Alma Mater Studiorum – Università di Bologna

DOTTORATO DI RICERCA IN FISICA

Ciclo XXX

Settore Concorsuale di afferenza: 02/D1

Settore Scientifico disciplinare: FIS/07

MODELING OF BIRTH-DEATH AND DIFFUSION PROCESSES IN BIOLOGICAL COMPLEX ENVIRONMENTS.

Presentata da: Silvia Vitali

Coordinatore Dottorato:

Prof.ssa Silvia Arcelli

Supervisore:

Prof. Gastone Castellani

Co-supervisore:

Ph.D. Gianni Pagnini

Esame finale anno 2018

One kid says to me, "See that bird? What kind of bird is that?" I said, "I haven't the slightest idea what kind of a bird it is." He says, "It's a brown-throated thrush. Your father doesn't teach you anything!" But it was the opposite. He had already taught me: "See that bird?" he says. "It's a Spencer's warbler." (I knew he didn't know the real name.) "Well, in Italian, it's a Chutto Lapittida. In Portuguese, it's a Bom da Peida. In Chinese, it's a Chung-long-tah, and in Japanese, it's a Katano Tekeda. You can know the name of that bird in all the languages of the world, but when you're finished, you'll know absolutely nothing whatever about the bird. You'll only know about humans in different places, and what they call the bird. So let's look at the bird and see what it's doing—that's what counts." (I learned very early the difference between knowing the name of something and knowing something.)

Richard Feynman

Abstract

Silvia VITALI

Modeling of birth-death and diffusion processes in biological complex environments.

This thesis is centered on the theory of stochastic processes and their applications in biological systems characterized by a complex environment. Three case studies have been modeled by the use of the three fundamental tools of stochastic processes: the master equation (ME), the stochastic differential equation (SDE) and the partial differential equation (PDE). The choice of an approach respect to another is determined also by the nature of the problem, i.e. the scale at which we are interested to observe the system, micro- meso- or macro-scopic. The principal approach here applied to deal with complexity is the characterization of the system by means of probability distributions describing each a parameter of the model or the introduction of fractional order derivatives to include non-local and memory effects maintaining the linearity in the equations. Different mathematical methods have been applied to obtain analytical solutions of the three original models proposed, related in particular to the theory of Laplace, Fourier and Mellin transform. In Chapter 1 we briefly review the theory of stochastic processes to introduce the topics presented in the following chapters.

Birth-death processes are fundamental in modeling of population dynamics, as the characterization of relative species abundance (RSA) in ecology. Models derived from ecological community studies have been also used to describe the evolution of genomic elements, and in particular, the dynamics of transposable elements. In Chapter 2 we derive a birth-death process master equation to test if Long Interspersed Elements (LINEs) can be modeled according to the neutral theory of biodiversity. According to this theory, the structure of the collection of LINE subfamilies would be the result of stochastic drift, as opposed to differences in ecological traits between subfamilies. Our results show that although the neutral model fits well the overall LINE distribution in humans, significant deviations from it can be observed by stratifying LINE subfamilies by age groups. This suggests that at specific times during the evolution of the mammalian genome multiple concurrently active LINE subfamilies might have been in direct competition. We further investigated how this competition could have been shaped by the LINE 5'UTR structure and by the chromatin landscape.

Dealing with biological systems (but not only), the diffusion process is one of the most important topics for a physicist. Brownian and anomalous diffusions are widely observed in nature and studied by the use of both phenomenological and founding models, the last ones trying to explain the origin of the anomaly in the system under study. One of the key concepts when speaking about anomalous diffusion is the complexity of the system itself, independently of the particular approach

of the model in use. Biological systems are complex and stochastic systems by definition at any level: from gene expression and motion of molecules inside the cell till the ecological description of individuals and their dispersal. However, thanks to physics, it is well known that this complexity does not arise from something that is complex at any level. If the problem is decomposed in smaller and smaller bricks it is possible to see that complexity arises at a meso-macroscopic scale from the same fundamental interactions treaten by fundamental physics and that stochasticity itself is generated by the complexity of the system observed. Following this idea, it seems reasonable that anomalous diffusion can be read in terms of a superposition of simpler processes. In Chapter 3 we derive a model of anomalous diffusion based on a Langevin approach in which anomalous behavior arise in the asymptotic intermediate state as a consequence of the heterogeneity of the system, from the superposition of Ornstein-Uhlenback processes.

Anomalous diffusive behavior can be also described by the fractional generalizations of diffusion equation by the introduction of fractional derivatives. Fractional derivatives are non-local integral operators that generalize the standard integer derivative, suitable to describe systems in which memory and non-local effects are observed. In Chapter 4 we propose an extension of the cable equation, useful to describe anomalous diffusion phenomena as the signal conduction in spiny dendrites, by introducing a Caputo time fractional derivative. The same generalization can be derived within the continuous time random walk framework, building the model as a superposition of Markovian processes, each characterized by its own timescale generated by the random geometry of the system. The same model can be also derived from a generalized grey Brownian motion in which is introduced a non-stationary distribution of length scales. The fundamental solutions of the most common boundary problems are derived by the application of the Efros theorem of Laplace transforms and written in terms of Wright special functions.

Acknowledgements

This Ph.D. has been funded by the Italian Ministry of Education at University of Bologna (Alma Mater Studiorum), "DIFA - Department of Physics and Astronomy", and supported by "BCAM - Basque Center for Applied Mathematics" in Bilbao. I spent six months in BCAM funded by the University of Bologna through Marco Polo program, working under Gianni Pagnini supervision, who has then become the official supervisor of this Ph.D. Thesis. I thank all BCAM people for the great chance I had to work in the wonderful work environment they create, which allows very different research fields to meet (in particular in the kitchen), and for the informal but professional atmosphere full of learning opportunities. I would like to thank for their precious help and support my supervisors Gianni Pagnini and Prof. Gastone Castellani and for his invaluable knowledge and suggestions Prof. Francesco Mainardi. I'm also grateful to Nicola Neretti, Steven Criscione, Claudia Sala, Italo Faria do Valle, Enrico Giampieri, Daniel Remondini, Paolo Paradisi, Oleksii Sliurasenko, Andrea Trucchia and Vittoria Sposini for their important contributions to the research and useful discussions.

Contents

Abstract iii

Acknowledgements v

1 Elements of stochastic processes 1
1.1 The Brownian Motion . . . 1
1.2 Brief introduction on classical methods . . . 3
1.3 Fractional models . . . 6

2 Ecological modeling of Long Interspersed Elements 11
2.1 Introduction . . . 11
2.1.1 Linear one-step one-variable birth-death process . . . 13
2.2 The neutral model of ecology of Volkov . . . 15
2.3 The ecosystem of Transposable Elements . . . 16
2.4 The model . . . 21
2.4.1 Mechanism of competition . . . 23
2.4.2 Numerical simulations of the model . . . 23
2.5 Data sources . . . 24
2.6 ABC method implementation and discrimination between the models proposed . . . 24
2.7 Results and discussion . . . 27

3 Langevin approach to generate anomalous transport in complex environment 45
3.1 Introduction . . . 45
3.2 The Langevin equation . . . 47
3.2.1 The standard Ornstein-Uhlenbeck process . . . 47
3.2.2 The classical Langevin oscillator . . . 49
3.2.3 Normal diffusion and Einstein-Smoluchowsky relations . . . 50
3.2.4 Fokker-Planck equation and probability distributions . . . 51
3.3 Heterogeneous ensemble of Brownian particles . . . 52
3.4 Test particle in heterogeneous ensemble . . . 56
3.5 Building anomalous diffusion regimes . . . 59
Laplace transformation mapping of global properties . . . 60
3.5.1 Population of timescales τ_c and anomalous time scaling . . . 62
Superdiffusive case . . . 63
Subdiffusive case . . . 66
3.5.2 Population of velocity diffusion coefficient D_V and deviations from Gaussian distribution . . . 70
3.6 Simulations of the model . . . 73

3.7 Discussion 74

4 Anomalous diffusion in spiny dendrites **77**
4.1 Introduction 77
4.2 Two stochastic models of anomalous diffusion 83
4.3 Time fractional cable model 87
4.3.1 Solution of the Signaling Problem via Laplace Transform 87
4.3.2 The Green function for the Cauchy Problem 89
4.3.3 Response to injected current 91
4.4 Discussion 93

5 Conclusions **97**

A Appendix A **99**
A.1 Mean field approximation and reductiond to 1D problem 99
A.2 Supplementary figures 100

B Appendix B **107**
B.1 Examples of relations between $b(\tau)$ and $h(D_v)$ 107
B.2 Special case with harmonic Langevin oscillator 108
B.2.1 Solution via Laplace Transform 110
B.2.2 Failure of confinment 112

Bibliography **113**

List of Figures

2.1 Example of oracle comparison of the hybrid algorithm with the Gillespiel algorithm. 24
2.2 upper panels: $b_A, d_A, b_I, b_{AI}, d_I = 0.0, 0.00005, 0.5, 0, 0$; middle panels: $b_A, d_A, b_I, b_{AI}, d_I = 0.0, 0.00005, 0.5, 0.0, 0.1$; lower panel: $b_A, d_A, b_I, b_{AI}, d_I = 0.0, 0.00005, 0.5, 0.099, 0.1$. 25
2.3 upper panels: $b_A, d_A, b_I, b_{AI}, d_I = 0.001, 0.0011, 0.09, 0, 0$; middle panels: $b_A, d_A, b_I, b_{AI}, d_I = 0.001, 0.0011, 0.09, 0, 0.001$; lower panels $b_A, d_A, b_I, b_{AI}, d_I = 0.001, 0.0011, 0.09, 0.0005, 0.001$. 26
2.4 **Simulations of competition in accumulation model.** Simulation of RSA of size of 10^4 elements. Simulation of a process with $b = 0.0001$, $d = 0.0011$. At each step of the simulation with a Gillespiel algorithm a competitor may arise with $p = 0.25$. The resulting distribution have means around 10 and 8 log2 of the counts respectively. 26
2.5 **ABC results of neutral model fit of Human RSA.**Comparison of the expected values from the mean parameters obtained with the ABC (black) to the Preston plot of the RSA and to the expected values obtained by non-linear least squares method (blue) on the same histogram(left panel);heat map of the posteriors of the parameters (right panel), log2 mean value (x axes), influx values (y axes). 31
2.6 **ABC results of neutral model fit of Mouse RSA.**Comparison of the expected values from the mean parameters obtained with the ABC (black) to the Preston plot of the RSA and to the expected values obtained by non-linear least squares method (blue) on the same histogram(left panel);heat map of the posteriors of the parameters (right panel), log2 mean value (x axes), influx values (y axes). 31
2.7 **ABC results of mixture model fit of Human RSA.**Comparison of the expected values from the mean parameters obtained with the ABC (black) to the Preston plot of the RSA and to the expected values obtained by non-linear least squares method (blue) on the same histogram(left panel);heat map of the posteriors of the parameters of the abund elements(center panel), heat map of the posteriors of the parameters of the rare elements(right panel), log2 mean value (x axes), influx values (y axes). 32

2.8 **ABC results of mixture model fit of Mouse RSA.**Comparison of the expected values from the mean parameters obtained with the ABC (black) to the Preston plot of the RSA and to the expected values obtained by non-linear least squares method (blue) on the same histogram(left panel);heat map of the posteriors of the parameters of the abund elements(center panel), heat map of the posteriors of the parameters of the rare elements(right panel), log2 mean value (x axes), influx values (y axes). 33

