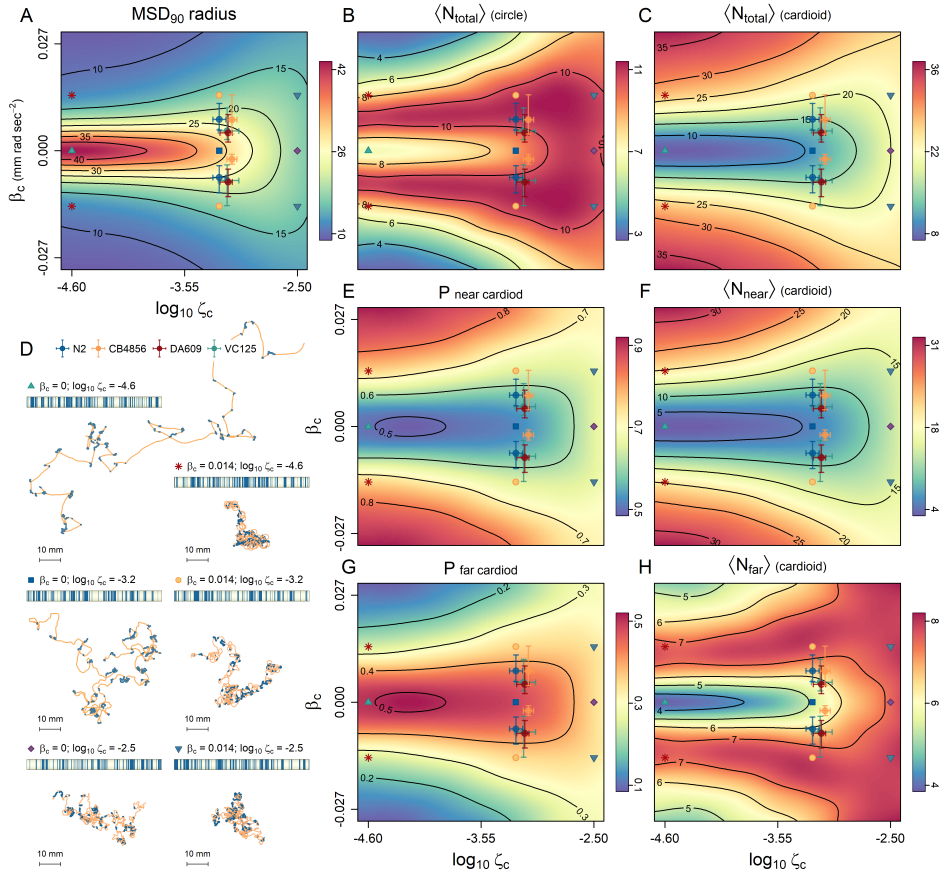


Figure 7.10: Search performance of the non-stationary model as a function of the curvature bias β_c and the sinuosity ζ_c. All model parameters were fixed at the values of the average N2 individual, as reported in Tables 6.1-6.2, except those defining the rate of crawls and sharp turns over time (see Eq. (6.10) and Eq. (6.10) in Chapter 6), which were set to yield a constant frequency of sharp turns (i.e., the average of the N2 strain). (A) Radius of the mean square displacement (i.e., $r = \sqrt{\mathrm{MSD}/\pi}$), evaluated at the end of the simulation (90 minutes). (B, C) Average number of encounters in the circle and the cardioid, respectively. (D) Representative examples of the tracks generated by the model at relevant combinations of β_c and ζ_c. (E, F) Probability and average number of near encounters within the cardioid. (G, H) Probability and average number of far encounters within the cardioid. All heatmaps are smoothed representations of the raw simulations, which comprised $10 \times 10 = 100$ combinations of different β_c and ζ_c values. Average values depicted in the figures were obtained using ensembles of 1000 synthetic trajectories, moving within circular and cardioid geometries of radius $r = 23$ mm. Colored crosses correspond to the empirical average values of β_c and ζ_c for strains N2, CB4856, DA609 and VC125. These values were computed from trajectory data of the experiments detailed in Chapter 3.

Impact on search performance of loops and sharp turns

Following the exact same approach, we analyzed the impact on search performance of the curvature bias (β_c), acting in combination with a constant sinuosity value, and the frequency of sharp turns (here controlled through parameter b_c). As in the previous section, we modified the *C. elegans* non-stationary model into a stationary model. All the parameters, including ζ_c, were fixed at the values of the average N2 individual (Tables 6.1-6.2), except for: (i) β_c and b_c, that were scanned, and (ii) the parameters defining the rate of crawls and sharp turns over time, which were set to be $a_c = a_s = k_c = k_s = 0$, and $b_s = 0.035$.

Like in the previous analysis, we conducted N-passage time simulations on both circular (Figure 7.11B) and cardioid (Figure 7.11C, E-H) geometries, with radii scaled to $r = 23$ mm. Other radii ranging from $r = 14$ to $r = 35$ mm yielded similar results. Beyond these spatial scales, the impact of β_c and b_c became less apparent.

The results (Figure 7.11) helped us to gain intuition about the contrasting effects of loops (smooth and directionally persistent) and sharp turns (abrupt and directionally random). We show that the radius of the MSD is maximum at small curvature and sharp turn rate values. In the circular geometry, a small amount of curvature improved the number of hits, and sharp turns show some relatively optimal values if combined with certain amount of curvature. So loops, prevent somehow the negative effects of sharp turns in the circular geometry (Figure 7.11B). In the cardioid geometry we got opposite results. Large curvature values improved search performance, when combined with small sharp turn rates (Figure 7.11C). Increased sharp turn rates showed a positive impact on search efficiency only if the curvature bias was zero. In any other case, the larger the rate of sharp turns the smaller the search efficiency. In the cardioid geometry, the overall efficiency pattern is driven by the near hits (Figure 7.11C, F). Noteworthy, far hits are more successful with large values of curvature and small sharp turn rates (Figure 7.11H).

All in all, the presence of a curvature bias seemed to be a key feature to augment near target detection by increasing spatial overlap and revisits. This feature governed and increased the overall search efficiency. On the contrary, sharp turns tend to smooth out the positive effect of the curvature bias. Therefore, they are only beneficial at low rates, and if combined with specific curvature values. Only at the limiting case of $\beta_c = 0$ (no curvature) increased sharp turn rates have a slight positive influence on the search efficiency.

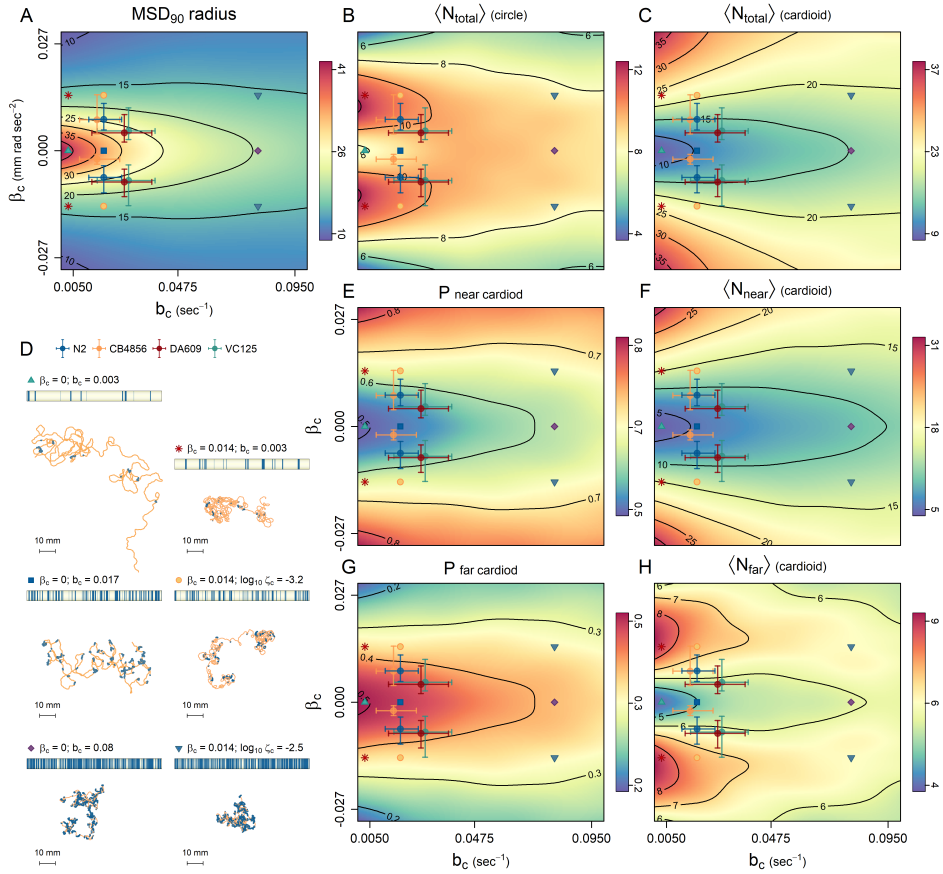


Figure 7.11: Search performance of the non-stationary model as a function of the curvature bias β_c and the frequency of sharp turns, parameterized by the baseline rate of crawls b_c. All model parameters were held constant at the values of the average N2 individual, detailed in Tables 6.1-6.2, except for those governing the rates of crawls and sharp turns over time (see Eq. (6.10) and Eq. (6.10) in Chapter 6), which were adjusted to maintain a constant frequency of sharp turns based solely on the baseline rate of crawls b_c. (A) Radius of the mean square displacement (i.e., $r = \sqrt{\text{MSD}/\pi}$), evaluated at the end of the simulation (90 minutes). (B, C) Average number of encounters in the circle and the cardioid, respectively. (D) Representative examples of the tracks generated by the model at relevant combinations of β_c and b_c. (E, F) Probability and average number of near encounters within the cardioid. (G, H) Probability and average number of far encounters within the cardioid. All heatmaps are smoothed representations of the raw simulations, which comprised $10 \times 10 = 100$ combinations of different β_c and b_c values. Average values depicted in the figures were obtained using ensembles of 1000 synthetic trajectories, moving within circular and cardioid geometries of radius $r = 23$ mm. Colored crosses correspond to the empirical average values of β_c and b_c for strains N2, CB4856, DA609 and VC125. These values were computed from trajectory data of the experiments detailed in Chapter 3.

7.4.4 Fitness as a consequence of search performance

Here, we present the results of the foraging experiments conducted to complement the synthetic search simulations with real data (see Sections 7.3.2-7.3.6). These experiments involved placing individual *C. elegans* of the strain N2 on large foraging plates (24.5 × 24.5 cm) containing 55 patches of bacteria, spatially distributed to mirror either the symmetric (homogeneous landscape) or the asymmetric (heterogeneous landscape) search condition. Importantly, the experimental scales of the landscapes were large enough for the worms to require exploitation-exploration trading, specially in the heterogeneous landscape. Navigating between clusters while exploiting patches within them demanded a sampling strategy to improve the chances of discovering new patches and exploring distant, uncharted clusters.

We counted both the number of patches discovered by the worms and the number of eggs laid in each type of landscape. While significant differences were observed in the number of patches discovered by the worms (Figure 7.12A), we did not observe differences in the overall number of laid eggs (Figure 7.12B). Nonetheless, we did observe a consistent positive and linear correlation between the number of patches found and the number of laid eggs in the two environments (Figure 7.12C). Interestingly, the increase in the number of laid eggs per patch found seemed to be slightly smaller in the heterogeneous landscape than in the homogeneous landscape.

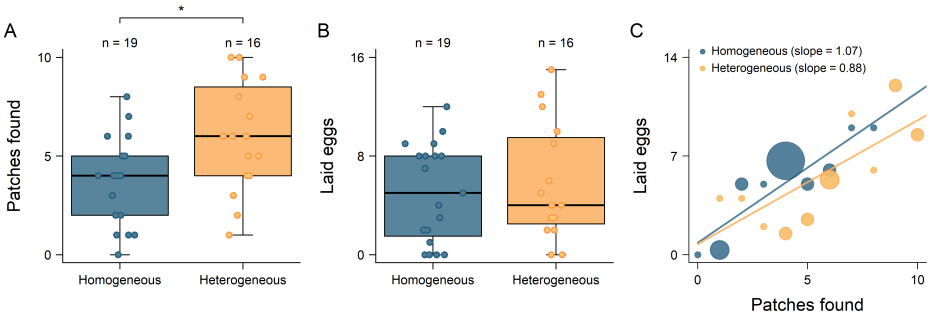


Figure 7.12: Results of the foraging experiments conducted with *C. elegans* (N2) and *E. coli* OP50 in homogeneous (symmetric search condition) and heterogeneous (asymmetric search condition) landscapes. (A) Total number of patches discovered by the worms, either completely or partially consumed, across the two experimental landscapes under evaluation. (B) Overall number of eggs deposited in each landscape. (C) Average number of eggs laid per patch discovered. The size of the dots is proportional to the number of data-points used to compute the averages (i.e., $n = 1, 2, 3, 6$), and the colored lines show a linear model fitted to data (Homogeneous: $R^2 = 0.80$, Heterogeneous: $R^2 = 0.51$). In all cases, the foraging landscapes comprised 55 patches containing 0.75 µL of *E. coli* OP50 suspended in foraging buffer at OD = 0.2. Comparison analyses were made using t-tests with a significance level $\alpha = 0.05$ (R package *rstatix*), and significant differences were found for the comparison in panel A.

7.5 Discussion

We demonstrated that turning dynamics, rather than speed, constitutes the main driver of search performance in *C. elegans*. At the local scale, we found that the curvature bias and sinuosity of *C. elegans* movement patterns compensate detection errors by introducing a certain amount of revisits ends enhancing spatial coverage. At the large scale, the frequency decay of sharp turns over time, and its consequential superdiffusion (Chapter 6), allows *C. elegans* to explore by trading proximal and distal resource encounters. All in all, *C. elegans* unfolded intensive-extensive or local-to-large scale sampling patterns while keeping a constant looping behavior that promoted spatial overlap and sampling revisits across scales.

These findings were corroborated by both synthetic simulations and empirical observations. Synthetic simulations highlighted the significance of considering the spatial distribution of targets (i.e. searcher-to-target distances) and the imposition of time constraints when assessing search efficiency. Additionally, a quantitative analysis of the relationship between superdiffusion and search performance using *C. elegans*-inspired random walk movement models, revealed that accelerated spreading (i.e., $\alpha > 1$) can play a pivotal role in animal search behavior. In alignment with previous findings (Hills et al., 2004; Gray et al., 2005; Klein et al., 2017; López-Cruz et al., 2019), our results indicate that the transition from local to large scale search patterns, constitutes a fundamental aspect of the search strategy of *C. elegans*. Furthermore, we also found that the tendency to smoothly turn clockwise (or anticlockwise) is a better buffer against error detection than increasing the rate of random sharp turns, as the probability of revisiting the space grows more effectively.

The study of intensive-extensive sampling patterns within superdiffusive stationary models like Lévy walks constitutes a crucial component of contemporary random search theory (Viswanathan et al., 1996, 2011; Bartumeus et al., 2014, 2016; Méndez et al., 2014). *C. elegans* manifests a random walk behavior exhibiting similar properties to a stationary Lévy walk with sinuosity (Lévy-Lan, $\zeta_c > 0$). This model stands out as one of the most effective search models we tested, wherein the balance between intensive and extensive sampling can be finely tuned through a level of sinuosity that maximizes proximal encounters. A comparable stationary model has been proposed for the flagella-driven motion of marine planktonic cells (Bartumeus and Levin, 2008).

However, in the case of *C. elegans*, the transition from intensive to extensive sampling behavior unfolds in a significantly more ordered manner over time. While non-stationary sampling patterns are typically associated with relocation experiments (Flavell et al., 2020), the results of foraging experiments with *C. elegans*

suggest that in this instance, such patterns might actually reflect a generic patch-leaving behavioral template.

Based on the trajectory analyses conducted throughout the thesis and the results obtained from the synthetic simulations presented in this chapter, it is plausible to posit that the non-stationary turning dynamics is the main determinant explaining why *C. elegans* found more food in heterogeneous (patchy) landscapes compared to homogeneous ones. Specifically, the curvature bias leading to looping behavior and the time-decaying sharp turns (resulting in superdiffusion) enabled the worms to effectively exploit bacterial patches within clusters while simultaneously traversing the considerable distances separating different clusters of patches in the heterogeneous landscape. We propose that the worm is propelled by an internal memory of recent past feeding experiences, which generates a motor template aiding the worm in initially exploiting and adhering to a patch before gradually or abruptly (as happen in relocation experiments) departing once resources have been depleted.

Additionally, it is noteworthy that *C. elegans* appeared to monitor the quantity of food consumed over time and adjust its investment in reproduction (egg laying) accordingly. This finding suggests a rudimentary form of parental care, whereby the worm allocates more resources to offspring production in response to higher environmental resource availability, likely aimed at enhancing the fitness of the progeny. Although indirectly, the foraging experiments have contributed to advancing our comprehension of the potential role that superdiffusion may play in *C. elegans* search behavior within more realistic environmental settings.

Further research is warranted to enhance statistical power and elucidate the behavioral mechanisms underlying foraging at the landscape level. Conducting analogous foraging experiments with *C. elegans* defective mutants in hunger sensation and memory retention could provide invaluable insights into the underlying biological mechanisms that shape optimal foraging and their potential eco-evolutionary ramifications. Specifically, we aim to delve deeper into the impact on exploitation-exploration tradeoffs of non-stationary versus stationary generative mechanisms of superdiffusion driven by a combination of smooth and sharp turning dynamics.

Chapter 8

Discussion

This thesis offers a comprehensive understanding of *Caenorhabditis elegans* movement and search behavior in a cue-deprived environment. From a biological perspective we found that *C. elegans* search trajectories are primarily built upon two distinct behavioral features: (i) a memory relaxation of past experience in a plentiful resource patch, reflected in the decaying frequency of sharp turns over time, and (ii) the curvature bias inherent in crawls, often resulting in complex looping patterns. From a physics perspective, the main variability observed in the trajectories stem from the non-stationary dynamics of the velocity vector, mainly influenced by time-decreasing fluctuations in angular velocity. Building upon these two complementary perspectives of the same phenomenon, the one from biology and the one from physics, we could reinforce and broaden the interface between behavioral ecology and statistical physics, providing a self-contained case study that addresses fundamental aspects in both fields. For example, biological determinants require to modify the basic and most fundamental movement descriptions of physics, for a number of reasons: (i) a systematic curvature bias; (ii) sequences of combined turns and speeds leading to distinct behavioral modes; (iii) independent modulation of sharp turns and crawling events; (iv) variations in the straightness of different crawling types; (v) the presence of time-dependent behavioral transitions; and (vi) the generation of complex movement patterns by stationary behavioral components. Therefore, while the run-and-tumble model has been employed to describe the movement of many organisms, it is too simple to describe the search movement of *C. elegans*.

It is crucial, at this point, to revisit and discuss a fundamental assumption of this work: the notion that the analyzed data reflects a genuine exploratory process. Specifically, a movement process characterized by very limited information. We posit that in an environment devoid of resources, and thus lacking clear gradients or signals toward food, worm movement is predominantly influenced by changes in

internal states rather than external stimuli. This is not to suggest the absence of external stimuli; rather, the lack of information about resources is indeed a clear external stimulus that can activate certain internal modules. Likely, there might be also microscale landscape irregularities that could locally explain certain changes of behavior. Indeed, we do not really know what type of default chemo-sensory circuitry is operating when the worm moves in an agar plate deprived of bacteria, and how it is connected to metabolic internal states.

Among internal drivers, we might think on metabolic cascades related to starvation (Pradhan et al., 2019; Flavell et al., 2020), but it is also true that in *C. elegans* starvation may occur at broader time scales than 90 minutes. Probably, a combination of both information processing—via the accumulation of unsuccessful food encounters over time—and changes in the metabolic state of the worm, determines how the rate of sharp turns and crawls change over time, ultimately shaping the overall superdiffusive spreading pattern observed through the 90 minutes of experiment. An interpretation that would be in clear alignment with previous research indicating multiple regulatory mechanisms acting at the neuronal, metabolic and genetic level, ensure the execution of area restricted search behavior right after food removal with the subsequent transition to global search or dispersal patterns (Hills et al., 2004; Gray et al., 2005; Gloria-Soria and Azevedo, 2008; Hakvey, 2009; López-Cruz et al., 2019; Pradhan et al., 2019). Notably, all individuals under study, including both wild type (N2, CB4856) and mutant strains (DA609, VC125), exhibited signatures of this ordered behavioral template. However, the two metabolically knock-out strains (DA609, VC125) showed more subtle changes of behavior across time, suggesting a dampened response to the lack of information or resources. For example, we observed the memory rates k_c and k_s (measured in sec^{-1}) were (DA609 $[k_c = -0.0029,\ k_s = -0.0005]$, VC125 $[k_c = -0.0017,\ k_s = -0.0013]$) for the two mutant strains, compared to (N2 $[k_c = -0.0021,\ k_s = -0.0017]$, CB4856 $[k_c = -0.0021,\ k_s = -0.0024]$) for the two wild type strains. Likewise, the overall coverage area at the end of the experiment (measured in mm^2) was smaller in the mutant strains (DA609=2204, VC125=1211), than in the wild type strains (N2=2603, CB4856=2665).

Further research is needed to better understand the source and scope of intra- and inter-strain variability in the execution of this motor template. Given the short life span of *C. elegans*, such a well-defined and long stereotyped behavior, which last more than one hour, raise also fundamental questions regarding the concept of free will (Heisenberg, 2009) and its manifestation in living organisms (Sapolsky, 2023). Indeed, motor biological constraints and natural selection are major forces shaping successful templates from a much wider pool of available options. For instance, the search behavior of *C. elegans* involves long transient superdiffusion despite the

occurrence of a systematic curvature bias, one of the most salient features of its trajectories. This curvature imposes looping behavior, spatial overlapping and slow spreading. However, we observed that sharp turns can break these loops and enable the worm to spread faster when executed with a moderate frequency. Considering the built-in limitations of *C. elegans* spreading capabilities, the modulation of crawling and sharp turns was close to an optimal (or marginally optimal) memory decay, which maximized the non-linear spreading response in a context of lack of resources. In other words, among all possible modulations of sharp turns decay over time, the one implemented by *C. elegans* is compatible with an exploration process exploiting at best superdiffusive properties. Other modulations could lead, for example, to the emergence of sub-diffusive or diffusive spreading patterns. Importantly, our work illustrate the challenges of using Hurst decomposition analysis to harness the determinants of diffusion in complex and non-stationary movement processes. For example, non-stationarity on turns may not be detected and emerge as long-range correlations in the spreading velocity.

Through search simulations and foraging experiments in resource rich landscapes we found that the temporal order of these behavioral patterns, which maximize superdiffusion, enables *C. elegans* to optimize its search performance in scenarios where both nearby and faraway targets may be accessible (i.e., asymmetric conditions). *C. elegans* allocates its sampling effort in a highly structured manner. First by capitalizing on the searching nearby, and then transitioning to broader exploratory patterns. In our case, strong site fidelity at the early exploration times accounted for previous experiences of having food in the surrounding area, but it is also important to take safety and risk-aversion into consideration.

In the asymmetric search conditions, encompassing both nearby and faraway targets, optimal superdiffusion is evident in stationary random walk models such as Lévy walks with a scaling exponent of $\mu = 2$ (Bartumeus et al., 2002; Sims et al., 2008; Viswanathan et al., 2011; Bartumeus et al., 2014). Thus, it appears that *C. elegans* exhibits analogous properties akin to optimal Lévy walks. In Lévy walks, the stochastic temporal order arises from sampling a power law distribution of move lengths. Area-restricted search behavior predominantly manifests during the initial movement steps, characterized by highly probable, small move lengths, whereas extensive search activity becomes more prominent subsequently, coinciding with the occurrence of larger and less likely move lengths. Further investigation is warranted to elucidate the significance of temporal ordering in the production of optimal superdiffusion. For instance, a comparative analysis of search efficiency between two exploratory processes, one time-dependent and the other entirely random, that share the same anomalous diffusion scaling exponent and mean squared displacement (MSD) within a fixed time-frame, could offer valuable insights. There

are many biological pathways in order to reach certain stochastic properties that are deemed useful in exploratory processes, others are useless or counterproductive. By leveraging *C. elegans* as a model organism for the study of search, we have gained valuable insights into the fundamental principles of search ecology, connecting biological facts with optimal search theory.

Our work sets the seed to perform future experiments and orient new questions related to search. How different would be patch leaving behavior (not one coming from a relocation process but actual patch departure behavior)? How different would be the movement of *C. elegans* performing taxis in a resource gradient? Could we build information gradients and produce transitions from taxis to exploratory movements? How would parameters in our *C. elegans* movement model be changed? The potential of working with *C. elegans* also makes it possible to think about combining movement tracking with in vivo expression (optogenetics), or pan-neuronal imaging. We hope to be the ones opening a new field of research in *C. elegans* focused on large scale spatial sampling and search behavior as something different from classic feeding behavior experiments at the Petri Dish scale.

Chapter 9

Conclusions

In the following paragraphs we outlined the main conclusions obtained throughout the course of this research, structured in the three specific objectives (Ob) defined in Chapter 2:

Ob1 - *Characterize the spreading pattern of C. elegans in the context of relocation experiments and diffusion theory.*

- The time scaling analysis of absolute increment displacements, squared increment displacements, and time-averaged mean square displacement is well suited to delineate the phenomenological origin of anomalous diffusion in *C. elegans*. However, getting a more mechanistic understanding of this process requires an explicit analysis of turning dynamics.
- Hurst exponent decomposition indicates non-stationary dynamics and long range correlations mediate anomalous diffusion in *C. elegans*.
- Hurst exponent decomposition suggests *C. elegans* has three distinct spreading regimes. However, its mean square displacement curve reflects a more complex picture, showing at least four distinct regimes.
- A heavy-tailed distribution of crawl displacements (or durations) can not explain superdiffusion in *C. elegans*, due to the strong effect that both the curvature bias and the sinuosity have on diffusivity.
- The rate of sharp turns and crawls exhibit time-dependent dynamics that lead to non-stationary spreading. Thereby, elucidating the emergence of the Moses effect.

- The memory decay in the rate of sharp turns and crawls over time causes a long-transient superdiffusive regime in *C. elegans*. This relaxation of preconditioning experiences (from well-fed to no-food) elicits a transition from intensive to extensive search patterns, causing a non-linear spreading.
- *C. elegans* individuals under study show optimal, or marginally suboptimal, memory relaxation times that maximize superdiffusion.

Ob2 - *Understand C. elegans movement across scales, identifying the mechanistic links between microscopic, mesoscopic, and macroscopic trajectory patterns.*

- Instantaneous speed, either in the form of a distribution or evaluated as a temporal series, reflects the twofold nature of *C. elegans* movement behavior, composed of sharp turns and crawls.
- The annotation algorithm developed, based exclusively on speed statistics, captures both sharp turns and crawls in detail.
- At the microscopic scale, sharp turns are associated with low speeds and high angular velocities; while crawls, are associated with high speeds and low angular velocities.
- At the mesoscopic scale, sharp turns are associated with short durations, small displacements, and big turning angles; while crawls, are associated with long durations, big displacements, and small turning angles.
- *C. elegans* increases the rate of sharp turns with time while decreases the rate of crawls. Meaning, sharp turns tend to be shorter and crawls tend to be larger over time. Ultimately, these time-dependent dynamics lead the frequency of strong reorientation events decrease over time.

Ob3 - *Quantitatively asses the search efficiency of C. elegans movement in key limiting case scenarios.*

- Turning dynamics, comprising the curvature bias, the sinuosity and the frequency of sharp turns, constitutes the main driver of search performance in *C. elegans*.
- The accelerated spreading produced by the frequency decay of sharp turns over time, enables *C. elegans* to increase nearby sampling effort without losing performance for finding far resources.

- Looping behavior allows *C. elegans* tackling detection errors and uncertainty while searching. Such effect is already evident in symmetric search conditions, but it is even stronger in asymmetric search conditions.
- The specific temporal order of *C. elegans* search behavior constitutes an optimal rule of thumb to find both near and far targets. Such optimization is greater when the transition from spatially intensive to extensive search patterns generates superdiffusion. Contrary, disturbing the order of this transition leads to increasingly inefficient movement patterns.

Technical appendices

Appendix A

Alternative ways of decomposing the Hurst exponent

A.1 The absolute spreading velocity

Utilizing methods delineated in Aghion et al. (2021) and Vilk et al. (2022), we computed the absolute spreading velocity employing the formula $|\mathbf{v}_i| = |d_{i+1} - d_i|/\Delta$. Where, i represents the observation time, with each observation separated from the next by a regular time stamp of size Δ, and d denotes the distance to the origin of the trajectory under examination. Analogous to our treatment of absolute increment displacements $|\mathbf{d}_i|$ (Chapter 4, Section 4.2.4), bold vector notation is utilized to signify spreading velocity can exhibit both positive (outward from the origin) and negative (returning toward the origin) values. In the literature (Aghion et al., 2021; Vilk et al., 2022; Meyer et al., 2022) what the spreading velocity is commonly referred as the average velocity, or simply velocity. In this regard, we observe potential for confusion in certain contexts, as the spreading velocity is fundamentally different from the conventional understanding of velocity. Specifically, it disregards the dynamics of turns (angles), which are intrinsic to the concept of velocity. Notably, $\mathbf{v}_i$ reflects alterations in the radial component of a trajectory projected in polar coordinates, i.e., $x(t) = r(t)\cos(\phi(t))$ and $y(t) = r(t)\sin(\phi(t))$. Consequently, a walk tracing the perimeter of the unit circle, for instance, would register $\mathbf{v}_i = 0$, despite moving. To underscore this distinction, we designated $\mathbf{v}_i = 0$ as the spreading velocity, signifying its role as a measure of the rate at which a moving object departs from the origin.

A.1.1 Non-stationarity: Moses effect

The derivation of the Moses exponent M in terms of the absolute spreading velocity was first proposed by Aghion et al. (2021), found:

$$\langle |\mathbf{v}_i| \rangle = \left\langle \frac{\Delta}{t} \sum_{i=1}^{t/\Delta} |\mathbf{v}_i| \right\rangle \propto t^{\,M-1/2} \,, \tag{A.1}$$

where $\langle \cdot \rangle$ denotes the population ensemble average, i indicates the observation time, and $|\mathbf{v}_i|$ is the absolute spreading velocity $|\mathbf{v}_i| = |d_{i+1} - d_i|/\Delta$. Interestingly, the factor Δ/t makes the slope flat when $M = 0.5$, making more visual and intuitive the idea that $M = 0.5$ yields stationary behavior. Note that the conceptual change in going from Eq. (4.5) to Eq. (A.1) is transforming displacement into velocity increments, which involves a change of sign from plus to minus in the exponents at the right-hand side of Eq. (A.1) because time needs also to be multiplied in the right-hand side as t^{-1}.

A.1.2 Long-range correlations: Joseph effect

The Joseph exponent J can be characterized from the asymptotic behavior of the spreading velocity correlations as:

$$\int_0^{\tau} \frac{\langle \mathbf{v}(t)\mathbf{v}(t+\tau) \rangle}{\langle \mathbf{v}^2(t) \rangle} \, d\tau \propto \tau^{\,2J-1} \qquad \text{or} \qquad \frac{\langle \mathbf{v}(t)\mathbf{v}(t+\tau) \rangle}{\langle \mathbf{v}^2(t) \rangle} \propto \tau^{\,2J-2} \,, \tag{A.2}$$

where $\mathbf{v}$ is the signed spreading velocity, τ is a time lag, and $\langle \cdot \rangle$ denotes the population ensemble average.

A.1.3 Extreme events: Noah effects

Finally, the Noah exponent L can be characterized in terms of the squared spreading velocity, as proposed by Aghion et al. (2021) and Vilk et al. (2022):

$$\langle \mathbf{v}^2(t) \rangle = \left\langle \frac{\Delta}{t} \sum_{i=1}^{t/\mathrm{N}} \mathbf{v}_i^2 \right\rangle \propto t^{\,2L+2M-2} \,. \tag{A.3}$$

where all the notation is equivalent to Eq. (A.1).

A.2 Random Walk Models

A.2.1 Brownian Walk

Following Berg (1983), we implemented the Brownian walk as a run-and-tumble model with a normal distribution of step lengths expressed in terms of speed $v = \mathcal{N}(\mu, \sigma)$, and a uniform distribution of angles $\phi = U(-\pi, \pi)$. In this model, the coordinates of the walk are obtained as follows:

$$\mathbf{r}_{i+1} = (x_i + \Delta v \cos\phi,\, y_i + \Delta v \sin\phi) \ , \tag{A.4}$$

where $\mathbf{r}_{i+1}$ is the position vector of the walk at the next step $i+1$, x_i and y_i are the Cartesian coordinates of the walk at the current step i, Δ is the time lag between steps, v is the module of the velocity vector (i.e., the speed), and ϕ is the absolute angle with respect to the origin of coordinates. The Hurst exponent decomposition of the resulting trajectories is shown in Figure A.1.

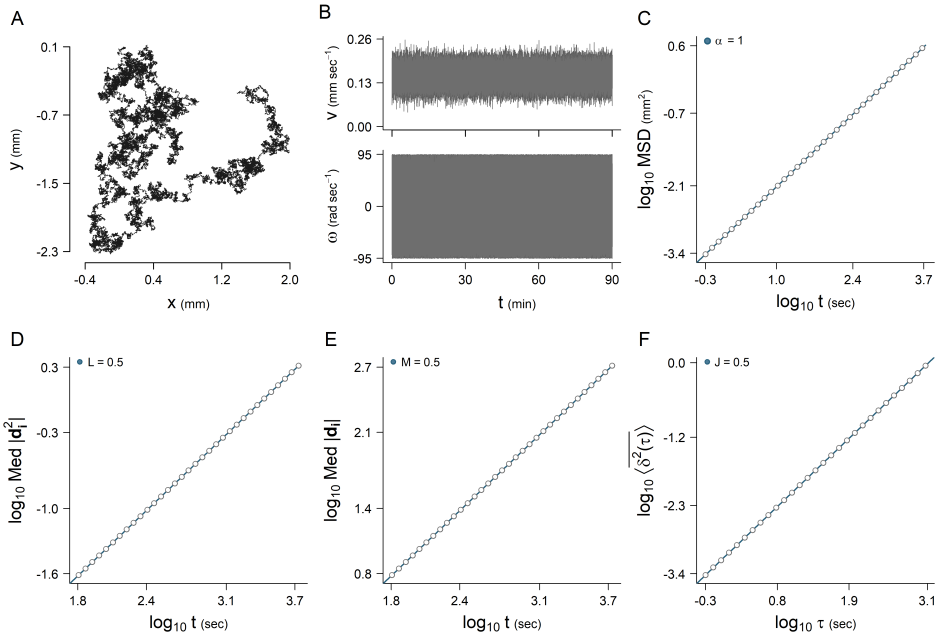


Figure A.1: Hurst exponent decomposition of the Brownian walk. (A) Illustrative example of a Brownian walk trajectory. (B) Temporal series of speed (upper panel) and angular velocity (lower panel) of the trajectory shown in panel A. (C) Mean square displacement with the anomalous diffusion exponent α. (D) Time scaling of squared increment displacements with the Noah exponent L. (E) Time scaling of absolute increment displacements with the Moses exponent M. (F) Ensemble time-averaged mean square displacement with the Joseph exponent J. All estimates were obtained from ensembles of $\mathrm{N} = 10^4$ trajectories of 90 minute length. The following parameters were used to generate the trajectories: (i) mean speed $\mu = 0.15$ mm $\cdot$ sec^{-1}, (ii) standard deviation of speed $\sigma = 0.025$ mm $\cdot$ sec^{-1}, (iii) time step $\Delta = 0.033$ seconds.

A.2.2 Lévy Models

Lévy walk

Based on Viswanathan et al. (2011), we implemented the paradigmatic Lévy walk as a run-and-tumble model with a power-law distribution of step lengths $p(l) = l^{-\mu}$, constant velocity v, time step Δ, and finally a uniform distribution of angles $\phi = U(-\pi, \pi)$. The coordinates of the model were obtained by following this routine:

1. Sample a random angle ϕ from a uniform distribution $U(-\pi, \pi)$ to determine the direction of movement.

2. Sample a random move length l from a virtually unbounded power-law distribution $p(l) = l^{-\mu}$.

3. Walk the distance $l = \mathrm{v}\Delta$ in the direction specified by ϕ, updating the position coordinates according to $\mathbf{r}_{i+1} = (x_i + \Delta \mathrm{v} \cos\phi,\, y_i + \Delta \mathrm{v} \sin\phi)$ until the duration τ is fulfilled.

4. Repeat steps 1-3 until the end of the simulation.

The Hurst exponent decomposition of the resulting trajectories is shown in Figure A.2.

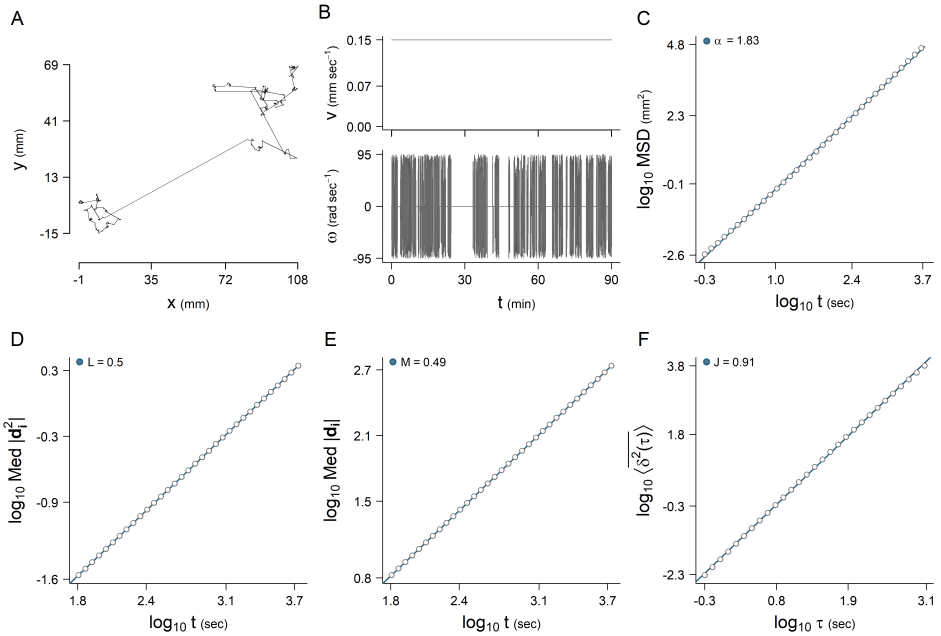


Figure A.2: Hurst exponent decomposition of the Lévy walk. (A) Illustrative example of a Lévy walk trajectory. (B) Temporal series of speed (upper panel) and angular velocity (lower panel) of the trajectory shown in panel A. (C) Mean square displacement with the anomalous diffusion exponent α. (D) Time scaling of squared increment displacements with the Noah exponent L. (E) Time scaling of absolute increment displacements with the Moses exponent M. (F) Ensemble time-averaged mean square displacement with the Joseph exponent J. All estimates were obtained from ensembles of $\mathrm{N} = 10^4$ trajectories of 90 minute length. The following parameters were used to generate the trajectories: (i) constant speed $\mathrm{v} = 0.15$ mm $\cdot$ sec^{-1}, (ii) scaling exponent of the power law $\mu = 2$ of displacement lengths, (iii) boundaries of the power law $\left[a = 0.0099, b = 7.5 \cdot 10^5\right]$ millimeters, (iv) time step $\Delta = 0.033$ seconds.