2.9 **Fit of the RSA of LINEs with a mixture model clusters different Mammalian Orders.** The set of optimized parameters obtained by fitting a mixture model of two negative binomials on 42 mammalian LINE RSAs are able to separate the most represented Taxonomic Orders: Υ_1 respect to Υ_2 and $1-x_1$ respect to $1-x_2$ respectively (upper panels). The couple of parameters for each of the two negative binomials (middle panels) allows as well a separation of different Orders, in particular in the case of the parameters describing the abundant group Υ_1, $1-x_1$. The parameters describing the pure neutral model (lower panel) do not clearly discriminate different Orders, however a trend similar to the couple of parameters of the abundant group of the mixture is observed. 34

2.10 **Statistical comparison of a mixture model of two negative binomials with respect to the neutral model of [Vol+03] during the evolution of the genome ecosystem in three Primates.** Data available from [Gio+07] have been used to rank order LINE Elements by their age of activity in Homo Sapiens, Chimpanzee and Rhesus Macaque genomes. The rank has been subdivided into intervals containing a fixed number of contiguous Elements ($N = 15$), each interval has been used as a sample ecosystem to fit both the neutral model and the mixture model. (a) The ABC model selection approach was used to compare the goodness of fit between the two models. The ABC model selection approach shows that a mixture model provides a better fit between rank positions 40 and 65 of the time ordered age of LINEs ("non-neutral time interval"). (b) Estimation of the mixture coefficient a during evolution. This coefficient represents the proportion of species associated to the regime in the mixture model, described by the negative binomial with a lower mean (i.e. LINE subfamilies with fewer Elements in the genome). When a is higher, it indicates the presence of a non negligible group of rare LINE species. (c) The percentage of LINE copies inserted in euchromatic regions displays a decreasing trend with time ordered age in human. (d) The copy number of LINE (euchromatin and heterochromatin insertions) displays a decreasing trend with time order as well. 35

2.11 **Statistical comparison of a mixture model of two negative binomials with respect to the neutral model of [Vol+03] during the evolution of the genome ecosystem in mouse and rat.** Data available from [Gio+07] have been used to rank order LINE Elements by their age of activity in Mouse and Rat genomes. The rank has been subdivided into intervals containing a fixed number of contiguous Elements ($N = 15$), each interval has been used as a sample ecosystem to fit both the neutral model and the mixture model. (a) The ABC model selection approach was used to compare the goodness of fit between the two models. The ABC model selection approach shows that a mixture model provides a better fit in the ancient Elements of the time ordered age of LINEs ("non-neutral time interval"). (b) Estimation of the mixture coefficient a during evolution. This coefficient represents the proportion of species associated to the regime in the mixture model, described by the negative binomial with a lower mean (i.e. LINE subfamilies with fewer Elements in the genome). When a is higher, it indicates the presence of a non negligible group of rare LINE species. (c) The percentage of LINE copies inserted in euchromatic regions displays a decreasing trend with time ordered age in human. (d) The copy number of LINE (euchromatin and heterochromatin insertions) displays a decreasing trend with time order as well. 36

2.12 **Evolution of estimated parameters for the neutral [Vol+03] and the mixture model [TZ13] in three Primates (Homo S., Chimpanzee, Rhesus M.) using a ABC method method (https://pymc-devs.github.io/pymc/).** (a) Birth-death rate parameter evolution, $x = b/d$, obtained for the neutral model; (b) Constant influx parameter evolution, $\Upsilon = S/b$, obtained for the neutral model; (c) Birth-death rate parameter evolution, $x = b/d$, obtained for the rare species regime in the mixture model; (d) Constant influx parameter evolution, $\Upsilon = S/b$, obtained for the rare species regime in the mixture model; (e) Birth-death rate parameter evolution, $x = b/d$, obtained for the common species regime in the mixture model; (f) Constant influx parameter evolution, $\Upsilon = S/b$, obtained for the common species regime in the mixture model. 37

2.13 **Evolution of estimated parameters for the neutral [Vol+03] and the mixture model [TZ13] in mouse and rat using a ABC method.** The analogous analysis reported in Fig. 2.12 for Primates has been performed for two Rodents (mouse, rat), for which data were available. (a) Birth-death rate parameter evolution, $x = b/d$, obtained for the neutral model; (b) Constant influx parameter evolution, $\Upsilon = S/b$, obtained for the neutral model; (c) Birth-death rate parameter evolution, $x = b/d$, obtained for the rare species regime in the mixture model; (d) Constant influx parameter evolution, $\Upsilon = S/b$, obtained for the rare species regime in the mixture model; (e) Birth-death rate parameter evolution, $x = b/d$, obtained for the common species regime in the mixture model; (f) Constant influx parameter evolution, $\Upsilon = S/b$, obtained for the common species regime in the mixture model; (g) Estimation of the mixture coefficient a during evolution. This coefficient represents the proportion of species associated to the second regime in the mixture model, described by the second negative binomial (rare species); (h) Comparison of Bayesian Information Criteria (BIC) of the neutral and mixture models. 38

2.14 **Space of parameters of LINEs evolution in a primates and murinides shows evidences of radiation events.** The space of parameters describing sliding window ecosystem of LINEs in human, chimpanzee, rhesus macaque (left panels) and mouse, rat (right panels) is shown. x and Υ parameters are correlated by the expected value (mean) of the distribution. Upper panels refer to the neutral model, middle panels refer to the group of the mixture model with highest copy number, lower panels refer to the group of rare Elements of the mixture model. Black circles indicated the most recent Elements, associated to Primate and Murinide radiation. It can be noted that two groups can be distinguished and it leads to a lower average copy number for primates and to higher average copy number for murinides. This is reflected in the distribution of copies in chromatine states, but in the case of human this transition is not significant as for mouse. As for the discrimination of host species in different taxonomic Orders, a mixture model seems more efficient to isolate the different dynamics. The group of rare Elements behave as a noise for the ecosystem and then does not produce significant informations to distinguish the two dynamics. . . 39

2.15 **Cluster species abundances in 46 mammalian reference genomes for LINEs in the non-neutral time interval.** The separation between abundant species (red) and rare species (blue) is maintained across all mammals included in this study, with the notable exception of the White Rhinoceros, which shows an opposite trend. 40

2.16 **5'UTR similarity between competing LINE retro-transposons in human.** In [KSB06] is suggested that different L1 Elements may coexist without competing if the 5'UTR is different, while a specie will overcome the others if the 5'UTR are very similar and might compete for the same factors. (a) The available consensus sequences of the 5'UTR of LINEs in the human genome have been aligned pairwise, with ClustalW2. (b) In several cases, the minimum distance is achieved between couple of Elements with similar ages and having high and low copy number respectively. 5'UTR regions have been selected for a subset of Elements with a well characterized 5'UTR sequence, and compared to all the other sequences available; the distance statistics and the label of the most similar sequences are reported in the plot. Significant similarity between 5'UTRs is observed for the following high and low copy numbers pairs: L1M2-L1M2c, L1MA9-L1M3b and L1M2-L1M3DE. 42

2.17 **Non linear correlation between number of insertion in euchromatin and heterochromatin states shows host genome adaptation mechanism.** Scatter plot in $log2$ scale of the number of insertions in euchromatin respect that in heterochromatin for each LINE specie in human (left panels) and mouse (right panels). The number of insertions in euchromatin and heterochromatin states result correlated by a power law with exponent ~ 1.2 (upper panels). The number of insertions in euchromatin and heterochromatin states result correlated by a power law with exponent. The group of most recent Elements (in red) is well separated from the others both in Human and Mouse by PCA (middle panels). This is much more evident when age variable is included in PCA (lower panels). 44

3.1 (color online) Statistical equivalence of the processes (3.41), (3.47) and (3.57). Panel $a)$ shows the density function of the center of mass $p(x_{\rm CM};t)$ and in the inset the variance $\langle X^2_{\rm CM}\rangle$; and panel $b)$ shows the density function $p(v_{\rm CM};t)$ and the variance $\langle V^2_{\rm CM}\rangle$. The PDFs of the process given by Eqs. (3.41) and (3.42) are presented via filled squares (■), Eqs. (3.46) and (3.47) are shown with empty triangles (△) and the process of Eq. (3.57) are the solid line. Different colors represent different times. Insets: green line refers to Eq. (3.41), blue line to Eq. (3.47) and red line refers to Eq. (3.57). Simulations have been performed with $\rho(\tau)$ and $p(m)$ given by Eqs. (3.63) and (3.59) with $\nu = 4/3\,, \eta = 3/4\,, \gamma_0 = 10^5$. 57
3.2 Simulations of Langevin model. 75

4.1 Figures taken from D. Johnston, S. Miao-Sin Wu. *Foundations of cellular neurophysiology*. MIT Press (1995). (Fig. 3.1). 80
4.2 Figure taken from D. Johnston, S. Miao-Sin Wu. *Foundations of cellular neurophysiology*. MIT Press (1995). (Fig. 4.6.). 80
4.3 Green function for the Signaling Problem 90
4.4 Step Response function for the Signaling Problem 90
4.5 Green function for the Cauchy Problem 91
4.6 Green function for the Second Kind Boundary Problem 93
4.7 Step Response function for the Second Kind Boundary Problem 93

A.1 **ABC results of neutral model fit of Cat RSA.**Comparison of the expected values from the mean parameters obtained with the ABC (black) to the Preston plot of the RSA and to the expected values obtained by non-linear least squares method (blue) on the same histogram(left panel);heat map of the posteriors of the parameters (right panel), log2 mean value (x axes), influx values (y axes). 101
A.2 **ABC results of neutral model fit of Chimpanzee RSA.**Comparison of the expected values from the mean parameters obtained with the ABC (black) to the Preston plot of the RSA and to the expected values obtained by non-linear least squares method (blue) on the same histogram(left panel);heat map of the posteriors of the parameters (right panel), log2 mean value (x axes), influx values (y axes). 101
A.3 **ABC results of neutral model fit of Cow RSA.**Comparison of the expected values from the mean parameters obtained with the ABC (black) to the Preston plot of the RSA and to the expected values obtained by non-linear least squares method (blue) on the same histogram(left panel);heat map of the posteriors of the parameters (right panel), log2 mean value (x axes), influx values (y axes). 101
A.4 **ABC results of neutral model fit of Dog RSA.**Comparison of the expected values from the mean parameters obtained with the ABC (black) to the Preston plot of the RSA and to the expected values obtained by non-linear least squares method (blue) on the same histogram(left panel);heat map of the posteriors of the parameters (right panel), log2 mean value (x axes), influx values (y axes). 102