Lévy flight

We modeled the so-called Lévy flight as explained in Viswanathan et al. (2011). To implement this model, one needs to basically follow the same steps as in the

Lévy walk; however, here all move length displacements last for the same duration Δ (usually $\Delta = 0$ assuming instantaneous move lengths, but can be any fixed time unit). Therefore, the coordinates of the flights are obtained by running the following routine:

1. Sample a random angle from a uniform distribution $\phi = U(-\pi, \pi)$.
2. Sample a random move length from a virtually unbounded power-law distribution $p(l) = l^{-\mu}$.
3. Update the position vector assuming $\mathbf{r}_{i+1} = (x_i + l\cos\phi,\, y_i + l\sin\phi)$. Similarly, time is updated by adding Δ seconds at each step.
4. Repeat steps 1-3 until the end of the simulation.

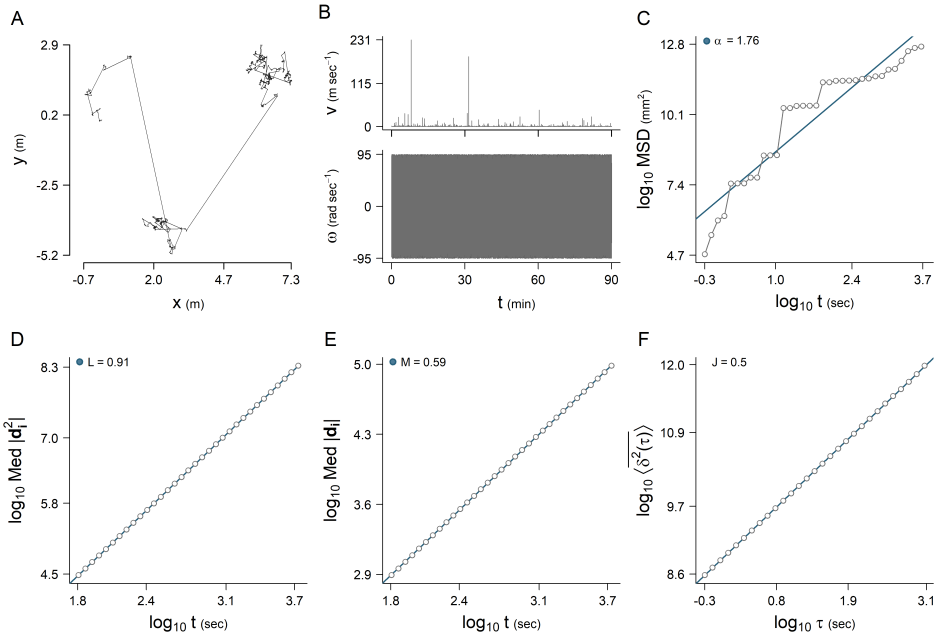


Figure A.3: Hurst exponent decomposition of the Lévy flight. (A) Illustrative example of a Lévy flight trajectory. (B) Temporal series of speed (upper panel) and angular velocity (lower panel) of the trajectory shown in panel A. (C) Mean square displacement (MSD) with the anomalous diffusion exponent α. Notably, the MSD curve exhibits extreme events, characterized by vertical increments, indicating wild accelerations, or nearly horizontal increments, corresponding to almost pauses. (D) Time scaling of squared increment displacements with the Noah exponent L. (E) Time scaling of absolute increment displacements with the Moses exponent M. (F) Ensemble time-averaged mean square displacement with the Joseph exponent J. All estimates were obtained from ensembles of N $= 10^5$ trajectories of 90 minutes length. The following parameters were used to generate the trajectories: (i) scaling exponent of the power law of displacement length $\mu = 2$, (ii) boundaries of the power law $\left[a = 0.0661, b = 5 \cdot 10^9\right]$ millimeters, and (iii) time step $\Delta = 0.033$ seconds. Note that due to the wild variability generated by the power law, the units in Panels A and B (speed) have been changed from millimeters to meters.

Appendix B

Mathematical functions

B.1 Probability density functions

Here we report a summary description of the probabilistic models used to fit the distribution of speed, angular velocity and duration associated with sharp turns and crawls. All mathematical expressions provided intend to document the fitting process we carried and not a strict mathematical definition of the functions. For example, we defined the support of mixture distributions just in the real positive line because in all cases, one of the two components i.e., Weibull, gamma, beta or exponential; is not defined for negative numbers.

Weibull Normal distribution

Support	$x \in [0, \infty)$				
Parameters	Shape	Scale	Mean	Variance	Probability
	$\varphi \in \mathbb{R}_{>0}$	$\theta \in \mathbb{R}_{>0}$	$\mu \in \mathbb{R}_{>0}$	$\sigma^2 \in \mathbb{R}_{>0}$	$p \in (0, 1)$
PDF	$f(x) = \underbrace{p\left(\frac{\varphi}{\theta}\left(\frac{x}{\theta}\right)^{(\varphi-1)} e^{-\left(\frac{x}{\theta}\right)^{\varphi}}\right)}_{\text{Weibull}} + \underbrace{(1-p)\left(\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}\right)}_{\text{Normal}}$				

Used to fit the distribution of speed.

Gamma Normal distribution

Support	$x \in [0, \infty)$				
Parameters	Shape	Scale	Mean	Variance	Probability
	$\kappa \in \mathbb{R}_{>0}$	$\vartheta \in \mathbb{R}_{>0}$	$\mu \in \mathbb{R}_{>0}$	$\sigma^2 \in \mathbb{R}_{>0}$	$p \in (0, 1)$

PDF

$$f(x) = p \underbrace{\left(\frac{1}{\Gamma(\kappa)\, \vartheta^{\kappa}}\, x^{\kappa-1}\, e^{-\frac{x}{\vartheta}} \right)}_{\text{Gamma}} + (1-p) \underbrace{\left(\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \right)}_{\text{Normal}}$$

Used to fit the distribution of speed.

Beta Normal distribution

Support	$x \in [0, \infty)$				
Parameters	Shape	Shape	Mean	Variance	Probability
	$\alpha \in \mathbb{R}_{>0}$	$\beta \in \mathbb{R}_{>0}$	$\mu \in \mathbb{R}_{>0}$	$\sigma^2 \in \mathbb{R}_{>0}$	$p \in (0, 1)$

PDF

$$f(x) = p \underbrace{\left(\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\, x^{1-\alpha}\, (1-x)^{1-\beta} \right)}_{\text{Beta}} + (1-p) \underbrace{\left(\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \right)}_{\text{Normal}}$$

Used to fit the distribution of speed.

Exponential Normal distribution

Support	$x \in [0, \infty)$			
Parameters	Scale $\theta \in \mathbb{R}_{>0}$	Mean $\mu \in \mathbb{R}_{>0}$	Variance $\sigma^2 \in \mathbb{R}_{>0}$	Probability $p \in (0, 1)$
PDF	$f(x) = p \underbrace{\left(\frac{1}{\theta} e^{-\frac{x}{\theta}}\right)}_{\text{Exponential}} + (1-p) \underbrace{\left(\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}\right)}_{\text{Normal}}$			

Used to fit the distribution of speed.

Normal distribution

Support	$x \in (-\infty, +\infty)$		
Parameters	Mean $\mu \in \mathbb{R}$	Variance $\sigma^2 \in \mathbb{R}_{>0}$	
PDF	$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$		

Used to fit the distribution of angular velocity.

Cauchy distribution

Support	$x \in (-\infty, +\infty)$		
Parameters	Location (mean) $x_0 \in \mathbb{R}$		Scale $\eta \in \mathbb{R}_{>0}$
PDF	$f(x) = \frac{1}{\pi}\left[\frac{\eta}{(x-x_0)^2+\eta^2}\right]$		

Used to fit the distribution of angular velocity.

Exponential distribution

Support	$x \in [0, +\infty)$
Parameters	Scale $\theta \in \mathbb{R}_{>0}$
PDF	$f(x) = \frac{1}{\theta} e^{-\frac{x}{\theta}}$

Used to fit the distribution of duration of sharp turns and crawls.

Double exponential distribution

Support	$x \in [0, \infty)$		
Parameters	Probability $p \in (0, 1)$	Scale $\theta \in \mathbb{R}_{>0}$	Scale $\beta \in \mathbb{R}_{>0}$
PDF	$f(x) = p\left(\frac{1}{\theta} e^{-\frac{x}{\theta}}\right) + (1-p)\left(\frac{1}{\beta} e^{-\frac{x}{\beta}}\right)$		

Used to fit the distribution of duration of sharp turns and crawls.

Gamma distribution

Support	$x \in (-\infty, +\infty)$		
Parameters	Shape $\kappa \in \mathbb{R}_{>0}$		Scale $\vartheta \in \mathbb{R}_{>0}$
PDF	$f(x) = \frac{1}{\Gamma(\kappa)\,\vartheta^{\kappa}}\; x^{\,\kappa-1}\, e^{-\frac{x}{\vartheta}}$		

Used to fit the distribution of duration of sharp turns and crawls.

Weibull distribution

Support	$x \in (-\infty, +\infty)$		
Parameters	Shape $\varphi \in \mathbb{R}_{>0}$		Scale $\theta \in \mathbb{R}_{>0}$
PDF	$f(x) = \frac{\varphi}{\theta}\left(\frac{x}{\theta}\right)^{(\varphi-1)} e^{-(\frac{x}{\theta})^{\varphi}}$		

Used to fit the distribution of duration of sharp turns and crawls.

Truncated power law

Support	$x \in [a, b]$		
Parameters	Scaling exponent $\mu \in (0, \infty)$	Memory $k \in \mathbb{R}$	Baseline rate $b \in \mathbb{R}_{>0}$
Function	$f(x) = e^{-\mu \log x - \log \frac{b^{1-\mu} - a^{1-\mu}}{1-\mu}}$		

Used to model times between reorientations in the Lévy walk and the Lévy-Langevin models.

B.2 Time dependent functions

Here we provide mathematical equations of time dependent functions used throughout the thesis.

Exponential (decay) with constant baseline

Support	$t \in [0, \infty)$		
Parameters	Excess initial rate $a \in \mathbb{R}_{>0}$	Memory rate $k \in \mathbb{R}$	Baseline rate $b \in \mathbb{R}_{>0}$
Function	$f(t) = ae^{kt} + b$		

Used to fit the instantaneous rate of crawls. The function decreases when $k < 0$, increases when $k > 0$, and stay constant at $a + b$ when $k = 0$.

Exponential (decay) in increasing form with constant baseline

Support	$t \in [0, \infty)$		
Parameters	Limiting rate $a \in \mathbb{R}_{>0}$	Memory rate $k \in \mathbb{R}$	Baseline rate $b \in \mathbb{R}_{>0}$
Function	$f(t) = a\left(1 - e^{kt}\right) + b$		

Used to fit the instantaneous rate of sharp turns. The function increases when $k < 0$ and stay constant at b when $k = 0$. For $k > 0$ it decreases, eventually reaching negative $f(t)$ values, which depending on the context might be undefined (like a negative rate).

Exponential decay

Support	$t \in [0, \infty)$	
Parameters	Initial rate $a \in \mathbb{R}_{>0}$	Decaying rate $\lambda \in \mathbb{R}_{>0}$
Function	$f(t) = ae^{-\lambda t}$	

Used to fit the correlations of speed and angular velocity.

Double exponential decay

Support	$t \in [0, \infty)$			
Parameters	1^{st} Init. rate $a_1 \in \mathbb{R}_{>0}$	1^{st} Dec. rate $\lambda_1 \in \mathbb{R}$	2^{nd} Init. rate $a_2 \in \mathbb{R}_{>0}$	2^{nd} Dec. rate $\lambda_2 \in \mathbb{R}$
Function	$f(t) = a_1\, e^{-\lambda_1 t}\, a_2\, e^{-\lambda_2 t}$			

Used to fit the correlations of speed and angular velocity.

Triple exponential decay

Supp.	$t \in [0, \infty)$					
Param.	1^{st} Init. rate $a_1 \in \mathbb{R}_{>0}$	1^{st} Dec. rate $\lambda_1 \in \mathbb{R}$	2^{nd} Init. rate $a_2 \in \mathbb{R}_{>0}$	2^{nd} Dec. rate $\lambda_2 \in \mathbb{R}$	3^{rd} Init. rate $a_3 \in \mathbb{R}_{>0}$	3^{rd} Dec. rate $\lambda_3 \in \mathbb{R}$
Func.	$f(t) = a_1\, e^{-\lambda_1 t}\, a_2\, e^{-\lambda_2 t}\, a_3\, e^{-\lambda_3 t}$					

Used to fit the correlations of speed and angular velocity.

Power law decay

Support	$t \in [0, \infty)$	
Parameters	Initial rate $a \in \mathbb{R}_{>0}$	Decaying rate $\alpha \in \mathbb{R}_{>0}$
Function	$f(t) = at^{-\alpha}$	

Used to fit the correlations of speed and angular velocity.

Exponential sine decay

Support	$t \in [0, \infty)$			
Parameters	Init. amp. $a \in \mathbb{R}_{>0}$	Dec. rate $\lambda \in \mathbb{R}$	Ang. freq. $\phi, \in \mathbb{R}_{>0}$	Phase. ang. $\varphi \in (-\pi, \pi)$
Function	$f(t) = a\, e^{-\lambda\, t} \cdot \cos(\phi\, t - \varphi)$			

Used to fit the correlations of angular velocity.

Power law sine decay

Support	$t \in [0, \infty)$			
Parameters	Init. amp. $a \in \mathbb{R}_{>0}$	Dec. rate $\alpha \in \mathbb{R}$	Ang. freq. $\phi, \in \mathbb{R}_{>0}$	Phase. ang. $\varphi \in (-\pi, \pi)$
Function	$f(t) = a\, t^{-\alpha} \cdot \cos(\phi\, t - \varphi)$			

Used to fit the correlations of angular velocity.

Appendix C

Analysis of speed and angular velocity correlations

Here we present a thorough analysis of speed and angular velocity correlations. While this material logically fits between Chapter 5 and Chapter 6, we recommend the reader to skip it until the very end of Chapter 6, in order to maintain the line of reasoning established when it was first referenced. The aim of this analysis was twofold. First, to characterize the decay of speed and angular velocity correlations in *C. elegans* trajectories. Second, to assess the impact they have on the mean square displacement (MSD), given that spreading velocity correlations can generate anomalous diffusion through the so-called Joseph effect (Chapter 4), which in turn can be linked to speed and/or angular velocity correlations (Dieterich et al., 2008; Méndez et al., 2014; Chen et al., 2017; Aghion et al., 2021; Vilk et al., 2022; Meyer et al., 2022).

C.1 Computing correlations

Correlations of speed and angular velocity were computed assuming the following correlation function:

$$C(x,\tau) \;=\; \langle \big(x(t) - \bar{X}\big) \cdot \big(x(t+\tau) - \bar{X}\big) \rangle \;, \tag{C.1}$$

where $x(t)$ is the variable of interest at time t, $\bar{X}$ is the average value of x, and τ is the time span (or time window) between the pairs of values being multiplied.

Since speed in the trajectories under study tended to increase with time, we detrended its signal before computing the correlations. To do so, we roll averaged speed values using a point-wise running window of 600 seconds width, and

subtracted the averaged speed time-dependent signal from the raw signal (Figure C.1A). Then, we computed the correlations as indicated in Eq. (C.1), but assuming $\bar{X} = 0$, as we had already detrended the signal. Here, contrary to what we found when detrending the speed of crawls (Section 6.4.2 from Chapter 6) the t-test comparing the standard deviation of speed before and after detrending indicated there was a significant difference between the two (p-value $= 0.007$, $\alpha = 0.05$). However, the values of the fluctuations with and without applying the detrending were strongly correlated (Pearson correlation coefficient, $\rho = 0.93$), and the magnitudes did not "dramatically" change (Figure C.1C).

We computed the correlations of angular velocity using the same protocol. However, as the average angular velocity remained relatively constant over time, we detrended its signal by subtracting the average value of the entire time series (i.e., using a constant value). Consequently, the fluctuations with and without detrending were identical (Figures C.1D-E).

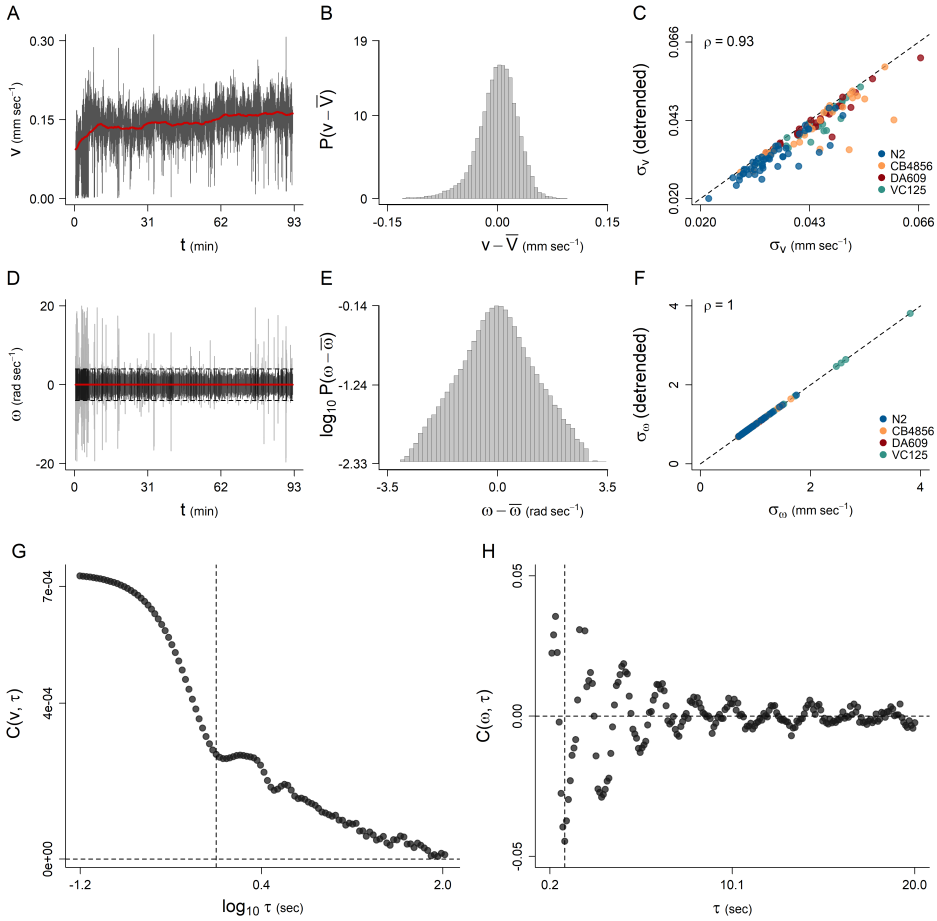


Figure C.1: Visual representation of the pipeline used to compute speed and angular velocity correlations for a representative N2 individual. (A) Temporal series of speed with the rolling average $\bar{V}$ depicted in red. (B) Histogram of detrended speed fluctuations. (C) Pearson correlation of the standard deviation of speed with and without detrending. (D) Temporal series of angular velocity with constant average $\bar{\omega}$ depicted in red. The gray region indicates data points above 99.5% or below 0.5% of the empirical distribution. These extremely large values were excluded from the analysis in order to better characterize the trend depicted in panel H. (E) Histogram of angular velocity fluctuations, or detrended angular velocity values. (F) Pearson correlation of the standard deviation of angular velocity with and without detrending. (G) Decaying trend of speed correlations. (H) Decaying trend of angular velocity correlations. In panels C, F, dashed lines indicate exact equality (1:1 line), while in panels G, H, they indicate the position of the zero coordinates.