A.5 **ABC results of neutral model fit of Rat RSA.**Comparison of the expected values from the mean parameters obtained with the ABC (black) to the Preston plot of the RSA and to the expected values obtained by non-linear least squares method (blue) on the same histogram(left panel);heat map of the posteriors of the parameters (right panel), log2 mean value (x axes), influx values (y axes). 102
A.6 **ABC results of neutral model fit of RhesuSrheMac3 RSA.**Comparison of the expected values from the mean parameters obtained with the ABC (black) to the Preston plot of the RSA and to the expected values obtained by non-linear least squares method (blue) on the same histogram(left panel);heat map of the posteriors of the parameters (right panel), log2 mean value (x axes), influx values (y axes). 102
A.7 **ABC results of mixture model fit of Cat RSA.**Comparison of the expected values from the mean parameters obtained with the ABC (black) to the Preston plot of the RSA and to the expected values obtained by non-linear least squares method (blue) on the same histogram(left panel);heat map of the posteriors of the parameters of the abund elements(center panel), heat map of the posteriors of the parameters of the rare elements(right panel), log2 mean value (x axes), influx values (y axes). 103
A.8 **ABC results of mixture model fit of Chimpanzee RSA.**Comparison of the expected values from the mean parameters obtained with the ABC (black) to the Preston plot of the RSA and to the expected values obtained by non-linear least squares method (blue) on the same histogram(left panel);heat map of the posteriors of the parameters of the abund elements(center panel), heat map of the posteriors of the parameters of the rare elements(right panel), log2 mean value (x axes), influx values (y axes). 103
A.9 **ABC results of mixture model fit of Cow RSA.**Comparison of the expected values from the mean parameters obtained with the ABC (black) to the Preston plot of the RSA and to the expected values obtained by non-linear least squares method (blue) on the same histogram(left panel);heat map of the posteriors of the parameters of the abund elements(center panel), heat map of the posteriors of the parameters of the rare elements(right panel), log2 mean value (x axes), influx values (y axes). 104
A.10 **ABC results of mixture model fit of Dog RSA.**Comparison of the expected values from the mean parameters obtained with the ABC (black) to the Preston plot of the RSA and to the expected values obtained by non-linear least squares method (blue) on the same histogram(left panel);heat map of the posteriors of the parameters of the abund elements(center panel), heat map of the posteriors of the parameters of the rare elements(right panel), log2 mean value (x axes), influx values (y axes). 104

A.11 **ABC results of mixture model fit of Rat RSA.**Comparison of the expected values from the mean parameters obtained with the ABC (black) to the Preston plot of the RSA and to the expected values obtained by non-linear least squares method (blue) on the same histogram(left panel);heat map of the posteriors of the parameters of the abund elements(center panel), heat map of the posteriors of the parameters of the rare elements(right panel), log2 mean value (x axes), influx values (y axes). 105
A.12 **ABC results of mixture model fit of RhesuSrheMac3 RSA.**Comparison of the expected values from the mean parameters obtained with the ABC (black) to the Preston plot of the RSA and to the expected values obtained by non-linear least squares method (blue) on the same histogram(left panel);heat map of the posteriors of the parameters of the abund elements(center panel), heat map of the posteriors of the parameters of the rare elements(right panel), log2 mean value (x axes), influx values (y axes). 105

List of Tables

2.1 Mean parameters obtained by the ABC method fitting the neutral model. 28
2.2 Mean parameters obtained by the ABC method fitting the mixture model part-1. 29
2.3 Mean parameters obtained by the ABC method fitting the mixture model part-2. 30

Dedicated to my family.
..in blood, in soul, and in paws.

Chapter 1

Elements of stochastic processes

Stochastic processes are at the basis of many research fields that desire a quantitative description when the complexity of the system forbids a deterministic approach. These fields range from economy, sociology to biology, chemistry, physics and industry problem solving itself. Despite the theory of probability have older roots, the origin of stochastic processes as we refer nowadays can be associated to the mathematical description of Brownian motion, which connected the microscopic description of molecules dynamics to the macroscopic theories of gas, heat conduction and fluid dynamics. In this chapter, we try to give a brief summary of the theory of stochastic processes, and its evolution toward the present time.

1.1 The Brownian Motion

We start recalling the theory of Brownian motion, which is deeply connected to this thesis work. Brownian motion has been first observed and investigated by the botanist Robert Brown in the XIX century, who was investigating the origin of the motion of pollen grains in water. This phenomenon has been described mathematically by Einstein [Ein05] and independently by Smoluchowski at the beginning of the XX century. Despite its apparent simplicity the work of Einstein contains as assumptions many of the fundamental concepts which will be rigorously developed years later and that will become the bases of the theory of stochastic processes as the Chapman-Kolmogorov equation, the Fokker-Plank equation, Kramers-Moyal expansions and so on.

The first assumption is the independence of each particle motion respect to the others as well as the motion of the same particle at different time points until the time interval is not too small. In this first assumption the concept of Markovianity and all that follows from it is almost defined.

The second assumption is that in a certain fixed time interval τ the particle moves of a certain quantity Δ with a certain probability $\phi(\Delta)$ and that such probability density does not change with time and it is the same for all the particles. This assumption contains the concept of thermal equilibrium, since the frequency of the shifts scaled by the timescale can be related to the distribution of the velocity of the Brownian particle at thermal equilibrium, and of ergodicity, meaning that a particle observed for enough long time is representative of all the others.

With these assumptions the fraction of particles dn that experiences a shift between Δ and $\Delta + d\Delta$ is $dn = N\phi(\Delta)d\Delta$, where N is the total number of particles in the system. We call the local concentration of particles $f = f(x,t)$. The variation

in time of the local concentration is generated by the displacement of the particles, weighted by the frequency of the shift, inside and outside the small volume (length) around the point x in the time interval τ is:

$$f(x, t+\tau) = \int_{-\infty}^{\infty} f(x+\Delta, t)\phi(\Delta)d\Delta\,. \tag{1.1}$$

Expanding the Taylor series until the first order in time and the second order in space, we obtain:

$$f + \tau\frac{\partial f}{\partial t} = f\int_{-\infty}^{\infty}\phi(\Delta)d\Delta + \frac{\partial f}{\partial x}\int_{-\infty}^{\infty}\Delta\phi(\Delta)d\Delta + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}\int_{-\infty}^{\infty}\Delta^2\phi(\Delta)d\Delta\,. \tag{1.2}$$

Because of symmetry considerations over the frequency of the shifts $\phi(\Delta)$ the linear term vanishes and it corresponds to diffusion equation:

$$\frac{\partial f(x,t)}{\partial t} = D\frac{\partial^2 f(x,t)}{\partial x^2}\,, \tag{1.3}$$

where the diffusion coefficient is defined as $D = \frac{1}{\tau}\int_{-\infty}^{\infty}\frac{\Delta^2}{2}\phi(\Delta)d\Delta$. Particle concentration represents the probability to find a particle per unit volume (or length in this case), thus the last expression corresponds to a special case of what will be called the Fokker-Planck equation.

As we will show in details in Section 3.2 Langevin introduced a mesoscopic description in phase space based on the concept of the mean field of the forces exerted on the particle. A mesoscopic particle ($\sim$ 100nm) moving in a viscous fluid feels the macroscopic drag exerted by the fluid and originated by the sum of the many microscopic interactions (mean field), but also the fluctuation generated by the hits at the microscopic level with the molecules that compose the medium in which it is immersed. The concept of thermal equilibrium of statistical mechanics is included by means of the mean kinetic energy of the particle, which is related to the temperature of the fluid in which is immersed. From this consideration, the powerful fluctuation-dissipation theorem arises [Kub66]. Solving the equation for the mean square displacement of the particle the same result obtained by Einstein is found, i.e. linear scaling with the time of the mean square displacement (MSD), with diffusion constant defined by particle and fluid physical characteristics:

$$D = \frac{k_B T}{6\pi\eta a}\,, \tag{1.4}$$

where a is the radius of the particle, η the viscosity of the fluid.

Einstein's and Smoluchowski's theories were verified experimentally by Perrin, Smoluchowski himself, Svedberg and Westgren [HS17]. Following the work of Einstein, Langevin, and Smoluchowski in 1914 Fokker and in 1917 Planck derived a partial differential equation for the Langevin equation which will become the well known and used Fokker-Planck equation, Kolmogorov and Wiener in 1930s developed the formalism for a rigorous mathematical treatment of the problem, Ito and Stratonovich developed two different but not independent approaches to define stochastic integrals [HS17].

What we can learn from this historical example is extremely deep. First of all the immense impact in the whole science that had the solution of a problem arose from the observation of natural phenomena, posed by the licit curiosity of a botanist. Secondly that over an intuition, as the ones in Einstein, Smoluchowski and Langevin

solutions, an entire theory could be formally developed, i.e. the theory of stochastic processes and stochastic differential equations. Third that every scientific theory needs experimental confirmation to acquire legitimacy.

Einstein's description in configuration space and Langevin's approach in phase space are reconciled by Smoluchowski in the overdamped limit, which correspond to an infinitely long time limit, and shows that velocity and coordinate variable decouples allowing a satisfactory description in the configuration space only. However, it has been pointed out by [Bod+16] that anomalous diffusion can be associated to the nonexistence of this limit, which entails long-time correlation appearance.

In this introduction, we try to give a very brief overview of stochastic processes and their generalization to fractional models.

1.2 Brief introduction on classical methods

A stochastic process describes, in general, the probabilistic time evolution of a system, i.e. the evolution in time of a random variable that fully describes the system under study. For a more rigorous derivation and details we resend the reader to [Kam81],[Gar90],[Ris89], [Fel71], [Kar19].

Consider a random variable $X(t)$ and measure its value $x_1, x_2, .., x_n$ at the time $t_1, t_2, ..t_n$. The joint probability density of this observation, if it exists, is:

$$P(x_1, t_1; x_2, t_2; ..; x_n, t_n)\,. \tag{1.5}$$

The conditional probability density of a set given another set is determined by:

$$P(x_1, t_1; x_2, t_2; ..; x_n, t_n | y_1, \tau_1; y_2, \tau_2; ..; y_m, \tau_m) = \frac{P(x_1, t_1; x_2, t_2; ..; x_n, t_n; y_1, \tau_1; y_2, \tau_2; ..; y_m, \tau_m)}{P(y_1, \tau_1; y_2, \tau_2; ..; y_m, \tau_m)}\,, \tag{1.6}$$

where no particular time ordering of the variables is required at this step.

If the value of a process a time t is totally independent of its values in past or future:

$$P(x_1, t_1; x_2, t_2; ..; x_n, t_n) = \prod_i P(x_i, t_i)\,. \tag{1.7}$$

A Markov process consists in a process in which probability is conditioned only by the most recent event. If we order the time series $t_1 > t_2 > ... > t_n > \tau_1 > ... > \tau_n$ we obtain:

$$P(x_1, t_1; x_2, t_2; ..; x_n, t_n | y_1, \tau_1; y_2, \tau_2; ..; y_m, \tau_m) = P(x_1, t_1; x_2, t_2; ..; x_n, t_n | y_1, \tau_1)\,, \tag{1.8}$$

through this assumption the joint probability of the first set can be rewritten as:

$$P(x_1, t_1; x_2, t_2; ..; x_n, t_n) = P(x_1, t_1 | x_2, t_2) P(x_2, t_2 | x_3, t_3) ... P(x_{n-1}, t_{n-1} | x_n, t_n) P(x_n, t_n)\,, \tag{1.9}$$

i.e. a Markov process is totally defined by the single state marginal probability, or probability to observe an event $B = (x_j, t_j)$ independently of the rest, and the one point to one point probability, or the probability to observe an event $A = (x_i, t_i)$ having observed in the past $B = (x_j, t_j)$. The Chapman-Kolmogorov equation explicits the independence in the conditioned probability of all the possible intermediate states for a Markov process:

$$P(x_1, t_1 | x_3, t_3) = \int P(x_1, t_1 | x_2, t_2) P(x_2, t_2 | x_3, t_3) dx_2\,. \tag{1.10}$$

The master equation (ME) is an alternative form of the Chapman-Kolmogorov equation, in which the evolution of a stochastic system is written in terms of transition probability between the possible states of the system. The probability density to observe a state $P(x,t)$ correspond to a time dependent transition probability $T_{t-t_0}(x|x_0)$ from the initial state ($t_0 = 0$) of the system to the state observed:

$$P(x,t) = T_{t-t_0}(x|x_0)\,, \tag{1.11}$$

writing the time dependent transition probability in terms of transition probability per unit time for two generic states $W(x|x')$, i.e. from x' to x, and substituting this relation in the Chapman-Kolmogorov equation it results:

$$\frac{\partial P(x,t)}{\partial t} = \int \left\{W(x|x')P(x',t) - W(x'|x)P(x,t)\right\} dx'\,. \tag{1.12}$$

In case of discrete states of the system ME is written in terms of a sum:

$$\frac{\partial P_n(t)}{\partial t} = \sum_{n'=0}^{\infty} \left\{W_{n,n'}P_{n'}(t) - W_{n',n}P_n(t)\right\}\,, \tag{1.13}$$

where $W_{n,n'}$ is the time independent transition probability from state n' to n and $P_n(t)$ the probability of the system to be in the state n at time t. ME transition matrix represent a short time transition probability and can be computed by any available method valid for the short time.

Another fundamental property of the ME is that for many systems it can be proven that the system reaches its stationary configuration at least in the longtime limit. The evolution in time of a stochastic variable can be approximated by a deterministic law that can be derived from the ME. This corresponds to a macroscopic law, as the drag in the Langevin equation, to which stochastic fluctuations can be added. When the expectation value (average) of the variable is representative of the system (linear processes), the deterministic equations can be derived by mean field approach. A systematic approximation of ME allows separating several orders of the internal noise in terms of a power series expansion respect to a suitable parameter of the system. These approaches are also useful to reduce the time of computation simulating the time-dependent evolution of a system, a similar approach has been used to simulate LINE dynamics in Chapter 2.

A special kind of ME is the Fokker Planck equation describing diffusion processes, which is also used to approximate more complex processes to the second order moment of transition probability matrix. Define $W(x';r) = W(x'|x)$ where $r = x' - x$ and expand ME in terms of a power series of r, allowed if W is sharply peaked in r but varies slowly wit x' and if P varies slowly with x. Each term of the expansion will result associated to the corresponding order of the moments of $W(x';r)$:

$$a_i(x) = \int r^i W(x,r)dr\,. \tag{1.14}$$

The truncation at the first two moment correspond to the Fokker Planck equation with $A(x) = a_1(x)$ and $B(x) = a_2(x)$.

A special case that result particularly relevant for our purpose is:

$$\frac{\partial P(x,t)}{\partial t} = -\frac{\partial}{\partial x}(A(x)P(x,t)) + \frac{1}{2}\frac{\partial^2}{\partial x^2}(B(x)P(x,t))\,, \tag{1.15}$$

also called Smoluchowski equation, with $B(x) > 0$. The right term can be written as a probabilistic flux by introducing a continuity equation for probability, then we have:

$$J(x,t) = A(x)P(x,t) - \frac{1}{2}\frac{\partial}{\partial x}(B(x)P(x,t))\,, \tag{1.16}$$

in which we can distinguish a drift term and a fluctuation or diffusion term. When $A(t) = A_0 + A_1 x$, with $A_1 < 0$, and $B(t) = B_0$ the corresponding Fokker-Planck equation describes the Ornstein Uhlenback (OU) process that will be treated more in details in Chapter 3.

The Langevin approach is an alternative to Fokker-Planck equation, describing the evolution of the variable through the physical macroscopic law and adding a noise term to account statistical fluctuations, i.e. by defining a stochastic differential equation. Langevin equation will be the basis of the model developed in Chapter 3. For this reason we leave a detailed presentation of the Langevin approach to the dedicated Section 3.2.

Consider a generic Langevin equation:

$$\dot{x} = A(x) + C(x)L(t)\,, \tag{1.17}$$

where $L(t)$ is a delta correlated noise term with zero mean and variance σ.

The presence of the noise term introduces a series of challenges to formally define integrals and function transformation of the variables since the noise term is in general strongly discontinuous. To deal with this problem the concept of stochastic integral has been introduced:

$$S_n = \int_0^t G(t)dW(t) = \sum_{i=1}^{n} G(\tau_i)[W(t_i) - W(t_{i-1})]\,, \tag{1.18}$$

where t has been divided in n intervals, $dW = W(t_i) - W(t_{i-1})$, is the stochastic increment in the time interval associated to some white noise $L(t)$, namely the Wiener increment, and the function G is calculated in $\tau_i \in [t_{i-1}, t_i]$, leading to different results depending on this choice. The two main approaches are the Ito stochastic integral, for which $\tau_i = t_{i-1}$ and the Stratonovich stochastic integral, for which $\tau_i = \frac{t_{i-1}+t_i}{2}$. These approach leads to different but related results, so we can pass from a description to the other by proper variable transformation. While Ito calculus is the most natural choice from a mathematical point of view, Stranotovich results more suitable from a physical point of view, where ordinary calculus can be applied in the transformation of the variables.

In relation to the Fokker-Planck equation, the following relation to Langevin equation in Eq.(1.18) exists. Computing the moments with the Stratonovich integral the corresponding Fokker-Planck equation reads:

$$\frac{\partial P(x,t)}{\partial t} = -\frac{\partial}{\partial x}\left[\left(A(x) + \frac{1}{2}\sigma C(x)\frac{\partial C(x)}{\partial x}\right)P(x,t)\right] + \frac{1}{2}\frac{\partial^2}{\partial x^2}\left(C(x)^2 P(x,t)\right)\,. \tag{1.19}$$

The same results can be infact obtained by transforming the variable $x(t) \to y = \int \frac{dx}{C(x)}dx$ with ordinary calculus, which is allowed by the implicit use of Stratonovich approach, to recover a more handling form of the Langevin equation:

$$\dot{y} = A'(y) + L(t)\,, \quad A' = \frac{A(x)}{C(x)}\,. \tag{1.20}$$

1.3 Fractional models

Fractional models arise from the need to give a parsimonious and simple mathematical description to many anomalous behaviors observed in real systems. In particular long-time correlation of the variable, or memory, the dependence of the observable on the age of the system, i.e. aging, and finally the emergence of non-Gaussian distributions. Last but not least the anomalous power law time scaling of the MSD, which is often a consequence of one or more of the previous features. The three fundamental approaches used to deal with this new challenges are based on generalizations of the same tools introduced in the previous sections: generalized ME and continuous time random walk (CTRW), in which the sojourn time of the particle is a random variable as well as the length of the jumps; generalizations of Langevin equation by the introduction of colored noise and/or memory kernels; fractional generalizations of the Fokker-Planck equation. Particularly relevant for this purpose has been the introduction of fractional calculus, in which power-law memory kernels naturally arise in the integrodifferential operators, i.e. fractional derivatives and integrals, from the generalization to non-integer values of the Cauchy formula [GM97]. For more rigorous mathematical treatments and details we refer to [Bal+12; KRS08; BIHS; MLP01; Mur11; GM97]. In Chapter 3 we will show that these approaches are not always necessary but long time correlations may arise by the superposition of simpler processes, i.e. OU processes, without memory due to the complexity of the system itself.

The most recurrent definition of anomalous diffusion is based on the nonlinear time scaling of the MSD:

$$\langle x(t)^2 \rangle \sim t^\alpha \,. \tag{1.21}$$

Following this approach, we distinguish sublinear regimes, i.e. subdiffusion, observed as example in disordered solids and in biological, in porous and in fractal media; and superlinear regimes, namely superdiffusion, observed as an example in turbulent plasma, transport in polymers and in many living organisms motion. Between the most common models applied to describe these phenomena we recall Levy flights and Levy walks, particular types of CTRW [ZDK15]; fractional Brownian motion (fBm), which can be derived by a generalization of the Langevin equation [MP08a; Lut01]; the generalized grey Brownian motion (ggBm) [MP08b]; and the space and/or time fractional diffusion equation [MLP01].

CTRW can be described by a generalized ME, that is defined by the introduction in the gain-loss expression of a time dependent transition matrix $K(x, x'; t - t')$, representing the probability per unit time to make a transition from x to x' during the time $t - t'$, instead of the time independent transition matrix per unit time $W(x|x')$:

$$\frac{\partial P(x,t)}{\partial t} = \int_0^t \sum_{x'} \left\{ K(x, x'; t - t') P(x', t') - K(x', x; t - t') P(x, t') \right\} dx' dt' \,. \tag{1.22}$$

The main idea behind the CTRW is that the sojourn times, i.e. the time passed between two jumps of the particle, are i.i.d. random variables as well as the jumps length. This characteristic generates the issue that we don't know a priori how many jumps the particle has performed to reach the point (x, t) starting from (x', t'). To derive the fundamental ME for CTRW is necessary to define first the conditional probability density to be in (x, t) after $n + 1$ steps given the initial condition $(x_0, t_0 = 0)$.

This can be derived by the recursive relation:

$$q_{n+1}(x,t|x_0,0) = \int_{-\infty}^{\infty}\int_0^t \Psi(x-x',t-t')q_n(x',t'|x_0,0)dx'dt' \,, \tag{1.23}$$

where the conditional probability to be in (x,t) after $n+1$ steps is written as integral form in terms of the conditional probability to be in a generic point (x',t') one step before. $\Psi(x-x',t-t')$ represents infact the probability to perform a jump of length $x-x'$ in the time interval $t-t'$. Assuming independence of jump lengths $x-x'$ and waiting times $t-t'$ random variables, the distribution $\Psi(x-x',t-t')$ can be written as the product of the two independent densities:

$$\Psi(x-x',t-t') = \lambda(x-x')\psi(t-t') \,. \tag{1.24}$$

The total conditional probability is obtained by summation over all the possible number of steps performed,

$$q(x,t|x_0,0) = \sum_n q_n(x,t|x_0,0) \,. \tag{1.25}$$

Substituting the recursive relation and the initial conditions it results:

$$q(x,t|x_0,0) = \int_{-\infty}^{\infty}\int_0^t \Psi(x',t')q(x-x',t-t'|x_0,0)dx'dt' + \delta(t)\delta_{x,x_0} \,. \tag{1.26}$$

CTRW can describe both Markovian and not Markovian processes depending on the distribution of waiting times, however, exponential distribution is the only one leading to a Markovian process. Power law waiting time probability and jump lengths density defined by an even function with finite moments generate fractional subdiffusion, which is non-Markovian and non-Gaussian. Instead exponentially distributed waiting times with Levy distributed jump lengths with power-law asymptotic produces the superdiffusive process Levy flights. These two last processes can be related respectively to the time fractional and space fractional diffusion equation fundamental solutions [MLP01].