C.2 Speed correlations

We characterized speed correlations by fitting a battery of time dependent functions (i.e., power law, exponential, double exponential and triple exponential) that explicitly accounted for fast and slow decaying trends (Figure C.2A). The fits were performed using non-linear least squares and model selection was based on AIC (Figure C.2B). All functions used can be consulted in Appendix B.2.

Based on model selection we concluded the best overall function to characterize the speed correlations was the triple exponential decay (Figures C.2A, B):

$$f(t) = a_1\, e^{-\lambda_1\, t} + a_2\, e^{-\lambda_2\, t} + a_3\, e^{-\lambda_3\, t}\ , \tag{C.2}$$

where the sum of a_1, a_2 and a_3 is the initial correlation; and λ_1, λ_2 and λ_3 are the three different decaying rates.

Strain averaged parameters of the triple exponential fitted to each individual are reported in Table C.1. Of note, despite the complexity observed in the decaying trend of speed correlations, they effectively approach zero in approximately 100 seconds. Suggesting speed correlations may not significantly influence the anomalous diffusion exponent of the mean square displacement in *C. elegans*.

Table C.1: Strain-averaged parameters of the triple exponential function fitted to speed correlations. Columns a_1, a_2 and a_3, report the initial correlation of the three exponential decays and are expressed in $\mathrm{mm}^2\,\mathrm{sec}^{-2}$. Columns λ_1, λ_2 and λ_3 indicate their corresponding rate and are expressed in sec^{-1}. Finally, N reports the number of individuals used to compute the averages.

Strain	a_1	λ_1	a_2	λ_2	a_3	λ_3	N
N2	0.00065	1.72	0.00031	0.331	0.00021	0.1671	48
CB4856	0.00115	1.37	0.00040	0.386	0.00033	0.5888	30
DA609	0.00108	1.09	0.00049	0.274	0.00046	0.2926	24
VC125	0.00122	1.21	0.00038	0.154	0.00030	0.2501	24

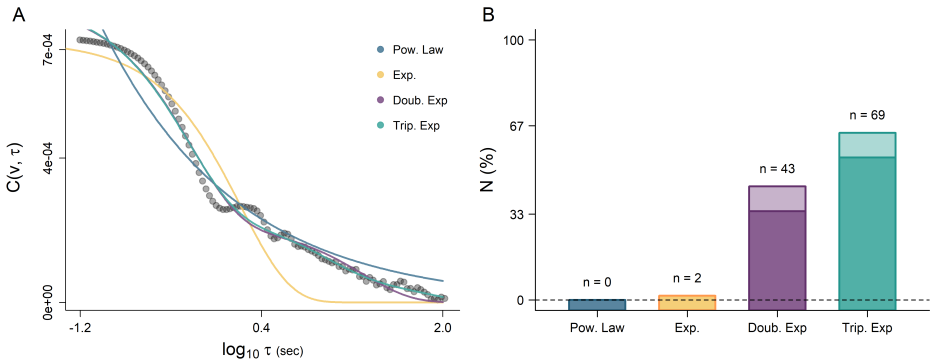


Figure C.2: Speed correlations fit and model selection. (A) Decaying trend in the speed correlations (expressed in $\mathrm{mm}^2\,\mathrm{sec}^{-2}$) of an N2 individual with corresponding model fits. The legend names indicate power law, exponential, double exponential, and triple exponential fits. (B) Model selection for the whole population (N = 126). Counts above bars indicate the number of times model selection was significant ($\Delta\mathrm{AIC} > 2$) and correspond to dark colors, while lighter colors indicate $\Delta\mathrm{AIC} < 2$ per model. Models with $\Delta\mathrm{AIC} < 2$ are not significantly different.

C.3 Angular velocity correlations

Angular velocity correlations showed a damping oscillatory wave, except in some cases in which they decayed monotonically. To characterize these trends we fitted a battery of time dependent functions (i.e., power law, power law sine, exponential, exponential sine) that explicitly accounted for both simple and oscillatory decays (Figure C.3A). All these functions were fitted using non-linear least squares and model selection was based on AIC (Figure C.3B). Their corresponding analytical expressions can be consulted in Appendix B.2.

Our results showed the best overall fit was the exponential sine function, followed by the power law sine. Given that the obtained curves with the two models where practically identical (Figure C.3A), we took the exponential sine function as a unique model, defined as:

$$f(t) = a\, e^{-\lambda\, t} \cdot \cos\left(\phi\, t - \varphi\right) \,, \tag{C.3}$$

where a is the initial amplitude; λ the decaying rate; ϕ the angular frequency and φ the initial phase angle, when $t = 0$. Strain averaged parameters of this model are reported in Table C.2.

Table C.2: Strain-averaged parameters of the exponential sine function fitted to angular velocity correlations. Column a indicates the initial correlation, expressed in $\text{rad}^2\,\text{sec}^{-2}$, while column λ indicates the decaying rate, expressed in sec^{-1}. In turn, ϕ represents the angular frequency and φ the phase angle, both expressed in rad. Finally, N show the number of individuals used to compute the averages.

Strain	a	λ	ϕ	φ	N
N2	0.130	0.421	3.01	0.38	48
CB4856	0.133	0.551	3.39	0.53	30
DA609	0.146	0.810	3.54	0.68	24
VC125	0.202	0.897	3.19	0.59	24

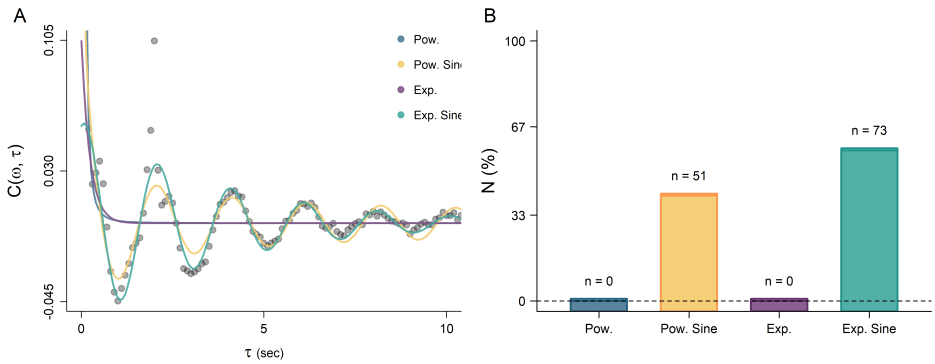


Figure C.3: Angular velocity correlations fit and model selection. (A) Representative example of an N2 individual exhibiting a damping oscillatory wave in angular velocity correlations, expressed in $\text{rad}^2\,\text{sec}^{-2}$. The legend names indicate models used to fit the trend: power law, power law sine, exponential and exponential sine. (B) Model selection for the whole population (N = 126). Counts above bars indicate the number of times model selection was significant ($\Delta\text{AIC} > 2$) and correspond to dark colors, while lighter colors indicate $\Delta\text{AIC} < 2$ per model. Models with $\Delta\text{AIC} < 2$ are not significantly different.

C.4 Are correlations in speed and angular velocity the main drivers of anomalous diffusion in *C. elegans*?

Based on the results obtained in Chapter 5 we hypothesized that anomalous diffusion in *C. elegans* does not emerge because of complex long correlations, but instead it comes from the temporal unfolding of sharp turns and crawls. To investigate this hypothesis, we built a shuffling protocol designed to explicitly assess the relationship between the correlations of speed and angular velocity and the mean square

displacement of the trajectories under study. Our protocol involves disrupting existing correlations at two different levels, within and between behavioral modes. We first disrupted within sharp turns and crawls velocities (module and angle), and second, we shuffled blindly all the trajectory (as if no behavioral modes would exist). Therefore, the temporal order of sharp turns and crawls is also lost.

Breaking the correlations of speed and angular velocity

Time series of speed and angular velocity were shuffled in a range of proportions, from 0% to 100%. Two protocols were used, one in which data was shuffled without disturbing the boundaries of sharp turns and crawls, and the other involving a blind shuffle of the whole data, loosing both the correlation properties but also the temporal order of crawls and turns. The resulting time series of shuffled speed and/or angular velocity were then used to generate artificial tracks, based on the fact that: (i) $x(t+dt) = x(t)+v\cos(\phi)dt$, and (ii) $y(t+dt) = y(t)+v\sin(\phi)dt$, where $x(t)$ and $y(t)$ are the coordinates of a given track at time t, v is the module of the velocity (i.e., the speed), and ϕ is the absolute direction (Figure C.4). Afterwards, we grouped them by shuffling proportion (i.e., from 0% to 100%) and shuffling protocol (i.e., by state or blind), and computed the mean square displacement curves.

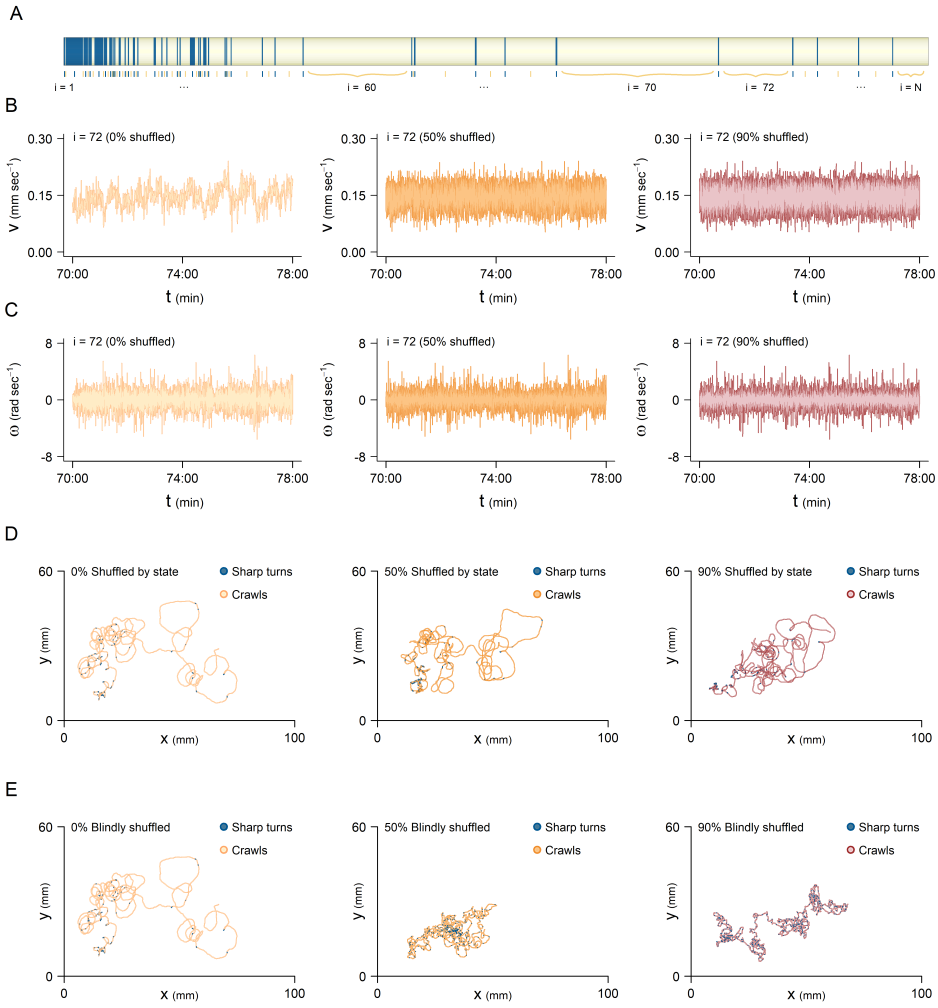


Figure C.4: Shuffling protocol of speed and angular velocity applied to a representative N2 individual. (A) Ethogram displaying indexed sharp turns (blue) and crawls (yellow). (B) Temporal series of speed associated with the crawl $i = 72$, shuffled at three different levels: 0% (no shuffling), 50% (half randomly shuffled), and 90% (ninety percent shuffled). (C) Equivalent time series for angular velocity. (D) Trajectories resulting from the shuffling protocol by behavioral state, preserving the temporal order (indexes) of behavioral modes depicted in panel A. (E) Trajectories resulting from a blind shuffle, which disrupts not only correlations but also the temporal order, and the whole concept of sharp turns and crawl behavioral states.

Computing the mean square displacement (MSD)

Following methods outlined in Chapter 4, we computed the mean square displacement as $\text{MSD}(t) = \frac{1}{N}\sum_{i=1}^{N} |\mathbf{r}_i(t) - \mathbf{r}_i(0)|^2$ using strain-based ensembles. In all cases, both the anomalous diffusion exponent α and the effective diffusion coefficient D_α were estimated by fitting a log-log piece-wise linear model with 4 regimes over the MSD curve, similar to Riahi et al. (2019). The fitting procedure was based on least-squares and was implemented using the R package *segmented* (Muggeo, 2003). Although a different number of regimes could be easily found, 4 was a good compromise between simplicity and accuracy to obtain reliable estimates of the anomalous diffusion exponent α, and at the same time, capturing the complexity of the MSD patterns in *C. elegans* (Figure C.5).

To assess whether the correlations of speed and angular velocity were responsible for anomalous diffusion via the Joseph effect (i.e., long-range spreading velocity correlations), we computed the ensemble time-averaged mean square displacement (TAMSD) and estimated the Joseph exponent as $J = \alpha/2$ (Meyer et al., 2022). In practice, we calculated the TAMSD for each individual and then averaged the resulting curves again by pooling individuals from the same strain. Subsequently, we fitted a log-log piece-wise linear model with three regimes and obtained the estimate of J from the slope of the final regime. For a more detailed description of the methodology used to estimate this exponent, go to Chapter 4. This pipeline was applied across a range of shuffling proportions from 0% to 100%, encompassing two types of shuffling protocols (by behavioral state or blind), and the two variables under study: speed and angular velocity.

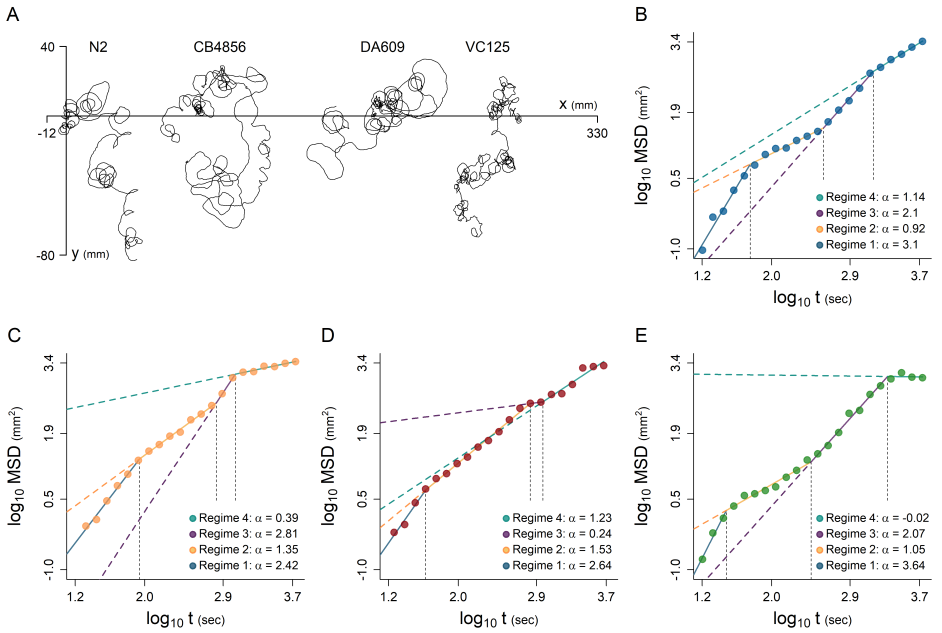


Figure C.5: Mean square displacement (MSD) analysis of *C. elegans* trajectories recorded during a relocation experiment without food. (A) Illustrative trajectories depicting the movement patterns of the wild-type strains N2 and CB4856, along with defective mutants DA609 and VC125. (B-E) MSD curves for each strain: N2 (blue), CB4856 (orange), DA609 (red), and VC125 (green). In all cases the MSD curve was parameterized by fitting a log-log piece-wise linear model with four distinct regimes, and the estimates of the anomalous diffusion exponent α are reported. Ensemble sizes used for MSD computation were N = 48 (N2), N = 30 (CB4856), N = 24 (DA609), and N = 24 (VC215).

Disentangling the effect of correlations on the anomalous diffusion exponent.

Our results demonstrate that anomalous diffusion in *C. elegans* is more contingent upon the behavioral states represented by sharp turns and crawls (Figure C.6) than on velocity correlations occurring within each of the behavioral states.

Specifically, we found that correlations in speed and angular velocity within behavioral states did not significantly impact on the anomalous diffusion scaling exponent α. This result implies that what governs the scaling exponent are not the correlations at these level. Conversely, shuffling across behavioral states, hence, altering both the boundaries and the internal structure of the behavioral states noticeably affects its value. This is true, for the angular velocity and at the large scales. With blind shuffling we observed a significant simplification of the MSD curve, for both speed and angular velocity, leading to a diminished or even absent presence of the distinct regimes observed in the original curve (see Figure C.6E).

These outcomes reaffirm two key points: firstly, that turning dynamics (i.e., angular velocity) play a pivotal role in *C. elegans* spreading behavior, and secondly, that the temporal unfolding of sharp turns and crawls is the main driver of these dynamics.

From a physical standpoint, our findings suggested that the temporal unfolding and modulation of sharp turns and crawls, and not velocity correlations, contributed to the emergence of long-range spreading velocity correlations (Joseph effect), as observed in Chapter 4. Crawls and sharp turns exhibit notable differences in average speed, turning variability, and duration. Thus, long correlations may naturally arise from the sustained movement dynamics in prolonged sharp turns or crawls and its change through time.