Lévy flights process is characterized by divergent second and higher moments. This feature is difficult to be related to experimental data because it is associated with an infinite velocity of the particle. A solution to this issue has been the introduction of a constant velocity v for the particles. The resulting process is the Levy walk, widely applied in many fields of science to model anomalous transport. For an extensive review, we refer to [ZDK15]. Since the particle remains bounded in the ballistic cone all the moments remain finite, but long jumps still may characterize the statistics to allow superdiffusion. The waiting times distribution before jumps, $\psi(t)$, becomes the waiting times distribution before changing the direction of the motion. The frequency of velocity changes $\nu(x,t)$ is described by:

$$\nu(x,t) = \int_{-\infty}^{\infty} dx' \int_0^t \phi(x',\tau)\nu(x-x',t-\tau)d\tau \,, \tag{1.27}$$

where

$$\phi(x,t) = \frac{1}{2}\delta(|x|-vt)\psi(t) \,, \tag{1.28}$$

guarantees that particles remain confined in the ballistic cone. The probability density $P(x,t)$ is then written:

$$P(x,t) = \int_{-\infty}^{\infty} dx' \int_0^t \Phi(x',\tau) P(x-x', t-\tau) d\tau \,, \tag{1.29}$$

where

$$\Phi(x,\tau) = \frac{1}{2}\delta(|x| - v\tau)\Psi(\tau) \,, \tag{1.30}$$

and $\Psi(\tau)$ is the analogous of the survival probability, i.e. the probability that the particle does not change direction of motion in the time interval τ.

The generalized Langevin equation (GLE) describes non-Markovian and/or non-Gaussian processes by the introduction of a colored noise $\xi(t)$ correlated through a dissipative memory kernel $\mathcal{K}(t)$, by the second fluctuation dissipation theorem [Kub66]:

$$m\frac{\partial v(t)}{\partial t} = -\int_0^t \mathcal{K}(t-t')v(t')dt' + \xi(t) \,, \tag{1.31}$$

where holds the relation:

$$\langle \xi(t)\xi(0)\rangle = k_b T \mathcal{K}(t) \,. \tag{1.32}$$

Fractional Brownian motion is a non-Markovian Gaussian process with stationary increments, showing anomalous time scaling of the MSD. It is defined by [JM10] as the integral of a fractional Gaussian noise. The corresponding stochastic process can be described through GLE [Lut01]. The white noise is then substituted by a power law correlated noise $\xi^H(t)$ characterized by zero mean, and Hurst exponent $H \in [0,1]$:

$$\langle \xi^H(t)\rangle = 0 \,, \tag{1.33}$$

$$\langle \xi^H(t)\xi^H(0)\rangle = k_b T \gamma(t) \,, \tag{1.34}$$

where

$$\gamma(t) = \frac{D_\alpha}{k_b T} \cdot t^{-\alpha} \,, \quad \alpha = 2H \,. \tag{1.35}$$

Then we have:

$$m\frac{dv(t)}{dt} = -\int_0^t \gamma(t-t')v(t')dt' + \xi^H(t) \,. \tag{1.36}$$

The corresponding diffusion equation is derived substituting the diffusion coefficient D with a time dependent coefficient of diffusion [Ade76]:

$$D(t) = k_b T \mathcal{L}^{-1}\left[s\widetilde{\gamma}(s)\right]^{-1} = D_\alpha t^{\alpha-1} \,, \tag{1.37}$$

with $0 < \alpha < 1$ or $1 < \alpha < 2$.

The corresponding FPE becomes:

$$\frac{\partial P(x,t)}{\partial t} = D_\alpha t^{\alpha-1}\frac{\partial^2}{\partial x^2}P(x,t) \,, \tag{1.38}$$

The generalized gray Brownian motion (ggBM), or Erdély-Kober fractional diffusion, [MP08b; MM09; DG+16] is a parametric class that defines Hurst self-similar with stationary increments processes (H-SSSI), dependent over two parameters, $\alpha \in (0,2), \beta \in (0,1]$, which allow to describe both sub and super diffusion regimes. ggBm can be read as a generalization of Gaussian processes since for the special

case $\beta = 1, \alpha \in (0,2)$ it corresponds to fBm and then to a Gaussian process, finally if $\beta = \alpha = 1$ BM is retrieved. The case $\alpha = \beta \in (0,1)$ corresponds to time fractional diffusion, which marginal density probability is governed by the time-fractional diffusion equation.

Self-similarity means that the process $X(at) = a^H X(t)$, where H is the Hurst exponent. Existence of stationary increments corresponds to invariance under time shifts transformation, $X(t+t') - X(t)$.

The ggBm can be defined by the random variables product:

$$X_{H,\beta} = \sqrt{\Lambda_\beta} X_H(t)\,, \tag{1.39}$$

where the random variable Λ_β is the Mainardi-Wright function $M_\beta(\lambda)$ with $\lambda \geq 0$ and $0 < \beta \leq 1$, and the stochastic process X_H is an fBM with $2H = \alpha$. The marginal density for $X_{H,\beta}$ respect to Λ is equivalent to the integral subordination relation:

$$P(x,t) = \int_0^\infty \mathcal{G}(x,\tau)\phi(\tau,t)d\tau\,, \tag{1.40}$$

where $\mathcal{G}(x,\tau)$ is a Gaussian with variance τ and $\phi(\tau,t) = \frac{1}{2t^{\alpha/2}} M_\beta(\frac{\tau}{t^{\alpha/2}})$, then all the moments related to this process are finite.

The evolution equation for the ggBM PDF is expressed in term of the time Erdély-Kober fractional derivative [Pag14]:

$$\frac{\partial P}{\partial t} = \frac{\alpha}{\beta} t^{\alpha-1} D^{\beta-1,1-\beta}_{\alpha/\beta} \frac{\partial^2 P}{\partial x^2}\,. \tag{1.41}$$

The fundamental solution of Eq. (1.41) corresponds to the the integral form in Eq. (4.52),and reads explicitly:

$$P(x;t) = \frac{1}{2t^{\alpha/2}} M_{\beta/2}\Big(\frac{|x|}{t^{\alpha/2}}\Big)\,. \tag{1.42}$$

The space-time fractional diffusion is a generalization of the diffusion equation by the introduction of Caputo time-fractional derivative and Riesz-Feller space fractional derivatives:

$${}_tD^\beta_* P(x;t) = {}_xD^\alpha_\theta P(x;t), \quad -\infty < x < +\infty, \quad t \geq 0\,, \tag{1.43}$$

with the real parameters are restricted to the ranges: $0 < \alpha \leq 2$, $|\theta| \leq \min\{\alpha, 2-\alpha\}$, $0 < \beta \leq 1$ or $0 < \beta \leq \alpha \leq 2$. Particular cases are space fractional diffusion when $\beta = 1$ and time fractional diffusion when $\theta = 0, \alpha = 2$. Standard diffusion is retrieved for $\beta = 1$ and $\alpha = 2$. The fundamental solution for Eq. (1.43) [MLP01] is defined by means of these parameters

$$\mathcal{G}^\theta_{\alpha,\beta}(x;t) = \int_0^\infty L^\theta_\alpha(x,u) M_\beta(u,t) du\,, \tag{1.44}$$

where the Levy distribution $L^\theta_\alpha(x,u)$ corresponds to the fundamental solution of the space fractional diffusion, i.e. $\beta = 1$, while the Mainardi-Wright distribution $M_\beta(u,t)$ is the fundamental solution of the special case $\theta = 0, \alpha = 1$, i.e. time fractional diffusion. This integral form is often expressed through the Mellin-Barnes integral representation or in terms of H-Fox function [MLP01; MPS05]. An alternative approach to define the stochastic process described by the symmetric space-time fractional diffusion, i.e. $\theta = 0$, is analogous to one applied for ggBm. The process is

defined by the marginal density of a product of two independent random variables:

$$X_{\alpha,\beta}(t) = \sqrt{\Lambda_{\alpha/2,\beta}} G_{2\beta/\alpha}(t), \quad 0 < \beta \leq 1, \quad 0 < \alpha \leq 2\,, \tag{1.45}$$

where $G_{2\beta/\alpha}(t)$ is a H-SSSI Gaussian process with power law variance $t^{2\beta/\alpha}$ and $\Lambda_{\alpha/2,\beta}$ is an independent constant non-negative random variable distributed according to the pdf $K^{-\alpha/2}_{\alpha/2,\beta}(\lambda), \quad \lambda \geq 0$. This stochastic process is a generalization of the Gaussian processes fBm and Bm, and it is uniquely determined by its mean and its autocovariance structure.

In conclusion, we may consider those operations that act on fractionalization or stretching of the time variable introduce memory in the system and characterize non-Markovian processes, this is done by non-Poissonian waiting times distribution in CTRW, by colored noise in Langevin approach. Instead non-Gaussian behavior is obtained by the introduction of the fractional component in the space variable, as a for jumps distribution in CTRW, random length scale in ggBM. The introduction of time and space fractional derivatives in the diffusion equation instead accounts for both contemporary, the shape of the fundamental solution and the anomalous time scaling of the variance (if it exists), because they act on the self-similarity properties of the solution [MLP01].

There exists often several different approaches to describe a process, depending on the scale we are observing the system. We may be interested in properties of the single particle or ensemble trajectories, for which a random walk approach is enough. We may be also interested in the dynamics of the system in which macroscopic forces may enter explicitly in concurrence with microscopic random fluctuations for which Langevin approach is suggested. Finally, we may be interested in the evolution of the system as an ensemble, the ME or PDE is the cheapest approach.

Chapter 2

Ecological modeling of Long Interspersed Elements

Birth-death processes are fundamental in modeling of population dynamics, as the characterization of relative species abundance in ecology. Models derived from ecological communities studies have been used to describe the evolution of genomic elements, and in particular, the dynamics of transposable elements. Here we apply this approach to test if Long Interspersed Elements can be modeled according to neutral theory of biodiversity. According to this theory, the structure of the collection of LINE subfamilies would be the result of stochastic drift, as opposed to differences in ecological traits between subfamilies. Our results show that although the neutral model fits well the overall LINE distribution in humans, significant deviations from it can be observed by stratifying LINE subfamilies by age groups. This suggests that at specific times during the evolution of the mammalian genome multiple concurrently active LINE subfamilies might have been in direct competition. We further investigate how this competition could have been shaped by the LINE 5'UTR structure and by the chromatin landscape.

2.1 Introduction

Birth-death processes describe phenomena in which changes in the abundances of the players in the process are involved. These players can be chemical species and a birth-death process may describe chemical reactions in which one or more species change their concentration. They can also be a living being, and birth-death process describes reproduction and death phenomena and may include competition mechanisms between different species.

In general the most appropriate mathematical tool to describe a birth-death process is the master equation, which corresponds to a more handling integral form of the Chapman Kolmogorov equation [Gar90] in the case of Markov processes, fully determined by the initial conditions of the system:

$$\frac{\partial P_n(t)}{\partial t} = \sum_{n'=0}^{\infty} \left\{ W_{n,n'} P_{n'}(t) - W_{n',n} P_n(t) \right\} , \tag{2.1}$$

where $W_{n,n'}$ is the time independent transition probability from state n' to n and $P_n(t)$ the probability to observe the system in the state n, i.e. to observe n individuals, at time t.

It is not always easy to calculate the exact solution of a master equation neither it is always necessary to write it down explicitly to derive the solution. In general when detailed balance is satisfied, i.e. $W_{n,n'}P_{n'} = W_{n',n}P_n$, the stationary solution of the process can be derived and correspond to the distribution P_n. The calculation of the time dependent solution requires more efforts, a standard tool for its derivation is the generating function: $G(s,t) = \sum_n s^n \cdot P_n(t)$. Approximate (and sometimes exact) solution can be determined by the introduction of auxiliary descriptions, as an example by constructing the corresponding Fokker Planck equation, if allowed, or by switching to a mesoscopic description, i.e. stochastic differential equations (SDEs). In SDEs approach deterministic description, representing the macroscopic laws, and random fluctuations, characterizing the stochastic nature of the process, are independent, thus dissipation-fluctuation arguments must be introduced if the origin of these two components in the process is the same. For details about these standard procedures, we refer to the textbooks [Gar90; Kam81] . A more systematic approximation method is based on the series expansion of the master equation treated by [Kam81] in terms of a small parameter Ω^{-1}, if the assumption that for large Ω jumps become small, i.e. fluctuation becomes negligible, is satisfied.

Birth-death process is not strictly defined as a diffusion process because there is no diffusion of particles in space, however, the number of individuals can be seen as the variable that is spanned in time by the system. If the first and second moments of the process are finite it can be in fact described at least at first order approximation by a Fokker Planck equation. Mapping a diffusion process described by a standard Fokker Planck equation type in terms of a birth-death process is always possible instead, which could be useful to derive the stationary solution if detailed balance is satisfied. Tough it is not the case study of this Chapter, we recall for the reader that anomalous diffusion processes with infinite variance are widely described in the modern literature and applied to real datasets, particularly relevant the case of Levy Flights [ZDK15].

Following the notation of [Kam81] let's define the birth-death process:

$$\dot{P}_n(t) = (E^{-1} - 1)g_n P_n(t) + (E^{+1} - 1)r_n P_n(t)\,, \tag{2.2}$$

where $E^{\pm 1} f(n) = f(n \pm 1) - f(n)$ represents the Van Kampen operator for unitary increments, this class of processes is called one-step processes. The introduction of Van Kampen operators implicitly means that g_n, r_n can be eventually expressed as functions of the variable n. For sake of convenience, we will introduce the present topic by means of this particular case, however, it can be generalized to arbitrary increments of size Δ, or a distribution of them.

The operators g_n and r_n represent the transition probabilities W for the birth and death process respectively: $g_n = W_{n+1,n}$, $r_n = W_{n-1,n}$. If detailed balance condition $W_{n,n'}P_{n'} - W_{n',n}P_n = 0$ is satisfied, P_n corresponds to the stationary solution of the master equation. For the one-step birth death process in Eq.(2.2) it corresponds to:

$$P_n = P_0 \prod_{k=1}^{n} \frac{g_{k-1}}{r_k}\,. \tag{2.3}$$

The state $n = 0$ is particular relevant in case of birth-death processes, because a negative number of individuals is not meaningful. From the expression for the stationary solution, it becomes evident the role of this boundary. From detailed balance condition it comes out that $g_0 P_0 = r_1 P_1$. If $g_0 = 0$, i.e. there is not an external influx of individuals in the system, and there is not a further condition over the state $n = 1$,

the system presents an absorbent state zero. Detailed balance is not satisfied and $P_0 = 1$ is the asymptotic stationary solution: $\lim_{t\to\infty} P_n(t) = \delta_{n,0}$.

2.1.1 Linear one-step one-variable birth-death process

Let's define the master equation for the one variable linear birth-death process, which will be the bases for the case study presented in this Chapter:

$$\dot{P}_n(t) = (E^{-1} - 1)g_n P_n(t) + (E^{+1} - 1)r_n P_n(t)\,, \tag{2.4}$$

where $g_n = b \cdot n$ and $r_n = d \cdot n$. The parameters b, d represent the procapite birth death rates respectively. This process has an absorbent state in $n = 0$, meaning that the only stable stationary solution is $P_n^S = \delta_{0,n}$. Nevertheless, it can be interesting to determine the time-dependent solution of the problem and the survival probability of the system. The survival probability is the probability that the system does not reach the absorbent state until time t, and it is related to the first passage problem. For the present example the probability that the system reaches the absorbent state, i.e. probability of extinction, is represented by $P_0(t)$. Survival probability can be defined as $\Psi(t) = 1 - P_0(t)$. The state $n = 0$ is a natural boundary and it is possible to calculate the time-dependent probability for each state $n = 0, 1, 2, ..$ with the method of the generating function.

The generating function can be defined by the following relation:

$$G(s,t) = \sum_n s^n \cdot P_n(t)\,, \tag{2.5}$$

from which probability density can be retrieved:

$$P_n(t) = \frac{1}{n!}\frac{\partial^n G(s,t)}{\partial s^n}|_{s=0}\,. \tag{2.6}$$

The master equation in Eq.(2.2) can be rewritten as the sum of all the allowed increments:

$$\dot{P}_n(t) = \sum_k (E^k - 1)c_k(n)P_n(t)\,, \tag{2.7}$$

that in the case of Eq.(2.4) correspond to $k = \pm 1$, where $c_k(n)$ are the corresponding transition probability. By the use of this last notation the time evolution equation for the generating functions becomes:

$$\begin{aligned}\frac{\partial G(s,t)}{\partial t} &= \sum_n s^n \cdot \frac{\partial P_n(t)}{\partial t} \\ &= \sum_k (s^{-k} - 1)\sum_n s^n \frac{c_k(n)}{n!}\frac{\partial^n}{\partial s^n} G(s,t)|_{s=0}\,.\end{aligned} \tag{2.8}$$

The transition probability can be expanded as $c_k(n) = \sum_a c_k^a n^a$, considering that $n^a s^n = (s\frac{\partial}{\partial s})^a s^n$, the time derivative of the generating function can be written in the form:

$$\begin{aligned}\frac{\partial G(s,t)}{\partial t} &= \sum_k (s^{-k}-1)\sum_a c_k^a \sum_n s^n \frac{n^a}{n!}\frac{\partial^n}{\partial s^n} G(s,t)|_{s=0} \\ &= \sum_k (s^{-k}-1)\sum_a c_k^a (s\frac{\partial}{\partial s})^a \sum_n s^n \frac{1}{n!}\frac{\partial^n}{\partial s^n} G(s,t)|_{s=0} \\ &= \sum_k (s^{-k}-1)\sum_a c_k^a (s\frac{\partial}{\partial s})^a G(s,t)\,.\end{aligned} \tag{2.9}$$

For the birth death process in Eq.(2.4) it reduces to:

$$\frac{\partial G(s,t)}{\partial t} = (1-s)(d-bs)\frac{\partial G(s,t)}{\partial s}\,. \tag{2.10}$$

This equation can be solved by the standard method of characteristics. The characteristic curves in the (s,t)-plane are determined by

$$-\,dt = \frac{ds}{(1-s)(d-bs)}\,. \tag{2.11}$$

By integration, the corresponding expressions for the characteristic curves are calculated

$$C = \frac{1-s}{d-bs}e^{-(d-b)t}\,. \tag{2.12}$$

Along any characteristic curve the variation of the considered generating function is null, then

$$G(s,t) = \Omega(\frac{1-s}{d-bs}e^{-(d-b)t})\,, \tag{2.13}$$

is solution of the evolution equation for any arbitrary function $\Omega(C)$. Assuming that the number of individuals at time zero is $m > 0$, the arbitrary function Ω can be defined to fit the initial condition of the system:

$$G(s,0) = s^m = \Omega(\frac{1-s}{d-bs})\,, \tag{2.14}$$

substituting a fictitious variable $\varsigma = \frac{1-s}{d-bs}$ it results $\Omega(\varsigma) = (\frac{d\varsigma-1}{b\varsigma-1})^m$.

For a generic time $\varsigma = \frac{1-s}{d-bs}e^{-(d-b)t}$ the generating function of the linear birth death process becomes:

$$G(s,t) = \left(\frac{d-bs+d(s-1)(1-e^{-(d-b)t})}{d-bs+b(s-1)(1-e^{-(d-b)t})}\right)^m\,. \tag{2.15}$$

Suppose that the population starts from $m = 1$ the extinction probability corresponds to:

$$P_0(t) = \frac{1-e^{-(d-b)t}}{1-\frac{b}{d}e^{-(d-b)t}}\,, \tag{2.16}$$

while

$$P_1(t) = \left(1-\frac{b}{d}\right)^2 \cdot \frac{e^{-(d-b)t}}{\left(1-\frac{b}{d}e^{-(d-b)t}\right)^2}\,. \tag{2.17}$$

The critical time can be defined as the time for which $P_0(t_c) = P_1(t_c)$.

2.2 The neutral model of ecology of Volkov

Ecological theories can be roughly classified into two big groups: theories related to niche-assembly perspective and theories related to dispersal-assembly perspective [HA04]. Niche-assembly perspective is based on the occupation of specific ecological niches (space and resources) by each species in the community, the abundance of the species in a community is determined by the consequent subdivision of the resources, where each species is adapted to be the best competitor in its niche. Dispersal-assembly theories instead are based on the idea of species turnover, their presence or absence in a community is produced by random dispersal and stochastic local extinction. A notable example is the theory of island biogeography in which equilibrium is determined by a compensation of immigration and extinction events.

By this introduction, it seems that niche-assembly perspective is related to the concept of adaptation and competition between species that is determined by the availability of resources, while dispersal-assembly perspective is related to the random nature of the phenomena involved and the idea that causality has a deep impact in shaping a community.

Lotka-Volterra model of prey-predator competition is a very interesting example. From deterministic equation associated it results that an equilibrium can be found for the system in which the number of individuals of the competitors oscillates in time leading to coexistence of the two species. The number of individuals of a species respect to the other results in a motion over a circular or elliptical orbit. Including fluctuations in the system and switching back to a stochastic description it results that these orbits change in time expanding or shrinking. If by chance number of preys will reach zero, predators will extinguish soon. Vice-versa, if predators extinguish, the number of preys will continue to increase toward infinite. Mechanism of competition and the general behavior of the system can be predicted deterministically, however, the end of the story cannot and it is strictly related to the presence of a random component in the system.

The master equation enters as a powerful mathematical tool to describe dynamical models of ecology in which stochasticity of the process is considered as the principal explanation for the observed patterns. These processes are birth, death, speciation, and migration of the individuals in a community. In ecology, the master equation describes the probability to observe a species with n individuals in a community (the eco-system), i.e. the relative species abundance (RSA). Mechanisms of competition can be included in a multivariable problem, where each variable represents the number of individuals of a species.

Neutral theories belong naturally to dispersal-assembly perspective. They are based on the principle of species equivalence, i.e. all the species undergo the same dynamics, thus the shape of the community is determined by the nature of the stochastic dynamics. Neutrality is generally defined within species of the same trophic level.

A simple but effective neutral theory of ecology is based on the introduction of an immigration parameter in the one-step one-variable birth-death process in Eq.(2.4) [Vol+03]. The master equation of this process reads in its expanded form:

$$\frac{dP_n(t)}{dt} = b_{n-1}P_{n-1}(t) + d_{n+1}P_{n+1}(t) - (b_n + d_n)P_n \,. \tag{2.18}$$

The immigration parameter Υ has been included in the birth term $b_n = b \cdot (n + \Upsilon)$ while the death rate remains $d_n = d \cdot n$.

The presence of the immigration term eliminates the absorbent state and detailed balance condition holds. The stationary solution results:

$$P_n = P_0 \prod_{k=0}^{n-1} \frac{b_k}{d_{k+1}} = P_0 \prod_{k=0}^{n-1} \frac{b \cdot (\Upsilon + k)}{d \cdot (k+1)}\,, \tag{2.19}$$

P_n represents the probability to observe a species with n individuals, the zero class represents the probability to don't observe a species, then to be compared with RSA datasets the solution must be normalized cutting off the zero-class.