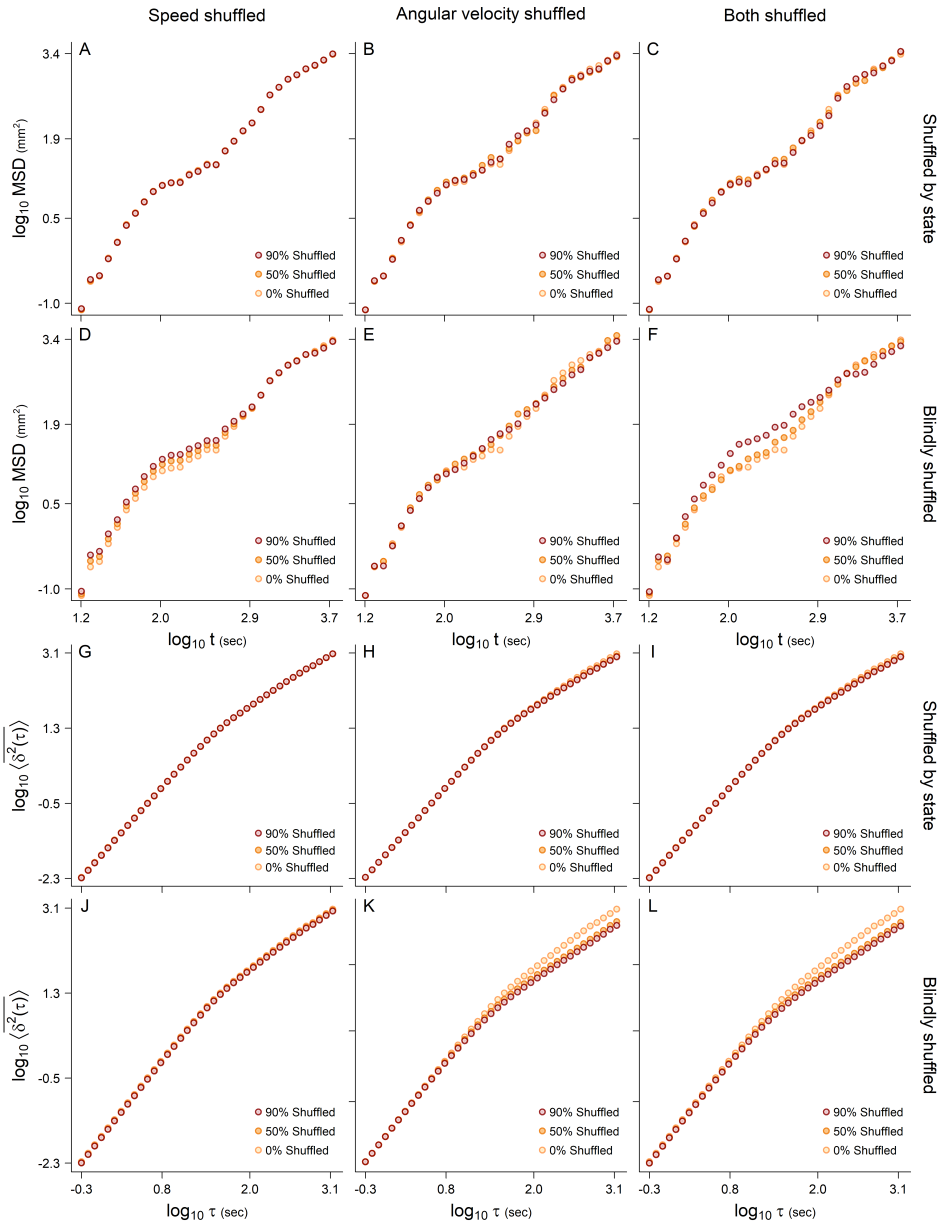


Figure C.6: Mean square displacement (MSD) (A-F), and ensemble time-averaged mean square displacement (TAMSD) (G-L), obtained by applying the shuffling protocol to trajectories of the N2 strain. Panels A-C display the MSD obtained from trajectories reconstructed after shuffling data without disturbing the boundaries of sharp turns and crawls. Panels D-F show the MSD curves of trajectories reconstructed after applying a blind shuffle. Panels G-L depict the TAMSD using the same shuffling techniques. In all cases, beige curves (0% shuffled) serve as a reference, representing trajectories derived from the original speed and angular velocity temporal series.

Appendix D

Methods and model validations

D.1 Validation of the behavioral annotation algorithm

In order to validate the classification algorithm developed in Chapter 5 we used a population of synthetic trajectories, with known behavioral labels, generated with the non-stationary model defined in Chapter 6. For each trajectory we applied the following pipeline: (i) compute the speed, (ii) fit the Weibull-normal distribution, (iii) run the behavioral annotation algorithm, and finally (iv) fit the distributions of angular velocity, from both sharp turns and crawls. This process was repeated 100 times for each individual under study, so that the total number of evaluated trajectories was $N = 12600$.

Since we knew the true behavioral labels of the model trajectories, we could assess the accuracy of the algorithm in terms of: (i) the similitude of obtained parameter space of speed and angular velocity for the synthetic and the empirical trajectories (compare Figures 5.9D, E and Figures D.1D, E), (ii) the labelling error of the annotation algorithm, namely, the number of wrongly annotated data points (Figure D.1F), (iii) the number of observed behaviors per trajectory (Figure D.1G), and (iv) the proportion of time devoted to sharp turns with respect to total amount of time (Figure D.1H).

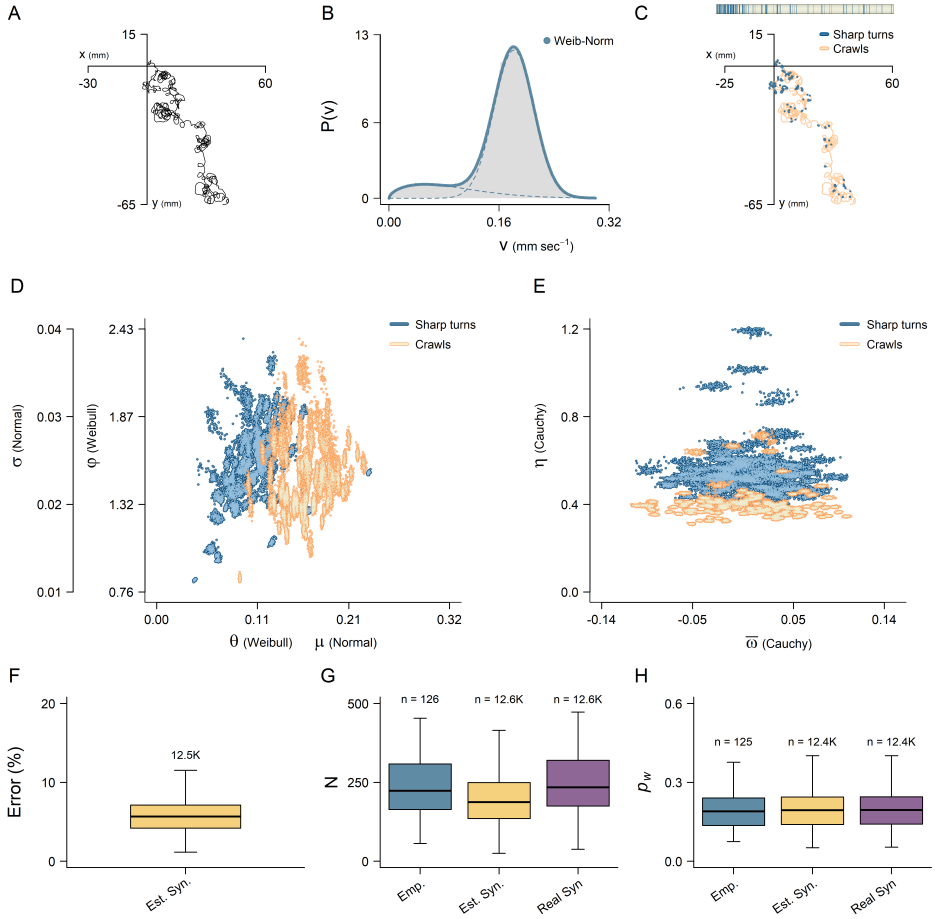


Figure D.1: Validation of the behavioral annotation algorithm. (A) Synthetic trajectory generated with the non-stationary model. (B) Weibull normal distribution fitted to instantaneous speed of the trajectory in A. (C) Resulting ethogram with the annotated trajectory. (D, E) Parameter space of speed and angular velocity distributions obtained with 100 realizations of each individual model, total N = 12600. (F) Labelling error of the behavioral annotation algorithm computed as the number of wrongly annotated data points. (G) Total number of sharp turns and crawls per trajectory. (H) Weibull probability weight of the speed distributions fitted to each individual realization. This quantity is equivalent to the proportion of time spent on sharp turns with respect to the total time. All results relative to synthetic trajectories correspond to the population shown in panels D-E.

D.2 Mesoscopic validation of movement models

Here we report the model validation at the mesoscopic scale, covering time spans from seconds to hours, for strains CB4856, DA609 and VC125.

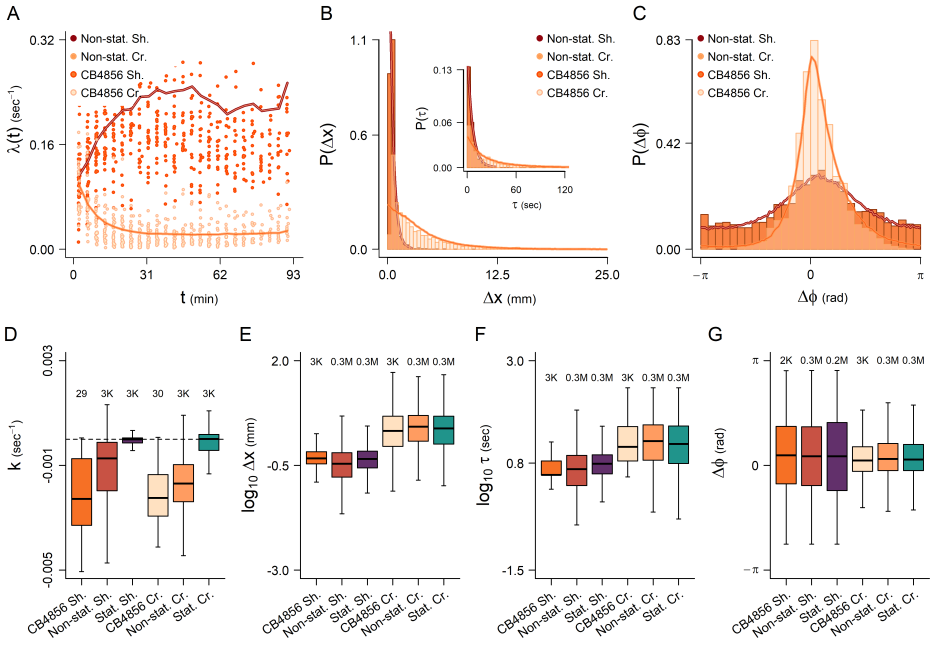


Figure D.2: Validation of the stationary and the non-stationary models for the strain CB4856 (30 individuals) at mesoscopic scale. Legend names indicate sharp turns (sh) and crawls (cr) of empirical data (CB4856), non-stationary, and stationary models (in this case also parameterized based on CB4856 individuals). (A) Temporal unfolding of sharp turns and crawls. (B) Distributions of net displacement and duration (inset). (C) Distribution of turning angle. (D) Boxplots comparing the memory rates of sharp turns (k_s) and crawls (k_c) in empirical and synthetic tracks. (E-G) Boxplots of the variables shown in panels B and C.

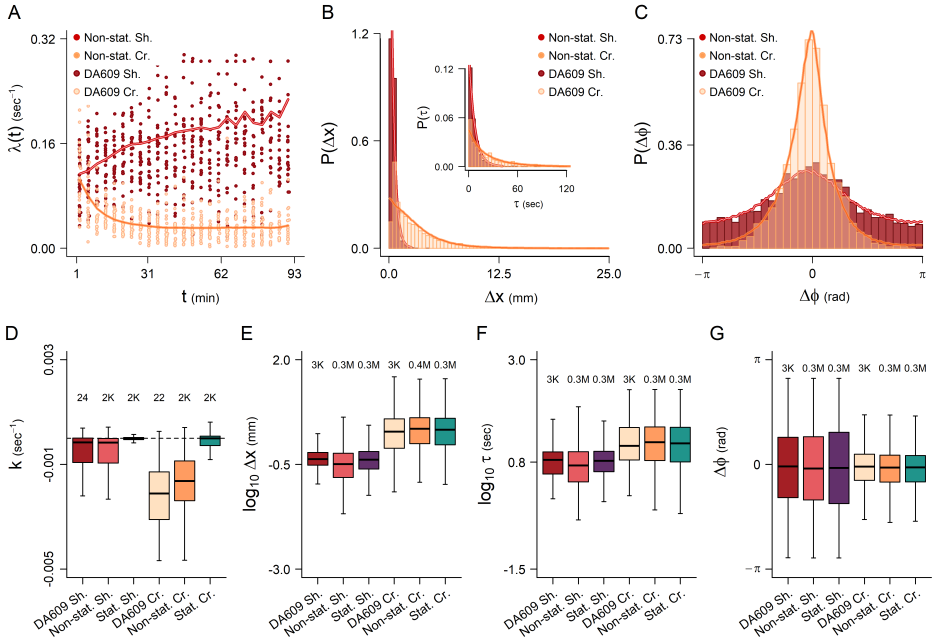


Figure D.3: Validation of the stationary and the non-stationary models for the strain DA609 (30 individuals) at mesoscopic scale. Legend names indicate sharp turns (sh) and crawls (cr) of empirical data (DA609), non-stationary, and stationary models (in this case also parameterized based on DA609 individuals). (A) Temporal unfolding of sharp turns and crawls. (B) Distributions of net displacement and duration (inset). (C) Distribution of turning angle. (D) Boxplots comparing the memory rates of sharp turns (k_s) and crawls (k_c) in empirical and synthetic tracks. (E-G) Boxplots of the variables shown in panels B and C.

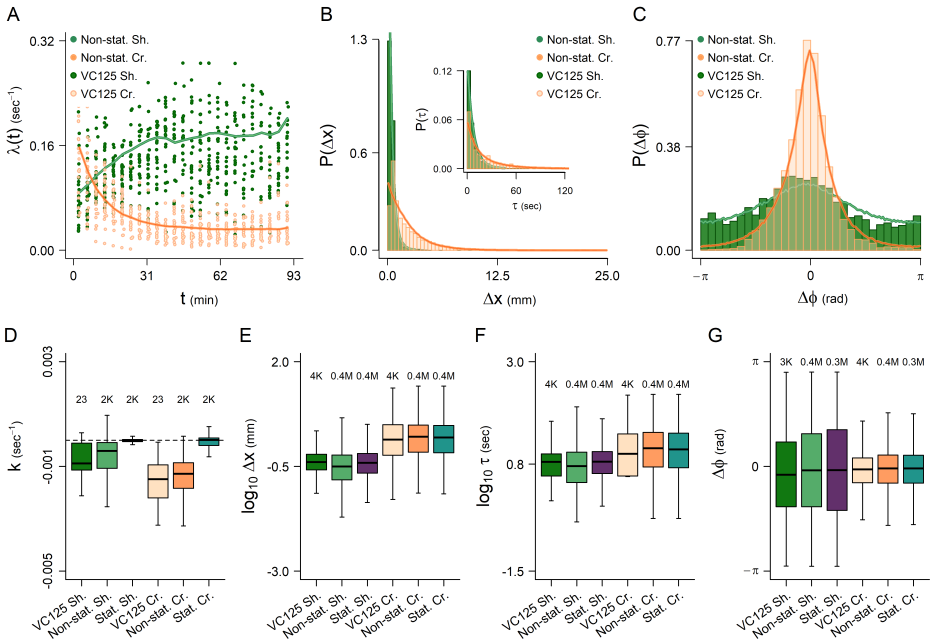


Figure D.4: Validation of the stationary and the non-stationary models for the strain VC125 (30 individuals) at mesoscopic scale. Legend names indicate sharp turns (sh) and crawls (cr) of empirical data (VC125), non-stationary, and stationary models (in this case also parameterized based on VC125 individuals). (A) Temporal unfolding of sharp turns and crawls. (B) Distributions of net displacement and duration (inset). (C) Distribution of turning angle. (D) Boxplots comparing the memory rates of sharp turns (k_s) and crawls (k_c) in empirical and synthetic tracks. (E-G) Boxplots of the variables shown in panels B and C.

D.3 Macroscopic validation of movement models

Replication of the empirical mean square displacement

Here we report the parameter values obtained for the mean square displacement (MSD) curves shown in Figure 6.13, which has been copied here to ease readability (Figure D.5). All values, including the transition time between regimes, were obtained by fitting a log-log piece-wise linear model to MSD data, and have been reported only for comparative purposes. Our objective was assessing the "power" of each model on reproducing the exact MSD curves of *C. elegans* given a certain number of simulations, in this case 1000 per empirical track. Therefore, reported values of both the anomalous diffusive exponent α, and the effective diffusion coefficient D_α, should not be considered proper estimates of the real parameter values, specially for the models. Due to the high number of fits (12 per strain) we did not

display the piece-wise linear models in the figure. However, we refer the reader to Figure 4.5 in order to get a sense of how they look like, and to understand how abnormal values like negative α or extremely large effective diffusion coefficients could appear.

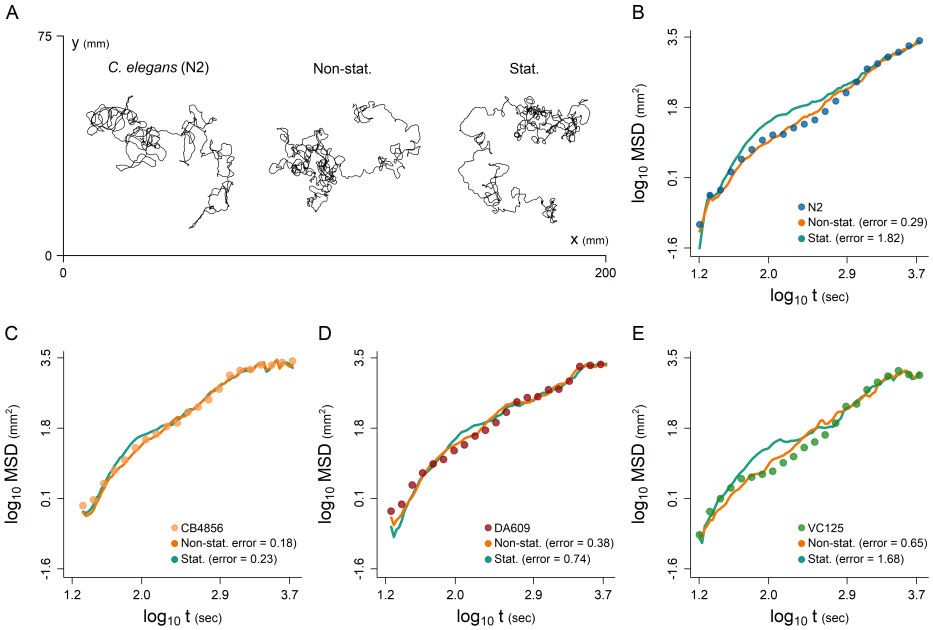


Figure D.5: Replication of the empirical mean square displacement (MSD) curves. (A) Example trajectory of an N2 individual with its corresponding artificial copies, generated with the stationary and the non-stationary models. In both cases, these copies were obtained from a population of 1000 trajectories by minimizing the sum of squares between the synthetic and the empirical point-wise distances to the origin. (B-E) MSD curves for the different strains i.e., N2, CB4856, DA609 and VC125, and their replicates. The error reported in panels B-E equals the sum of squares between empirical MSD and synthetic MSD curves.

Table D.1: Replication of the mean square displacement (MSD) of the strains N2 and CB4856 with the non-stationary and stationary models. In the column Regime, time is expressed in seconds while in the column Duration it is expressed in hours, minutes and seconds. The MSD, measured at the end of each regime, is expressed in mm^2, and the effective diffusion coefficient D_α in $mm^2\,s^{-1}$. The anomalous diffusion exponent α is dimensionless and is reported with the standard error of the estimate.

Strain/Model	Regime	MSD	$\alpha \pm \sigma_\alpha$	D_α	Duration
N2	$1599 < t \leq 5405$	2628.0	1.09 ± 0.077	$5.8 \cdot 10^{-2}$	01:03:23
Non-stat.	$2651 < t \leq 5405$	2449.9	1.06 ± 0.167	$6.5 \cdot 10^{-2}$	00:45:54
Stat.	$504 < t \leq 5405$	2650.3	1.26 ± 0.034	$1.4 \cdot 10^{-2}$	01:21:41
N2	$365 < t \leq 1599$	699.3	2.07 ± 0.061	$4.2 \cdot 10^{-5}$	00:20:34
Non-stat.	$362 < t \leq 2651$	1147.6	1.51 ± 0.039	$2.0 \cdot 10^{-2}$	00:38:09
Stat.	$76 < t \leq 504$	134.2	0.97 ± 0.050	$8.2 \cdot 10^{-2}$	00:07:08
N2	$66 < t \leq 365$	33.0	0.84 ± 0.047	$5.8 \cdot 10^{-2}$	00:04:59
Non-stat.	$53 < t \leq 362$	56.9	1.40 ± 0.043	$3.6 \cdot 10^{-3}$	00:05:09
Stat.	$21 < t \leq 76$	21.5	3.11 ± 0.093	$7.8 \cdot 10^{-6}$	00:00:55
N2	$16 < t \leq 66$	7.8	2.82 ± 0.061	$1.4 \cdot 10^{-5}$	00:00:50
Non-stat.	$16 < t \leq 53$	3.8	2.99 ± 0.080	$6.6 \cdot 10^{-6}$	00:00:37
Stat.	$16 < t \leq 21$	0.4	8.29 ± 0.645	$9.8 \cdot 10^{-13}$	00:00:05
CB4856	$1109 < t \leq 5405$	2605.7	0.36 ± 0.048	29.36^*	01:11:36
Non-stat.	$1566 < t \leq 5405$	2257.1	0.16 ± 0.081	145.7^*	01:03:59
Stat.	$1423 < t \leq 5405$	2419.6	0.31 ± 0.064	43.03^*	01:06:22
CB4856	$615 < t \leq 1109$	1471.9	2.58 ± 0.198	$5.2 \cdot 10^{-6}$	00:08:14
Non-stat.	$94 < t \leq 1566$	1857.1	1.58 ± 0.023	$4.1 \cdot 10^{-3}$	00:24:32
Stat.	$239 < t \leq 1423$	1605.1	1.61 ± 0.043	$3.4 \cdot 10^{-3}$	00:19:44
CB4856	$84 < t \leq 615$	322.6	1.36 ± 0.034	$1.3 \cdot 10^{-2}$	00:08:52
Non-stat.	$49 < t \leq 94$	21.8	2.17 ± 0.230	$2.9 \cdot 10^{-4}$	00:00:46
Stat.	$77 < t \leq 239$	90.8	0.92 ± 0.089	$1.5 \cdot 10^{-1}$	00:02:42
CB4856	$21 < t \leq 84$	21.5	2.48 ± 0.060	$9.3 \cdot 10^{-5}$	00:01:02
Non-stat.	$21 < t \leq 49$	5.2	3.20 ± 0.132	$5.3 \cdot 10^{-6}$	00:00:27
Stat.	$21 < t \leq 77$	32.0	3.34 ± 0.069	$4.1 \cdot 10^{-6}$	00:00:56

Note*: these abnormally high effective diffusion coefficients D_α appear in association with abnormally low anomalous diffusion exponents α (see Figures 4.5C, E for an illustrative example).

Table D.2: Replication of the mean square displacement (MSD) of the strain DA609 and VC125 with the non-stationary and stationary models. In the column Regime, time is expressed in seconds while in the column Duration it is expressed in hours, minutes and seconds. The MSD, measured at the end of each regime, is expressed in mm^2, and the effective diffusion coefficient D_α in $mm^2\,s^{-1}$. The anomalous diffusion exponent α is dimensionless and is reported with the standard error of the estimate.

Strain/Model	Regime	MSD	$\alpha \pm \sigma_\alpha$	D_α	Duration
DA609	$3694 < t \leq 5405$	2206.5	0.25 ± 0.525	61.69^*	00:28:31
Non-stat.	$665 < t \leq 5405$	3046.5	1.17 ± 0.042	$3.3 \cdot 10^{-2}$	01:19:00
Stat.	$107 < t \leq 5405$	2674.7	1.06 ± 0.016	$7.5 \cdot 10^{-2}$	01:28:18
DA609	$512 < t \leq 3694$	2002.5	1.11 ± 0.051	$5.5 \cdot 10^{-2}$	00:53:02
Non-stat.	$457 < t \leq 665$	262.7	-0.03 ± 0.660	79.87^*	00:03:28
Stat.	$54 < t \leq 107$	42.2	2.58 ± 0.224	$8.0 \cdot 10^{-4}$	00:00:53
DA609	$38 < t \leq 512$	223.7	1.51 ± 0.033	$4.5 \cdot 10^{-3}$	00:07:54
Non-stat.	$55 < t \leq 457$	265.7	1.58 ± 0.042	$4.3 \cdot 10^{-3}$	00:06:41
Stat.	$41 < t \leq 54$	7.1	1.97 ± 0.910	$6.8 \cdot 10^{-4}$	00:00:13
DA609	$18 < t \leq 38$	4.4	3.22 ± 0.210	$9.0 \cdot 10^{-5}$	00:00:20
Non-stat.	$18 < t \leq 55$	9.5	3.29 ± 0.113	$4.4 \cdot 10^{-6}$	00:00:37
Stat.	$18 < t \leq 41$	4.2	4.18 ± 0.170	$1.9 \cdot 10^{-7}$	00:00:23
VC125	$2486 < t \leq 5405$	1085.2	-0.40 ± 0.173	$10^{3.93*}$	00:48:39
Non-stat.	$3465 < t \leq 5405$	1043.5	-0.96 ± 0.316	$10^{6.01*}$	00:32:20
Stat.	$2116 < t \leq 5405$	1303.4	0.23 ± 0.151	$10^{1.66*}$	00:54:49
VC125	$197 < t \leq 2486$	1483.0	1.93 ± 0.033	$1.0 \cdot 10^{-4}$	00:38:10
Non-stat.	$344 < t \leq 3465$	1602.4	1.42 ± 0.031	$3.8 \cdot 10^{-3}$	00:52:01
Stat.	$512 < t \leq 2116$	1053.0	1.89 ± 0.088	$1.4 \cdot 10^{-3}$	00:26:35
VC125	$35 < t \leq 197$	11.0	0.91 ± 0.057	$2.2 \cdot 10^{-2}$	00:02:42
Non-stat.	$40 < t \leq 344$	60.7	1.62 ± 0.034	$1.2 \cdot 10^{-3}$	00:05:04
Stat.	$66 < t \leq 521$	75.0	0.77 ± 0.048	$1.5 \cdot 10^{-1}$	00:07:36
VC125	$16 < t \leq 35$	2.3	3.23 ± 0.173	$6.0 \cdot 10^{-6}$	00:00:19
Non-stat.	$16 < t \leq 40$	1.9	2.70 ± 0.111	$2.2 \cdot 10^{-5}$	00:00:24
Stat.	$16 < t \leq 66$	15.3	3.17 ± 0.082	$6.5 \cdot 10^{-6}$	00:00:50

Note*: these abnormally high effective diffusion coefficients D_α appear in association with abnormally low anomalous diffusion exponents α (see Figures 4.5C, E for an illustrative examples).

Individual versus average variability of the mean square displacement

Here we report the comparison of the mean square displacement (MSD) curves of the average individuals versus the particular individuals of the strains CB4856, DA609 and VC125, for the stationary and the non-stationary models.

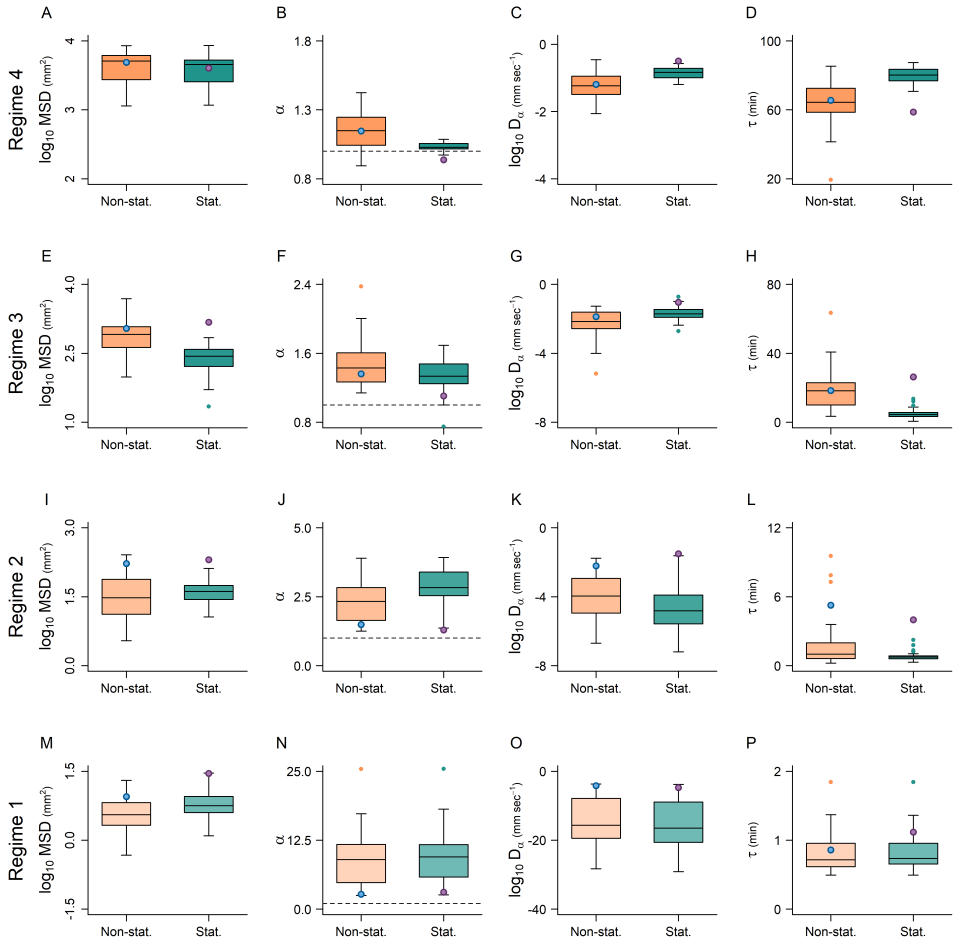


Figure D.6: Mean square displacement (MSD) of the average individual *versus* the particular individuals of the strain CB4856, modelled with the non-stationary and the stationary models. All boxplots were constructed by pooling individual-based estimates, obtained from ensembles of 1000 trajectories (total ensemble size N = 48000 per model). Blue and purple dots where obtained from the average individual (ensemble size N = 1000 per model). From left to right: MSD measured at the end of the regime, anomalous diffusion exponent with $\alpha = 1$ denoted as a dashed line, effective diffusion coefficient D_α, and duration of the regime. (A-D) Fourth (long-term) regime. (E-H) Third (midterm) regime. (I-L) Second (early transient) regime. (M-P) First (initial momentum) regime.

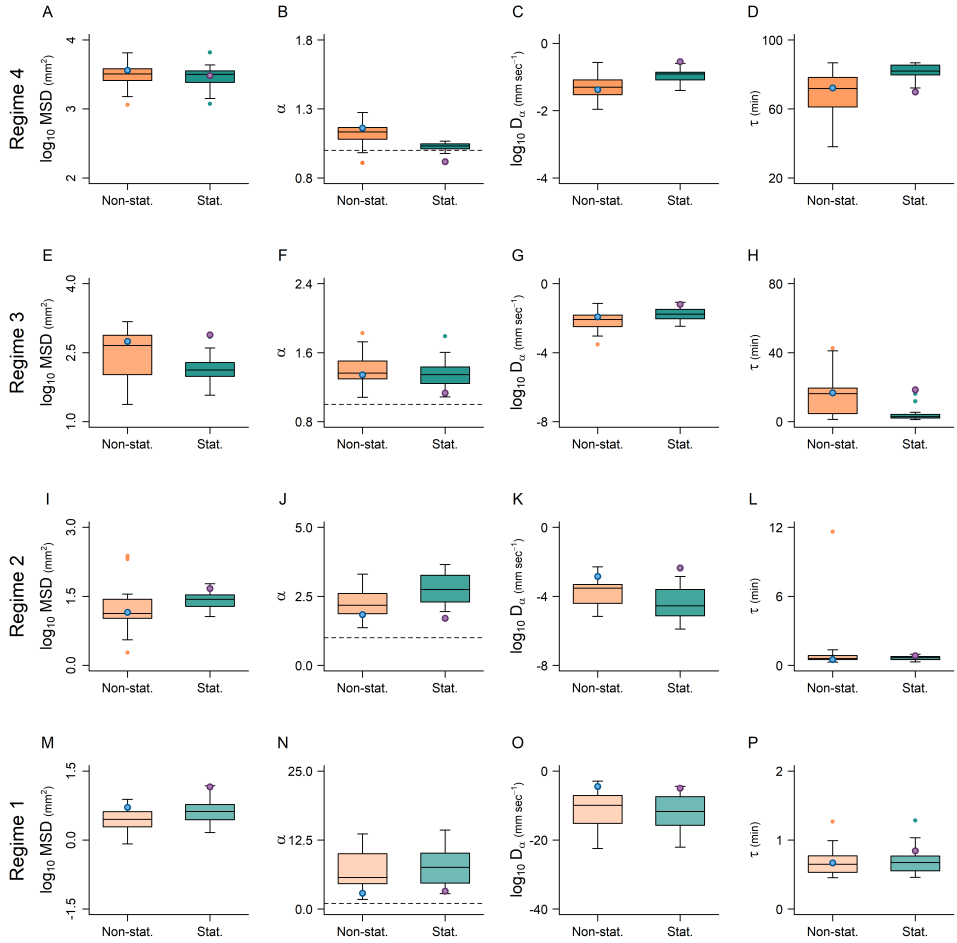


Figure D.7: Mean square displacement (MSD) comparison between the average individual and the particular individuals of the strain DA609, modelled with the non-stationary and the stationary models. All boxplots were constructed by pooling individual-based estimates, obtained from ensembles of 1000 trajectories (total ensemble size N = 48000 per model). Similarly, the big blue and purple dots where obtained from the average individual (total ensemble size N = 1000 per model). From left to right, each row of panels show: MSD measured at the end of the regime, anomalous diffusion exponent with $\alpha = 1$ denoted as a dashed line, effective diffusion coefficient D_α, and duration of the regime. (A-D) Characterization of the fourth (long-term) regime. (E-H) Characterization of the third (midterm) regime. (I-L) Characterization of second (early transient) regime. (M-P) Characterization of the first (initial momentum) regime.

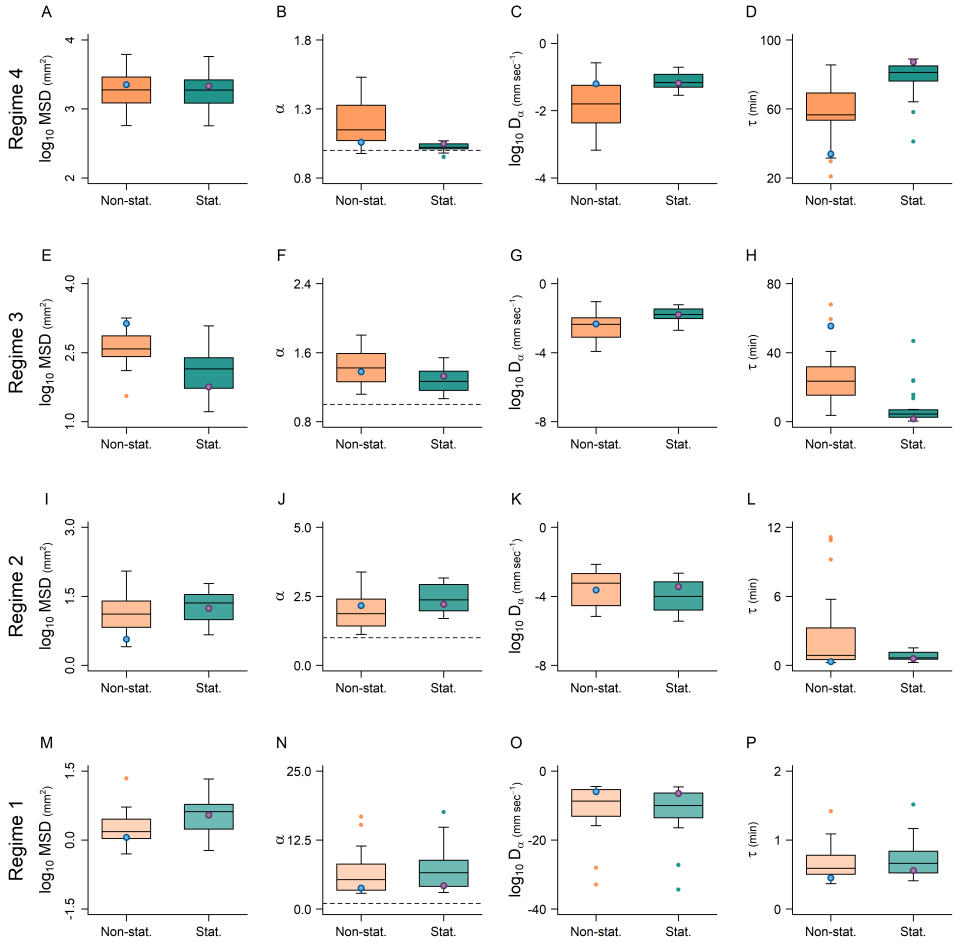


Figure D.8: Mean square displacement (MSD) comparison between the average individual and the particular individuals of the strain VC125, modelled with the non-stationary and the stationary models. All boxplots were constructed by pooling individual-based estimates, obtained from ensembles of 1000 trajectories (total ensemble size N = 48000 per model). Similarly, the big blue and purple dots where obtained from the average individual (total ensemble size N = 1000 per model). From left to right, each row of panels show: MSD measured at the end of the regime, anomalous diffusion exponent with $\alpha = 1$ denoted as a dashed line, effective diffusion coefficient D_α, and duration of the regime. (A-D) Characterization of the fourth (long-term) regime. (E-H) Characterization of the third (midterm) regime. (I-L) Characterization of second (early transient) regime. (M-P) Characterization of the first (initial momentum) regime.

D.4 Memory heatmaps of the mean square displacement

Raw memory heatmaps

Here we show the raw (non-smoothed) results of the simulations conducted to characterize the behavior of the MSD in function of the memory rates k_c and k_s.