The stationary solution results a Negative Binomial distribution

$$P_n = \frac{(1-\frac{b}{d})^{S/b}}{\Gamma(S/b)} \frac{(\frac{b}{d})^n}{n!} \Gamma(n + \frac{S}{b}) = \frac{(1-x)^{\Upsilon}}{\Gamma(\Upsilon)} \frac{(x)^n}{n!} \Gamma(n + \Upsilon)\,, \tag{2.20}$$

where $x = b/d$.

$$P_{RSA} = \langle \Phi_n \rangle = \Theta \frac{x^n}{n!} \Gamma(n + \Upsilon)\,, \tag{2.21}$$

where $\Theta = \frac{S'}{[(1-x)^{-\Upsilon} - 1]\Gamma(\Upsilon)}$ represents the Hubbel biodiversity parameter.

P_{RSA} represent the RSA of a community since it is the probability to find a species having a certain number n of individuals.

The appearance of the negative binomial distribution as stationary distribution is particularly relevant. Negative binomial is able to resemble two typical distribution used to describe RSAs, Log-Normal and Log-Series, introduced by the works of Fisher and Preston. Furthermore, the presence of the immigration parameter works as a density-dependent birth parameter. Let's define $\Upsilon = S/b$, $b_n = n(b + \frac{S}{n})$. The presence of parameter S/n gives a possible interpretation to what is called rare species advantage, i.e rare species have higher spreading capacity respect to the abundant ones.

In [Vol+03] this theory has been applied to model the relations between community and metacommunity organization, in particular, applied to the two different ecosystems of corals in the coral reefs and trees in tropical forests. For details to describe species turnover or probability of extinction of a single species we resend to [Vol+03; Aza+06]

As shown in [TZ13] the RSA of an ecological system can be described by a sum of many distributions to account as an example a niche partitioning, in the case that birth-death rate and immigration are the same in all the niche, the result returns into the previously described neutral model for meta-community [Vol+03].

2.3 The ecosystem of Transposable Elements

Discovered in maize by Barbara McClintock during 1940-1950 (Nobel Price 1983), Transposable Elements (TEs) are also known as selfish DNA or jumping genes. They are present in DNA of eukaryotic and prokaryotic organisms and often constitute a large fraction of many genomes (45% of the human genome, 37% of the mouse genome, more than the 80% in some plants like maize)[MLGP10]. Initially considered junk DNA, they are now known to play fundamental roles in the maintenance of genomic diversity and in the reshaping of gene regulatory networks [KL97],

[Fes08], [Kun+10],[Bié10],[RRM11],[Tes+12],[CO13]. Current TEs activity in humans has been also correlated with several genetic diseases and cancer [GP12],[CB09].

TEs are defined as DNA sequences that can change their position inside the genome through a process called transposition (or retrotransposition if the process is RNA mediated). This process can be replicative, i.e. producing another copy that will insert in another location in the genome (insertion), or not. The replicative process determines the invasion of the host genome with the consequent increase of the genome length. The proteins and all the molecular machinery necessary to transposition can, in principle, be autonomously produced by the TE replicating. If this skill is missing in one element, due to mutation inside its sequence or because the element cannot produce it at all, it is called non-autonomous but in some cases it can still transpose using a compatible machinery produced by other elements. After their first insertion, most of the copies lose the capacity of transposition.

A wide zoology of TE exists and a common nomenclature and phylogenetical classification based on sequence properties has been created to join many separated studies. TEs have been grouped into two large classes depending on their transposition intermediate (RNA or not). Class I TEs are characterized by reverse transcription via RNA intermediate (copy and paste), similar to the retrovirus behavior. These can be divided in many subclasses, in particular, we recall LTR (long terminal repeated) TEs, and non-LTR TEs, LINE and SINE. Class II TEs can transpose directly through cut and paste mechanism, that is often not replicative[MLGP10] [GP12]. Families and sub-families label smaller taxonomic differences between TEs and are defined by phylogenetic sequence identity and because they share specific sequence insertions, deletions or substitutions [Wic+07].

With the improvement of genomic sequencing techniques and the increase of sequencing data comprehensive of repetitive elements, there has been a growing demand for a comprehensive resource of sequence data and other basic information about TEs. The first such resource was the widely known 'RepBase', established in 1992 containing representative sequences and sequence fragments of 53 published human families of interspersed repeats. Extended in the following years it was succeeded by 'Repbase Update' (RU) in 1997. RU, in addition to compiling known elements, began the electronic publication of TEs unreported elsewhere [Jur00]. Furthermore an electronic peer-reviewed journal was launched (http://www.girinst.org) in the same years, still active it is updated monthly. Based on RU many server-based routine sequence analysis have been created for researchers: 'Pythia', CENSOR (http://www.girinst.org/censor/index.php), succeeded now by RepeatMasker (http://repeatmasker.genome.washington.edu) [Jur00]. Raw sequences of many TE and the informations over their insertion coordinates in the genome, similarity and copy numbers are available in such databases and tools.

The dynamics of TEs has been modeled on the basis of their transposition and excision (elimination from the genome) rates and their fitness effects on the host [CL89][VFB09] in a germline. The analogy between an ecological system and transposable elements has been proposed in [VFB09]. If a copy of a TE is considered as an individual, one TE species comprises closely genetically related TE copies that share the same interaction with their environment. The community of a genome contains all the copies of TEs irrespective of their subfamilies, families or classes, and it is analogous to the biotic portion of an ecosystem. The abiotic component is composed of the genes and various kind of non-coding sequences and the intracellular environment. The species richness can be defined as the number of TE species within a genome and the relative species abundance of a TE species as the number

of copies of that element relative to the total number of copies of all the elements present in that genome.

In the system of TEs, we consider birth (replication), death (excision) and immigration (horizontal transfer and mediation in time over possible interactions) as the random processes shaping the community. Recently the analogy between the distribution of RSA in ecological communities and the distribution of relative abundance of genetic sequences has been used to describe the dynamics of the population of 'genetic species' in the genome (long terminal repeat (LTR) retrotransposons, non-LTR retrotransposons, cut-and-paste DNA transposons, rolling-circle DNA transposons, self-synthesizing DNA transposons, satellites, simple repeats, tRNA, miRNA, snoRNA) and to test neutrality [SBD13].

The effect over the fitness of the population of TEs copy number has been considered in many models to study how TEs copies reach fixation in a population and what kind of mechanisms shape TEs abundances [CL89], [SKR05],[DCB05], [AK06], [LC05], [LC06], [ALC07]. In these approaches the equilibrium between selection pressure and birth rate of TEs as well as self-regulation mechanisms are generally considered to limit the number of insertions. Copies that reach fixation should have a neutral or eventually positive effect on the genome [BF01]. The phenomenon of domestication, i.e. the recruitment by the cell of some TEs to carry out tasks related to their activity as the encoded transposition proteins or regions regulating their expression in the genome, is widely observed in nature [Vol06][RRM11].

Contrasted patterns and variation in the amount and diversity of TEs may reflect some organization due to host specific selection pressure, both at genomic and population level, or be the result of stochastic forces at the level of the individual copies. The interdependence between TEs and the host genome and the replication mechanisms of the Elements suggest a strong parallelism between TEs dynamics in the genome and species community dynamics in their ecosystem [VFB09],[SBD13]. Both the niche theory, based on the partitioning of resources, space and time of action between competing species [CL03], and the neutral theory, in which stochastic mechanisms as demographic stochasticity, migration, and speciation are the most important forces shaping the community [HA04], have features suitable to describe the TEs ecosystem. Despite that, TEs ecosystem contains some peculiarities that differentiate it from standard ecosystems [VFB09]. TEs create and continuously shape their own environment because death copies, i.e. TE copies that lose any transposition ability, are the major part of the genomic landscape in which new copies may insert without deleterious effect on the cell functionalities. The selection at the level of the host induces TEs to evolve traits that constitute a selective disadvantage at the individual level, as for example a lower transposition rate [HB04][ALC07]. Mechanisms related to the molecular nature of TEs may occur.

Long Interspersed Elements (LINEs) are the most abundant class of TEs in mammals. They belong to the retroelements class, i.e. their replication is RNA mediated, as will be discussed later. Although some elements are still active in genomes with a potentially high impact, the number of active copies is very small compared to the population of inactive elements. In humans only L1Hs LINE Element is currently active, [al.01],[Wei+01],[MPC15]. Here we focus on the modeling of LINEs genomic distribution under the hypothesis of *competitive neutrality* [Lin+15], i.e. the absence of competitive differences among different types of entity. It means that all the copies of all elements in the community (the host genome) are characterized by the same transposition activity, sequence divergence, death rate [VFB09].

LINEs ecosystem in mammals results particularly suitable to be described under the neutral hypothesis. LINEs evolved often on a single lineage, in particular in

Primates [KSB06], with a subsequent appearance of active Elements, making competition between different Elements negligible. Coexistence of multiple L1 families or lineages is documented for ancient LINEs [Smi+95] and currently in mouse [MH01], where L1 frequently recruited novel 5'UTR sequences [SaB13], suggesting that simultaneous activity of non-homologous promoters does not introduce a competition between the elements. The genomic environment is unique to each of the TEs copy. Full-length L1 copies may differ randomly in their level of transposition activity [Bro+03],[Sel+06], and random processes at the individual level strongly impact the structure of the entire community. This stochasticity supports the neutral approach to describe the community dynamics.

LINEs retro-transposition

LINEs replication is RNA mediated. Full-length elements contain a promoter region (5'UTR), two protein coding regions (ORF1, ORF2) and a poly-A tail (3'UTR). The internal promoter allows a retrotransposon to generate autonomous duplicate copies at multiple locations in the genome. RNA polymerase II-mediated transcription of a genomic locus from an internal promoter that directs transcription initiation at the 5' boundary of the element. The produced RNA is exported to the cytoplasm, in which ORF1 (which encodes an RNA-binding protein) and ORF2 (which encodes a protein with endonuclease and reverse-transcriptase activities) are translated. Both proteins show a strong cis-preference; consequently, they preferentially associate with the RNA transcript that encoded them to produce a ribonucleoprotein (RNP) particle. After coming back into the nucleus the proteins on RNA can open a nick in DNA and produce a DNA copy of the template through a process termed target-primed reverse transcription (TPRT). The resulting new insertion is a low fidelity copy of the parent LINE element with frequent 5' truncations, often losing replication capacity.