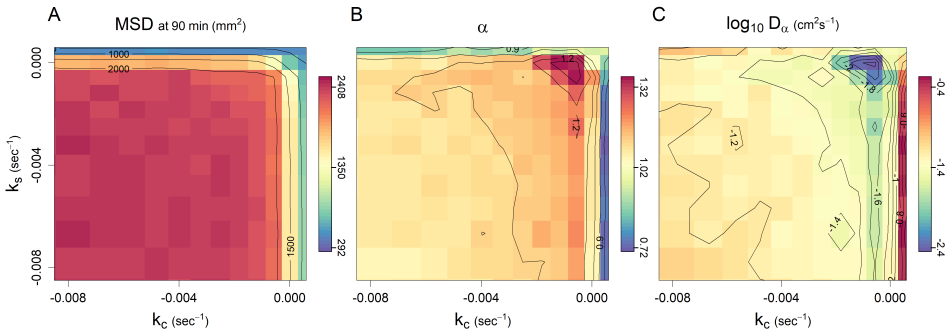


Figure D.9: Mean square displacement (MSD) of the non-stationary model as a function of the memory rates k_c and k_s. Other model parameters were fixed at the values of the average N2 individual, as reported in Tables 6.1-6.2, 6.4. (A) Overall spatial coverage in 90 minutes. (B) Anomalous diffusion exponent α. (C) Effective diffusion coefficient D_α. Each heatmap consists of a square matrix with $15 \times 15 = 225$ combinations of k_c and k_s values, and each pair of k_c and k_s (i.e., the "pixels") contains 10 independent MSD curves with 1000 trajectories, from which estimates of α and D_α were extracted. Then, the average values of: (i) the MSD measured at the end of the simulation (panel A), (ii) the anomalous diffusion exponent α (panel B), and (iii) the effective diffusion coefficient D_α (panel C), are depicted in the heatmaps. We implemented such a high-throughput protocol aiming to reduce the spurious variability associated with random number generators. In all cases, estimates of both α and D_α were obtained at the long-term limit (last observed regime) using a log-log piece-wise linear model with four regimes, as detailed in Section 6.2.

Memory heatmaps controlling $\lambda_c(t)$ and $\lambda_s(t)$

We characterized the behavior of the mean square displacement as a function of the memory rates k_c and k_s, while controlling the number of executed sharp turns throughout the simulation length (90 minutes). To achieve this, we maintained equal average rates of crawls $\bar{\lambda}_c = \langle \lambda_c(t) \rangle$, and sharp turns $\bar{\lambda}_s = \langle \lambda_s(t) \rangle$ for the different combinations of k_c and k_s explored (axes in Figures D.4A-C). These average rates were derived using the mean value theorem for integrals, which states the mean value of a continuous function over an interval $[a, b]$ is given by $\frac{1}{b-a} \int_a^b f(x)dx$. Therefore, we could obtain the average rate of crawls and sharp turns by numerical integration of Eq. (6.10) (crawls) and Eq. (6.11) (sharp turns). Specifically, we computed the values of $\bar{\lambda}_c$ and $\bar{\lambda}_s$ associated with the average N2 individual. We

then developed an optimization routine to determine the values of a_c and a_s that yield targeted $\bar{\lambda}_c$ and $\bar{\lambda}_s$ given arbitrary k_c and k_s. Of note, we could not evaluate positive k_s in this analysis because $k_s > 0$ make the integral of $a_s \left(1 - e^{k_s t}\right) + b_s$ divergent. To overcome this issue we set $k_s = 0$ and raised the value of b_s so that we could at least reach the targeted average rate of sharp turns. Consequently, heatmap values relative to $k_s > 0$ can be overlooked.

Through this exercise, we demonstrated that the anomalous diffusion exponent α is maximized as a result of optimal non-stationary trends in the rate of sharp turns and crawls. Specifically, the key element for generating superdiffusion is the temporal order in the rate (or duration) of sharp turns and crawls rather than overall number of executed sharp turns. Transitioning from long sharp turns and short crawls to short sharp turns and long crawls results in superdiffusion, while the reverse order leads to subdiffusion. Additionally, we observed that when fixing the average rate of sharp turns and crawls, the significance of the initial condition, defined by $a_c + b_c$ and b_s, diminishes. As shown in Figure D.4E, in such circumstances, a rapid memory decay (i.e., $k_c \ll 0$ and $k_s \ll 0$) is equivalent to a persistent or stationary memory (i.e., $k_c = 0$ and $k_s = 0$).

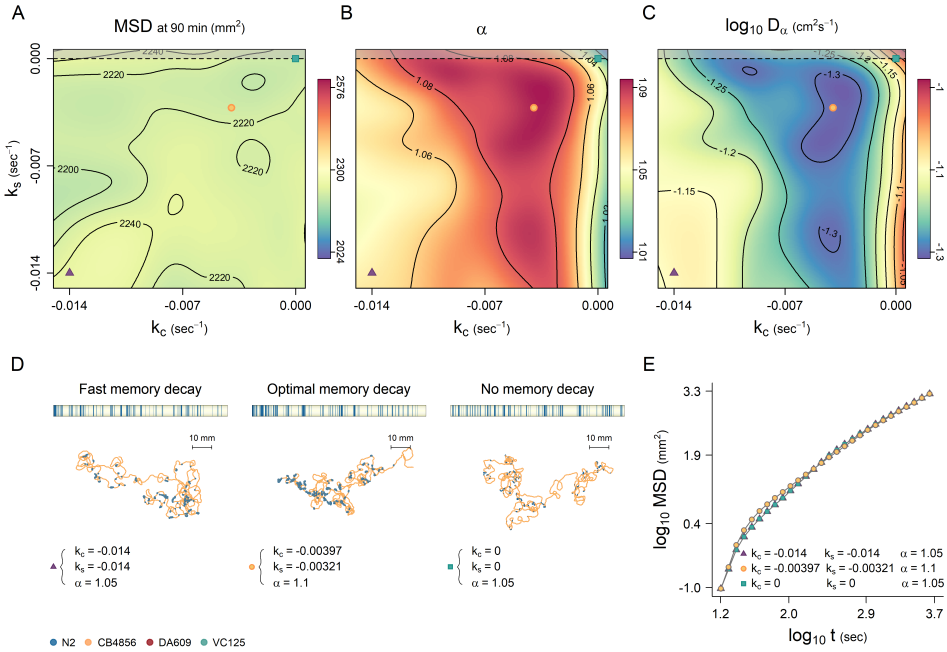


Figure D.10: Memory heatmaps of the mean square displacement (MSD) for fixed average rates of crawls and sharp turns. All model parameters except k_c and k_s were fixed at the values of the average N2 individual, as reported in Tables 6.1-6.2, 6.4. (A) Overall spatial coverage in 90 minutes. (B) Anomalous diffusion exponent α. (C) Effective diffusion coefficient D_α. (D) Examples of annotated trajectories with sharp turns (blue) and crawls (yellow) at relevant points of the parameter space. Note the ethograms show the same number of sharp turns and crawls in average, but their temporal trend is different, specially in the optimal memory decay. (E) MSD curves of k_c and k_s values of the tracks shown in panel D (ensemble size N = 5000). All heatmaps are smoothed representations of the raw simulations (Figure D.4), which comprised $15 \times 15 = 225$ combinations of different k_c and k_s values. In its turn, each pair of k_c and k_s contains a whole MSD curve, equivalent to those depicted in panel E, computed with an ensemble of 1000 trajectories from which estimates of α and D_α were extracted. In all cases, both α and D_α were fitted at the long-term limit (last observed regime) using a log-log piece-wise linear model with four regimes. Shaded regions ($k_s > 0$) where depicted just for illustrative purposes and might be considered artificial.

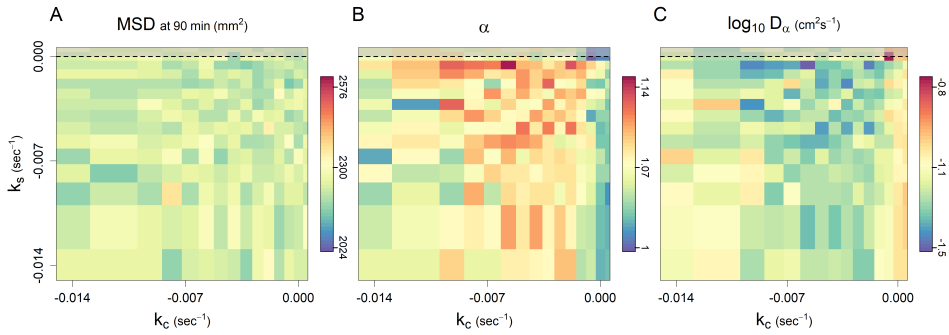


Figure D.11: Raw memory heatmaps of the mean square displacement (MSD) for fixed average rates of crawls and sharp turns. All model parameters except k_c and k_s were fixed at the values of the average N2 individual, as reported in Tables 6.1-6.2, 6.4. (A) Overall spatial coverage in 90 minutes. (B) Anomalous diffusion exponent α. (C) Effective diffusion coefficient D_α. Each heatmap consists of a square matrix with $15 \times 15 = 225$ combinations of k_c and k_s values, and each pair of k_c and k_s (i.e., the "pixels") contains 5 independent MSD curves with 1000 trajectories, from which estimates of α and D_α were extracted. Then, the average values of: (i) the MSD measured at the end of the simulation (panel A), (ii) the anomalous diffusion exponent α (panel B), and (iii) the effective diffusion coefficient D_α (panel C), are depicted in the heatmaps. In all cases, estimates of both α and D_α were obtained at the long-term limit (last observed regime) using a log-log piece-wise linear model with four regimes. Shaded regions ($k_s > 0$) where depicted just for illustrative purposes and might be considered artificial.

Bibliography

Aghion, E., Meyer, P. G., Adlakha, V., Kantz, H., and Bassler, K. E. (2021). Moses, Noah and Joseph effects in Lévy walks. *New Journal of Physics*, 23(2):023002.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6):716–723.

Ardiel, E. L. and Rankin, C. H. (2010). An elegant mind: Learning and memory in *Caenorhabditis elegans*.

Ariel, G., Rabani, A., Benisty, S., Partridge, J. D., Harshey, R. M., and Be'Er, A. (2015). Swarming bacteria migrate by Lévy walk. *Nature Communications*, 6:8396.

Bartumeus, F. (2009). Behavioral intermittence, Lévy patterns, and randomness in animal movement. *Oikos*, 118(4):488–494.

Bartumeus, F., Campos, D., Ryu, W. S., Lloret-Cabot, R., Méndez, V., and Catalan, J. (2016). Foraging success under uncertainty: search tradeoffs and optimal space use. *Ecology Letters*, 19(11):1299–1313.

Bartumeus, F., Catalan, J., Fulco, U. L., Lyra, M. L., and Viswanathan, G. M. (2002). Optimizing the encounter rate in biological interactions: Lévy versus brownian strategies. *Physical Review Letters*, 88(9).

Bartumeus, F., Giuggioli, L., Louzao, M., Bretagnolle, V., Oro, D., and Levin, S. A. (2010). Fishery discards impact on seabird movement patterns at regional scales. *Current Biology*, 20(3):215–222.

Bartumeus, F. and Levin, S. A. (2008). Fractal reorientation clocks: linking animal behavior to statistical patterns of search. *Proceedings of the National Academy of Sciences of the United States of America*, 105(49):19072–19077.

Bartumeus, F., Luz, M. G. D., Viswanathan, G. M., and Catalan, J. (2005). Animal search strategies: a quantitative random-walk analysis. *Ecology*, 86(11):3078–3087.

Bartumeus, F., Peters, F., Pueyo, S., Marrasé, C., and Catalan, J. (2003). Helical lévy walks: Adjusting searching statistics to resource availability in microzooplankton. *Proceedings of the National Academy of Sciences of the United States of America*, 100(22):12771–12775.

Bartumeus, F., Raposo, E. P., Viswanathan, G. M., and Luz, M. G. D. (2014). Stochastic optimal foraging: tuning intensive and extensive dynamics in random searches. *PLoS ONE*, 9(9):e106373.

Bazazi, S., Bartumeus, F., Hale, J. J., and Couzin, I. D. (2012). Intermittent motion in desert locusts: behavioural complexity in simple environments. *PLoS Computational Biology*, 8(5):e1002498.

Bell, W. J. (1991). *Searching behaviour: the behavioural ecology of finding resources*. Chapman and Hall, London, UK.

Benhamou, S. (2007). How many animals really do the Lévy walk? *Ecology*, 88(8):1962–1969.

Benhamou, S. and Collet, J. (2015). Ultimate failure of the Lévy foraging hypothesis: Two-scale searching strategies outperform scale-free ones even when prey are scarce and cryptic. *Journal of Theoretical Biology*, 387:221–227.

Berg, H. C. (1983). *Random walks in biology*. Princeton Unversity Press, New Jersey.

Berg, H. C. and Berry, R. M. (2004). *E. coli in motion*. Springer, New York, NY.

Bolker, B. and R Development Core Team (2023). *bbmle: Tools for General Maximum Likelihood Estimation*. R package version 1.0.25.1.

Brenner, S. (1974). The genetics of *Caenorhabditis elegans*. *Genetics*, 77(1):71–94.

Brown, A. E., Yemini, E. I., Grundy, L. J., Jucikas, T., and Schafer, W. R. (2013). A dictionary of behavioral motifs reveals clusters of genes affecting *Caenorhabditis elegans* locomotion. *Proceedings of the National Academy of Sciences of the United States of America*, 110(2):791–796.

Burnham, K. and Anderson, D. (2002). *Model selection and multimodel inference: a practical information-theoretic approach*. Springer, New York, NY.

Byerly, L., Cassada, R., and Russell, R. (1976). The life cycle of the nematode *Caenorhabditis elegans*: I. Wild-type growth and reproduction. *Developmental Biology*, 51(1):23–33.

Calhoun, A. J., Chalasani, S. H., and Sharpee, T. O. (2014). Maximally informative foraging by *Caenorhabditis elegans*. *eLife*, 3:e04220.

Campos, D., Bartumeus, F., Méndez, V., and Espadaler, X. (2014). Reorientation patterns in central-place foraging: internal clocks and klinokinesis. *Journal of the Royal Society Interface*, 11(90):20130859.

Cermak, N., Yu, S. K., Clark, R., Huang, Y. C., Baskoylu, S. N., and Flavell, S. W. (2020). Whole-organism behavioral profiling reveals a role for dopamine in statedependent motor program coupling in *C. elegans*. *eLife*, 9:e57093.

Chen, L., Bassler, K. E., McCauley, J. L., and Gunaratne, G. H. (2017). Anomalous scaling of stochastic processes and the Moses effect. *Physical Review E*, 95(4):042141.

Chupeau, M., Bénichou, O., and Voituriez, R. (2015). Cover times of random searches. *Nature Physics*, 11(10):844–847.

Cole, B. J. (1995). Fractal time in animal behaviour: the movement activity of *Drosophila*. *Animal Behaviour*, 50(5):1317–1324.

Conners, M. G., Michelot, T., Heywood, E. I., Orben, R. A., Phillips, R. A., Vyssotski, A. L., Shaffer, S. A., and Thorne, L. H. (2021). Hidden Markov models identify major movement modes in accelerometer and magnetometer data from four albatross species. *Movement Ecology*, 9(1):7.

Costa, W. S., der Auwera, P. V., Glock, C., Liewald, J. F., Bach, M., Schüler, C., Wabnig, S., Oranth, A., Masurat, F., Bringmann, H., Schoofs, L., Stelzer, E. H., Fischer, S. C., and Gottschalk, A. (2019). A GABAergic and peptidergic sleep neuron as a locomotion stop neuron with compartmentalized Ca^{2+} dynamics. *Nature Communications*, 10(1):4095.

Croll, N. A. (1975). Components and patterns in the behaviour of the nematode *Caenorhabditis elegans*. *Journal of Zoology*, 176(2):159–176.

Cáceres, M. O. and Budini, A. A. (1997). The generalized Ornstein-Uhlenbeck process. *Journal of Physics A: Mathematical and General*, 30(24):8427–8444.

Denny, M. and Gaines, S. (2002). *Chance in biology: Using probability to explore nature*. Princeton University Press, New Jersey, N.

Dieterich, P., Klages, R., Preuss, R., and Schwab, A. (2008). Anomalous dynamics of cell migration. *Proceedings of the National Academy of Sciences of the United States of America*, 105(2):459–463.

Einstein, A. (1905). Über die von der molekularkinetischen theorie der wärme geforderte bewegung von in ruhenden flüssigkeiten suspendierten teilchen. *Annalen der Physik*, 322(8):549–560.

Feder, J. (1988). *Fractals*. Physics of solids and liquids. Plenum Press, New York, NY.

Feller, W. (1935). Über den zentralen grenzwertsatz der wahrscheinlichkeitsrechnung. *Mathematische Zeitschrift*, 40:521–559.

Fischer, H. (2011). *A history of the central limit theorem from classical to modern probability theory*. Springer, New York, NY.

Flavell, S. W., Raizen, D. M., and You, Y. J. (2020). Behavioral states. *Genetics*, 216(2):315–332.

Fryxell, J. M., Hazell, M., Börger, L., Dalziel, B. D., Haydon, D. T., Morales, J. M., McIntosh, T., and Rosatte, R. C. (2008). Multiple movement modes by large herbivores at multiple spatiotemporal scales. *Proceedings of the National Academy of Sciences of the United States of America*, 105(49):19114–19119.

Frézal, L. and Félix, M.-A. (2015). The natural history of model organisms: *C. elegans* outside the petri dish. *eLife*, 4:e05849.

Fürth, R. (1920). Die Brownsche bewegung bei berücksichtigung einer persistenz der bewegungsrichtung. Mit anwendungen auf die bewegung lebender infusorien. *Zeitschrift für Physik*, 2(3):244–256.

Gallagher, T., Bjorness, T., Greene, R., You, Y. J., and Avery, L. (2013). The geometry of locomotive behavioral states in *C. elegans*. *PLoS ONE*, 8(3):e59865.

Gloria-Soria, A. and Azevedo, R. B. (2008). npr-1 regulates foraging and dispersal strategies in caenorhabditis elegans. *Current Biology*, 18(21).

Golding, I. and Cox, E. C. (2006). Physical nature of bacterial cytoplasm. *Physical Review Letters*, 96(9):183–194.

Gomez-Marin, A., Partoune, N., Stephens, G. J., and Louis, M. (2012). Automated tracking of animal posture and movement during exploration and sensory orientation behaviors. *PLoS ONE*, 7(8):e41642.

Gomez-Marin, A., Stephens, G. J., and Brown, A. E. (2016). Hierarchical compression of *Caenorhabditis elegans* locomotion reveals phenotypic differences in the organization of behaviour. *Journal of the Royal Society Interface*, 13(121):20160466.

Gray, J. M., Hill, J. J., and Bargmann, C. I. (2005). A circuit for navigation in *Caenorhabditis elegans*. *Proceedings of the National Academy of Sciences of the United States of America*, 102(9):3184–3191.

Gurarie, E., Fleming, C. H., Fagan, W. F., Laidre, K. L., Hernández-Pliego, J., and Ovaskainen, O. (2017). Correlated velocity models as a fundamental unit of animal movement: synthesis and applications. *Movement Ecology*, 5(1):13.

Hakvey, S. C. (2009). Non-dauer larval dispersal in *Caenorhabditis elegans*. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*, 312(3):224–230.

Heisenberg, M. (2009). Is free will an illusion? *Nature*, 459(7244):164–165.

Helms, S. J., Rozemuller, W. M., Costa, A. C., Avery, L., Stephens, G. J., and Shimizu, T. S. (2019). Modelling the ballistic-to-diffusive transition in nematode motility reveals variation in exploratory behaviour across species. *Journal of the Royal Society Interface*, 16(157):20190174.

Hills, T., Brockie, P. J., and Maricq, A. V. (2004). Dopamine and glutamate control area-restricted search behavior in *Caenorhabditis elegans*. *Journal of Neuroscience*, 24(5):1217–1225.

Hills, T. T., Todd, P. M., Lazer, D., Redish, A. D., Couzin, I. D., Bateson, M., Cools, R., Dukas, R., Giraldeau, L. A., Macy, M. W., Page, S. E., Shiffrin, R. M., Stephens, D. W., and Wolfe, J. W. (2015). Exploration versus exploitation in space, mind, and society. *Trends in Cognitive Sciences*, 19(1):46–54.

Hoaglin, D. C. (2003). John W. Tukey and data analysis. *Statistical Science*, 18(3):311–318.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6(2):65–70.

Humphries, N. E., Queiroz, N., Dyer, J. R., Pade, N. G., Musyl, M. K., Schaefer, K. M., Fuller, D. W., Brunnschweiler, J. M., Doyle, T. K., Houghton, J. D., Hays, G. C., Jones, C. S., Noble, L. R., Wearmouth, V. J., Southall, E. J., and Sims, D. W. (2010). Environmental context explains lÃ ©vy and brownian movement patterns of marine predators. *Nature*, 465(7301):1066–1069.

Jackson, C. H. (2011). Multi-state models for panel data: the msm package for R. *Journal of Statistical Software*, 38(8):1–28.

Jager, M. D., Weissing, F. J., Herman, P. M., Nolet, B. A., and Koppel, J. V. D. (2011). *Lévy* walks evolve through interaction between movement and environmental complexity. *Science*, 332(6037):1551–1553.

Kays, R., Crofoot, M. C., Jetz, W., and Wikelski, M. (2015). Terrestrial animal tracking as an eye on life and planet. *Science*, 348(6240):aaa2478.

Kim, I. S. and Dickinson, M. H. (2017). Idiothetic path integration in the fruit fly drosophila melanogaster. *Current Biology*, 27(15):2227–2238.

Klein, M., Krivov, S. V., Ferrer, A. J., Luo, L., Samuel, A. D., and Karplus, M. (2017). Exploratory search during directed navigation in *C. elegans* and *Drosophila* larva. *eLife*, 6:e30503.

Kramer, D. L. and McLaughlin, R. L. (2001). The behavioral ecology of intermittent locomotion. *American Zoologist*, 41(2):137–153.

Kubo, R. (1966). The fluctuation-dissipation theorem. *Reports on Progress in Physics*, 29(1):255–284.

Kölzsch, A., Alzate, A., Bartumeus, F., de Jager, M., Weerman, E. J., Hengeveld, G. M., Naguib, M., Nolet, B. A., and van de Koppel, J. (2015). Experimental evidence for inherent lévy search behaviour in foraging animals. *Proceedings of the Royal Society B: Biological Sciences*, 282(1807):20150424.

Langevin, P. (1908). Sur la theorie du mouvement brownien. *Comptes Rendus de l'Académie des sciences*, 146:530–533.

Laplace, P. S. (1785). Mémoire sur les approximations des formules qui sont fonctions de très grands nombres. *Mémoires de l'Académie Royale des Sciences de Paris*, année 1782:1–88.

Leccardi, M. (2005). Comparison of three algorithms for Lévy noise generation. In *Proceedings of fifth EUROMECH nonlinear dynamics conference*, pages 1–14.

Lee, S. H. and Kang, S. H. (2015). Characterization of the crawling activity of *Caenorhabditis elegans* using a Hidden Markov model. *Theory in Biosciences*, 134(3-4):117–125.

Lenz, F., Chechkin, A. V., and Klages, R. (2013). Constructing a stochastic model of bumblebee flights from experimental data. *PLoS ONE*, 8(3):e59036.

Lyapunov, A. M. (1901). Sur un théorème du calcul des probabilités. *Comptes rendus hebdomadaires des séances de l' Académie des Sciences de Paris*, 132:126–128.

Lévy, P. (1935). Propriétés asymptotiques des sommes des variables aléatoires indépendantes ou enchaînées. *Journal de mathématiques pures et appliquées*, 14:347–402.

López-Cruz, A., Sordillo, A., Pokala, N., Liu, Q., McGrath, P. T., and Bargmann, C. I. (2019). Parallel multimodal circuits control an innate foraging behavior. *Neuron*, 102(2):407–419.

Madirolas, G., Al-Asmar, A., Gaouar, L., Marie-Louise, L., Garza-Enríquez, A., Rodríguez-Rada, V., Khona, M., Bello, M. D., Ratzke, C., Gore, J., and Pérez-Escudero, A. (2023). *Caenorhabditis elegans* foraging patterns follow a simple rule of thumb. *Communications Biology*, 6(1):841.

Mandelbrot, B. B. and Wallis, J. R. (1968). Noah, Joseph, and operational hydrology. *Water Resources Research*, 4(5):909–918.

Meneely, P. M., Dahlberg, C. L., and Rose, J. K. (2019). Working with worms: Caenorhabditis elegans as a model organism. *Current Protocols Essential Laboratory Techniques*, 19(1):e35.

Metzler, R., Jeon, J. H., Cherstvy, A. G., and Barkai, E. (2014). Anomalous diffusion models and their properties: non-stationarity, non-ergodicity, and ageing at the centenary of single particle tracking. *Physical Chemistry Chemical Physics*, 16(44):24128–24164.

Metzner, C., Mark, C., Steinwachs, J., Lautscham, L., Stadler, F., and Fabry, B. (2015). Superstatistical analysis and modelling of heterogeneous random walks. *Nature Communications*, 6:7516.

Meyer, P., Barkai, E., and Kantz, H. (2017). Scale-invariant Green-Kubo relation for time-averaged diffusivity. *Physical Review E*, 96(6):062122.

Meyer, P. G., Adlakha, V., Kantz, H., and Bassler, K. E. (2018). Anomalous diffusion and the Moses effect in an aging deterministic model. *New Journal of Physics*, 20(11):113033.

Meyer, P. G., Aghion, E., and Kantz, H. (2022). Decomposing the effect of anomalous diffusion enables direct calculation of the Hurst exponent and model classification for single random paths. *Journal of Physics A: Mathematical and Theoretical*, 55(27):274001.

Mischiati, M., Lin, H. T., Herold, P., Imler, E., Olberg, R., and Leonardo, A. (2015). Internal models direct dragonfly interception steering. *Nature*, 517(7534):333–338.

Montroll, E. W. and Weiss, G. H. (1965). Random walks on lattices. II. *Journal of Mathematical Physics*, 6(2):167–181.

Moore, T. Y., Cooper, K. L., Biewener, A. A., and Vasudevan, R. (2017). Unpredictability of escape trajectory explains predator evasion ability and microhabitat preference of desert rodents. *Nature Communications*, 8(1):062122.

Muggeo, V. M. (2003). Estimating regression models with unknown break-points. *Statistics in Medicine*, 22(19):3055–3071.

Méndez, V., Campos, D., and Bartumeus, F. (2014). *Stochastic foundations in movement ecology*. Springer, Berlin, Heidelberg.

Nathan, R., Getz, W. M., Revilla, E., Holyoak, M., Kadmon, R., Saltz, D., and Smouse, P. E. (2008). A movement ecology paradigm for unifying organismal movement research. *Proceedings of the National Academy of Sciences*, 105(49):19052–19059.

Okubo, A., Levin, S. A., et al. (2001). *Diffusion and ecological problems: modern perspectives*. Springer, New York, NY.

Padfield, D., O'Sullivan, H., and Pawar, S. (2021). rtpc and nls.multstart: A new pipeline to fit thermal performance curves in R. *Methods in Ecology and Evolution*, 12(6):1138–1143.

Panlilio, M. and Ryu, W. S. (2013). Food matters: a shift in the *C. elegans* foraging strategy at long time scales. Master's thesis, University of Toronto, Department of Physics and the Donnelly Centre, 60 St George St., Toronto, Canada M5S1A.

Patterson, T. A., Thomas, L., Wilcox, C., Ovaskainen, O., and Matthiopoulos, J. (2008). State-space models of individual animal movement. *Trends in Ecology and Evolution*, 23(2):87–94.

Pearson, K. (1905). The problem of the random walk. *Nature*, 72(1865):294–294.

Peliti, M., Chuang, J. S., and Shaham, S. (2013). Directional locomotion of *C. elegans* in the absence of external stimuli. *PLoS ONE*, 8(11):e78535.

Pierce-Shimomura, J. T., Dores, M., and Lockery, S. R. (2005). Analysis of the effects of turning bias on chemotaxis in *C. elegans*. *Journal of Experimental Biology*, 208(24):4727–4733.

Pierce-Shimomura, J. T., Morse, T. M., and Lockery, S. R. (1999). The fundamental role of pirouettes in *Caenorhabditis elegans* chemotaxis. *Journal of Neuroscience*, 19(21):9557–9569.

Podestá, T. S. V., Rosembach, T. V., Santos, A. A. D., and Martins, M. L. (2017). Anomalous diffusion and q-Weibull velocity distributions in epithelial cell migration. *PLoS ONE*, 12(7):e0180777.

Pradhan, S., Quilez, S., Homer, K., and Hendricks, M. (2019). Environmental programming of adult foraging behavior in *C. elegans*. *Current Biology*, 29(17):2867–2879.

Pyke, G. H. (2015). Understanding movements of organisms: It's time to abandon the Lévy foraging hypothesis.

R Core Team (2023). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rankin, C. H., Beck, C. D., and Chiba, C. M. (1990). *Caenorhabditis elegans*: A new model system for the study of learning and memory. *Behavioural Brain Research*, 37(1):89–92.

Real, L. A. (1981). Uncertainty and pollinator-plant interactions: the foraging behavior of bees and wasps on artificial flowers. *Ecology*, 62(1):20–26.

Riahi, M. K., Qattan, I. A., Hassan, J., and Homouz, D. (2019). Identifying short- and long-time modes of the mean-square displacement: an improved nonlinear fitting approach. *AIP Advances*, 9(5):055112.

Roberts, W. M., Augustine, S. B., Lawton, K. J., Lindsay, T. H., Thiele, T. R., Izquierdo, E. J., Faumont, S., Lindsay, R. A., Britton, M. C., Pokala, N., Bargmann, C. I., and Lockery, S. R. (2016). A stochastic neuronal model predicts random search behaviors at multiple spatial scales in *C. elegans*. *eLife*, 5:e12572.

Ryu, W. S. and Samuel, A. D. (2002). Thermotaxis in *Caenorhabditis elegans* analyzed by measuring responses to defined thermal stimuli. *Journal of Neuroscience*, 22(13):5727–5733.

Sabino, P. and Petroni, N. C. (2021). Gamma-related Ornstein–Uhlenbeck processes and their simulation. *Journal of Statistical Computation and Simulation*, 91(6):1108–1133.

Saeki, S., Yamamoto, M., and Iino, Y. (2001). Plasticity of chemotaxis revealed by paired presentation of a chemoattractant and starvation in the nematode caenorhabditis elegans. *Journal of Experimental Biology*, 204(10):1757–1764.

Salvador, L. C., Bartumeus, F., Levin, S. A., and Ryu, W. S. (2014). Mechanistic analysis of the search behaviour of *Caenorhabditis elegans*. *Journal of the Royal Society Interface*, 11(92):20131092.

Sapolsky, R. M. (2023). *Determined: A Science of Life Without Free Will*. Penguin Press, New York, NY.

Schienbein, M. and Gruler, H. (1993). Langevin equation, Fokker-Planck equation and cell migration. *Bulletin of Mathematical Biology*, 55(3):585–608.

Schulenburg, H. and Félix, M.-A. (2017). The Natural Biotic Environment of Caenorhabditis elegans. *Genetics*, 206(1):55–86.

Shingai, R. (2000). Durations and frequencies of free locomotion in wild type and gabaergic mutants of *Caenorhabditis elegans*. *Neuroscience Research*, 38(1):71–83.

Sims, D. W., Humphries, N. E., Hu, N., Medan, V., and Berni, J. (2019). Optimal searching behaviour generated intrinsically by the central pattern generator for locomotion. *eLife*, 8:e50316.

Sims, D. W., Southall, E. J., Humphries, N. E., Hays, G. C., Bradshaw, C. J., Pitchford, J. W., James, A., Ahmed, M. Z., Brierley, A. S., Hindell, M. A., Morritt, D., Musyl, M. K., Righton, D., Shepard, E. L., Wearmouth, V. J., Wilson, R. P., Witt, M. J., and Metcalfe, J. D. (2008). Scaling laws of marine predator search behaviour. *Nature*, 451(7182):1098–1102.

Slezak, J., Metzler, R., and Magdziarz, M. (2018). Superstatistical generalised langevin equation: non-Gaussian viscoelastic anomalous diffusion. *New Journal of Physics*, 20(2):023026.

Spanier, B., Stürzenbaum, S. R., Holden-Dye, L. M., and Baumeister, R. (2005). *Caenorhabditis elegans* neprilysin NEP-1: an effector of locomotion and pharyngeal pumping. *Journal of Molecular Biology*, 352(2):429–437.

Srivastava, N., Clark, D. A., and Samuel, A. D. (2009). Temporal analysis of stochastic turning behavior of swimming *C. elegans*. *Journal of Neurophysiology*, 102(2):1172–1179.

Stephens, D. W., Brown, J. S., and Ydenberg, R. C. (2007). *Foraging: behavior and ecology*. University of Chicago Press, Chicago, IL.

Stephens, G. J., Johnson-Kerner, B., Bialek, W., and Ryu, W. S. (2008). Dimensionality and dynamics in the behavior of *C. elegans*. *PLoS Computational Biology*, 4(4):e1000028.

Stephens, G. J., Johnson-Kerner, B., Bialek, W., and Ryu, W. S. (2010). From modes to movement in the behavior of *Caenorhabditis elegans*. *PLoS ONE*, 5(11):e13914.

Stephens, G. J., Mesquita, M. B. D., Ryu, W. S., and Bialek, W. (2011). Emergence of long timescales and stereotyped behaviors in *Caenorhabditis elegans*. *Proceedings of the National Academy of Sciences of the United States of America*, 108(18):7286–7289.

Takahashi, H. and Scott, T. K. (1993). Intensity of hydrostimulation for the induction of root hydrotropism and its sensing by the root cap. *Plant, Cell & Environment*, 16(1):99–103.

Tucker, M. A., Böhning-Gaese, K., Fagan, W. F., Fryxell, J. M., Moorter, B. V., Alberts, S. C., Ali, A. H., Allen, A. M., Attias, N., Avgar, T., Bartlam-Brooks, H., Bayarbaatar, B., Belant, J. L., Bertassoni, A., Beyer, D., Bidner, L., Beest, F. M. V., Blake, S., Blaum, N., Bracis, C., Brown, D., Bruyn, P. J. D., Cagnacci, F., Calabrese, J. M., Camilo-Alves, C., Chamaillé-Jammes, S., Chiaradia, A., Davidson, S. C., Dennis, T., DeStefano, S., Diefenbach, D., Douglas-Hamilton, I., Fennessy, J., Fichtel, C., Fiedler, W., Fischer, C., Fischhoff, I., Fleming, C. H., Ford, A. T., Fritz, S. A., Gehr, B., Goheen, J. R., Gurarie, E., Hebblewhite, M., Heurich, M., Hewison, A. J., Hof, C., Hurme, E., Isbell, L. A., Janssen, R., Jeltsch, F., Kaczensky, P., Kane, A., Kappeler, P. M., Kauffman, M., Kays, R., Kimuyu, D., Koch, F., Kranstauber, B., LaPoint, S., Leimgruber, P., Linnell, J. D., López-López, P., Markham, A. C., Mattisson, J., Medici, E. P., Mellone, U., Merrill, E., MirandaMourão, G. D., Morato, R. G., Morellet, N., Morrison, T. A., Díaz-Muñoz, S. L., Mysterud, A., Nandintsetseg, D., Nathan, R., Niamir, A., Odden, J., O'Hara, R. B., Oliveira-Santos, L. G. R., Olson, K. A., Patterson, B. D., Paula, R. C. D., Pedrotti, L., Reineking, B., Rimmler, M., Rogers, T. L., Rolandsen, C. M., Rosenberry, C. S., Rubenstein, D. I., Safi, K., Saïd, S., Sapir, N., Sawyer, H., Schmidt, N. M., Selva, N., Sergiel, A., Shiilegdamba, E., Silva, J. P., Singh, N., Solberg, E. J., Spiegel, O., Strand, O., Sundaresan, S., Ullmann, W., Voigt, U., Wall, J., Wattles, D., Wikelski, M., Wilmers, C. C., Wilson, J. W., Wittemyer, G., Zieba, F., Zwijacz-Kozica, T., and Mueller, T. (2018). Moving in the anthropocene: Global reductions in terrestrial mammalian movements. *Science*, 359(6374):466–469.

Turchin, P. (1998). *Quantitative analysis of movement: measuring and modeling population redistribution in animals and plants*. Weimar and Now; 13. Sinauer Associates.

Ueda, T., Terayama, K., Kurihara, K., and Kobatake, Y. (1975). Threshold phenomena in chemoreception and taxis in slime mold *Physarum polycephalum*. *Journal of General Physiology*, 65(2):223–234.

Uhlenbeck, G. E. and Ornstein, L. S. (1930). On the theory of the Brownian motion. *Physical Review*, 36(5):823–841.

Upadhyaya, A., Rieu, J. P., Glazier, J. A., and Sawada, Y. (2001). Anomalous diffusion and non-Gaussian velocity distribution of hydra cells in cellular aggregates. *Physica A: Statistical Mechanics and its Applications*, 293(3-4):549–558.

Vilk, O., Aghion, E., Avgar, T., Beta, C., Nagel, O., Sabri, A., Sarfati, R., Schwartz, D. K., Weiss, M., Krapf, D., Nathan, R., Metzler, R., and Assaf, M. (2022). Unravelling the origins of anomalous diffusion: from molecules to migrating storks. *Physical Review Research*, 4(3):033055.

Viswanathan, G. M., Afanasyev, V., Buldyrev, S. V., Murphy, E. J., Prince, P. A., and Stanley, H. E. (1996). Lévy flight search patterns of wandering albatrosses. *Nature*, 381(6581):413–415.

Viswanathan, G. M., Luz, M. G. D., Raposo, E. P., and Stanley, H. E. (2011). *The physics of foraging: an introduction to random searches and biological encounters*. Cambridge, New York, NY.

von Smoluchowski, M. (1906). Zur kinetischen theorie der Brownschen molekularbewegung und der suspensionen. *Annalen der Physik*, 326(14):756–780.

Wakabayashi, T., Kitagawa, I., and Shingai, R. (2004). Neurons regulating the duration of forward locomotion in *Caenorhabditis elegans*. *Neuroscience Research*, 50(1):103–111.

Wu, H. I., Li, B. L., Springer, T. A., and Neill, W. H. (2000). Modelling animal movement as a persistent random walk in two dimensions: expected magnitude of net displacement. *Ecological Modelling*, 132(1-2):115–124.

Zheng, Y., Brockie, P. J., Mellem, J. E., Madsen, D. M., and Maricq, A. V. (1999). Neuronal control of locomotion in *C. elegans* is modified by a dominant mutation in the GLR-1 ionotropic glutamate receptor. *Neuron*, 24(2):347–361.