We start our analysis by developing the most parsimonious model of neutrality to which deviation from neutrality can be added and test their predictions on genomic data. We modeled the way LINEs populated mammalian genomes over the course of evolution as a birth-death process of two interacting species: full-length (active) and incomplete (inactive) LINE copies. The number of active and inactive copies of one Element in the host genome is represented by n_A and n_I respectively. Active copies can generate new active LINEs by retrotransposition at a rate equal to $b_A n_A$ (birth process of active copies). Over the course of time, mutations and the host selection pressure inactivate active copies at a rate equal to $d_A n_A$ (death process of active copies). Some transposition events are incomplete such that the inserted copy is incapable of autonomous retrotransposition; for example, L1 insertions are often 5'-truncated (e.g. Figure 6B of [Cri+14]). We refer to the rate at which this process occurs as $b_I n_A$. Furthermore, a transcribed incomplete copy can hijack the retrotransposition machinery of autonomous copies to duplicate in a new location, a process called trans-complementation that can occur at a rate equal to $b_{AI} n_A n_I$. This phenomenon has been observed, for example, in LINE-1 retroElements, although it should happen at a much smaller rate than retrotransposition in cis [Wei+01]. Then, the birth rate of incomplete copies is given by $(b_{AI} + \frac{d_A + b_I}{n_I}) n_A n_I$ (birth process of

inactive copies). The host selection may also prevent the fixation of copies that affect negatively its fitness reducing this rate. Finally, when many mutations accumulate, inactivated copies cannot be distinguished from random background genomic sequences and they essentially disappear from the genome. This can also occur through excision of large genomic regions as well as structural rearrangements. we refer to this process as the death rate of the incomplete copies d_I. This birth-death stochastic process can be described via the following two-dimensional master equation by using the step operators $E^{\pm} f(n) = f(n \pm 1)$ [Kam81]:

$$\begin{aligned} \frac{dP(n_A, n_I, t)}{dt} &= (E^-_{n_A} - 1) b_A n_A P(n_A, n_I, t) + (E^+_{n_A} E^-_{n_I} - 1) d_A n_A P(n_A, n_I, t) \\ &+ (E^-_{n_I} - 1) b_I n_A P(n_A, n_I, t) + (E^-_{n_I} - 1) b_{AI} n_A n_I P(n_A, n_I, t) \\ &+ (E^+_{n_I} - 1) d_I n_I P(n_A, n_I, t)\,, \end{aligned} \tag{2.22}$$

where the lower index in the step operator indicates the variable within the probability distribution on which the operator acts. Each term in the sum on the right-hand side of Eq.(2.23) models one of the biological processes described abundance in the ecosystem under study. The asymptotic stationary solution $P(n_I)$ represents the RSA (P_{RSA}) of the system, i.e. the probability to observe a species with a certain number of individuals in a community.

When equilibrium is reached for both active and inactive copies for enough long time the stationary solution for the RSA is a negative binomial distribution. In fact, taking n_A as a constant the equation for n_I reduces to the neutral model of ecology described in [Vol+03]. A special case of the model is obtained when excision and trans-complementation processes are neglected. In the case that equilibrium for active species does not hold ($b_A << d_A$) the asymptotic solution P_{RSA} is again a negative binomial distribution when the absorbent state for active copies is reached, with the notable difference that expected values of the parameters is different from the previous case. If trans-complementation does not produce a relevant contribution and the system is out of equilibrium the RSA is described by a negative binomial distribution with a number of failure parameter of order one ($\Upsilon \sim 1$). Instead, if the trans-complementation process is relevant the number of failures should be about $\Upsilon \sim (b_I + d_A)/b_{AI} >> 1$, much larger then one because up to literature trans-complementation events should be rarer than retrotransposition events in cis.

Negative binomial distribution is able to resemble two typical distribution used to describe RSAs in ecology: Log-Normal and Log-Series, widely discussed in ecology [HA04]. The distribution is normalized to one after removing the probability to observe zero individuals (the unobserved species). Mathematical details of the derivation and numerical simulations can be found in the Section 2.4.

We proposed competition between two Elements activated by the same promoter region as a stochastic process which may determine a deviation from the expected distribution, generating a bimodal behavior described by a mixture of negative binomials. The probability P_{poly} of a single polymerase to act on two LINE species activated by the same promoter region with n_1 and n_2 number of active copies respectively, is rescaled by the fraction of active copies belonging to that specie: $P_{poly,1} = P_{poly} \frac{n_1}{n_1+n_2}$, and vice-versa. The transposition rates b_A, b_I and b_{AI}, result to be rescaled as well by the same factor. Lower birth rate parameters will result in a lower copy number of LINEs in competition respect to Elements with full rate available. Neutrality hypothesis is maintained because all LINEs species are equivalent in the model, the accidental activation by the same promoter region introduces a disadvantage for the species that compete for the molecular machinery, breaking

temporarily the neutral assumption, until the extinction by chance of one of the competitors. The distribution arising from competition is a mixture of two negative binomials, where the mixture coefficient is related to the probability that two Elements compete.

We will show that the neutral model proposed is able to describe the general trend of the RSA in the 42 Mammalian genomes considered. However, the mixture model results more attractive to describe LINEs community because allows distinguishing host species belonging to different taxonomic orders. Furthermore, the hypothesis of competition between promoters has been supported by the evidence that rare LINEs active in the same age presents a higher level of similarity in the 5'UTR respect to the other Elements. By knowledge of the order of activity of many Elements [Gio+07], the evolution of the RSA through a sliding window method is also studied. It results that both in Primates that in Murinides a transition can be observed at the time of radiation. Chromatin state characterization of LINEs copies in human and mouse has also been considered. It will be shown that in human a gradual transition to less abundant Elements is observed while in mouse two well-separated cluster can be identified, one of them associated with LINEs active since the murine radiation.

2.4 The model

According to what is explained in the previous section we can model the birth and death stochastic process via the following two-dimensional master equation:

$$\begin{aligned}\frac{dP(n_A,n_I,t)}{dt} &= (E^-_{n_A}-1)b_A n_A P(n_A,n_I,t) + (E^+_{n_A}E^-_{n_I}-1)d_A n_A P(n_A,n_I,t)\\ &+ (E^-_{n_I}-1)b_I n_A P(n_A,n_I,t) + (E^-_{n_I}-1)b_{AI} n_A n_I P(n_A,n_I,t)\\ &+ (E^+_{n_I}-1)d_I n_I P(n_A,n_I,t)\,,\end{aligned} \tag{2.23}$$

where the lower index in the step operator indicates the variable within the probability distribution on which the operator acts. Despite that inactive element dynamics is conditioned by the presence of some active copies, their copy number is *not time-correlated* because number of active copies will perform a random walk until extinction while inactive copies accumulate in the genome. Hence we can apply the *mean field* approach (see Appendix A for details) to derive the following deterministic equations from Eq.(2.23):

$$\begin{cases}\frac{dn_A}{dt} = b_A\cdot n_A - d_A\cdot n_A\\ \frac{dn_I}{dt} = (b_I+d_A)\cdot n_A + b_{AI}\cdot n_A\cdot n_I - d_I\cdot n_I\end{cases} \tag{2.24}$$

Active copies are described by a pure birth-death process with absorbing state $n_A = 0$, while inactive copies undergo a birth-death process with an external influx $S = (b_I + d_A)\cdot n_A$.

The dynamics of active copies and inactive copies happen at different timescales. Respect to defective copies, active copies have higher probability to be expressed in the genome ($b_A \gg b_{AI}$), they are more sensitive to mutations, which may lead to inactivation, and they are more subjected to host selection pressure and defense mechanisms, like silencing, ($d_A \gg d_I$)[BF01].

Mean field n_A is expected to converge to the absorbent slowly because LINEs can persist in genomes for millions of years [AK06], generating relatively constant

numbers of new insertions over time. Assuming equilibrium between new insertions and host defenses, $\frac{dn_A}{dt} \approx 0$, the two differential equations can be reduced to the second one, that take the form:

$$\frac{dn_I}{dt} = b \cdot n_I - d \cdot n_I + S\,, \tag{2.25}$$

where $b = b_{AI} n_A$ and $d = d_I$ are constant birth and death rates respectively, and $S = (b_I + d_A) n_A$ can be seen as a constant external influx. The stationary solution $P(n_I)$ of the corresponding master equation considered in [Vol+03] is a negative binomial and represent the RSA of the ecosystem as shown in Eq.(2.21).

This approach allows a simple and straightforward interpretation of the parameters involved in the dynamics however it is necessary to stress a few aspects. The time variable involved in the equation does not represent an absolute time but contains a common time scale for the eco-genome-system under study. Defective copy and active copy are in competition for the molecular machinery, then take the parameters b_{AI}, b_A, b_I equal to constants means a perfect cis-preference of the active copies, which are always able to maintain their machinery for their own with the same rate. This assumption can be relaxed introducing a competition of the Michaelis-Menten kind with a weight to mime cis-preference.

The driving process behind the dynamics until trans-complementation becomes relevant is an accumulation of Poisson type with an expectation value of the order of the time passed by each state of the active element ($T(n_A = 1, 2, ...\infty)$) multiplied by the rate of birth of defective copies ($(b_I + d_A) n_A$). The key point to distinguish the pure influx model from the birth-death with influx model in Eq.(2.25) is the assumption $\frac{dn_A}{dt} \approx 0$, in order to leave the system on for enough time to reach the stationary state. The derivation of the analytical solution of the time-dependent most general process is not straightforward, however numerical simulation can be performed to check the asymptotic solution to be a negative binomial for some interesting set of parameters. The study of the time-dependent evolution of the system is beyond the scope of the present work since we are interested in the characterization of the RSA, composed by the population of extinct LINEs, and then described by the asymptotic solution by definition.

The survival probability $p(\tau)$ of the active population correspond to the survival probability of a pure birth-death model, and can then be find in most of standard textbooks of stochastic processes as example [Kam81].

For the initial state $n_A(t = 0) = 1$ it reads :

$$p(\tau) = 1 - \frac{1 - e^{(d_A - b_A)\tau}}{1 - (b_A/d_A) e^{(d_A - b_A)\tau}}\,. \tag{2.26}$$

In the limit of $b_A << d_A$ the survival probability is approximated by an exponential. This corresponds to a Gamma distribution with unitary shape parameter ($\Upsilon \sim 1$). The expectation number of insertions after extinctions for a pure accumulation process (inactive copies does not replicate themselves or die) can be then approximated to a Gamma-Poisson distribution, i.e. a Negative Binomial distribution.

In conclusion, if trans-complementation does not produce a relevant contribution we expect to observe an RSA following a negative binomial distribution with a number of failure parameter of order one. This is not the case of the distribution arising from Eq.2.25 since the number of failures corresponds to the influx parameter, which should be much larger than one because trans-complementation is much

rarer than regular retrotransposition events ($b_I >> b_{AI}$).

2.4.1 Mechanism of competition

Consider two LINE species activated by the same promoter region with n_1 and n_2 number of active copies respectively. The probability P_{poly} of a single polymerase to act on one of the two species in competition, starting the process of transposition, is rescaled by the fraction of active copies belonging to that species: $P_{poly,1} = P_{poly}\frac{n_1}{n_1+n_2}$, and vice-versa. The transposition rates, b_A, b_I and b_{AI}, result to be rescaled by the same factor $\frac{n_1}{n_1+n_2} < 1$, determining an asymmetric disadvantage for the two competing species, if $n_1 \neq n_2$ when competition starts.

Neutrality assumption is maintained at the bases. When two elements in competition compare occasionally, this mechanism acts directly on the birth rate parameters, breaking naturally neutrality assumption. Single lineage and frequent change of promoter is favored because both species involved are penalized by the competition, and a new arising species is penalized as much as possible if the other is yet abundant. After the extinction of the competitor, the winner returns to have the whole availability of the protein and possibly became abundant.

At the light of these considerations, we expect to distinguish two distributions, one containing the *losers* or (rare species) of the competition and information about the competition process, and one containing the *winners* (or abundant species) which is expected to be comparable with the neutral one.

2.4.2 Numerical simulations of the model

To test if the dynamical model can generate a negative binomial type distribution beyond the given assumption we performed numerical simulations. We used Gillespie algorithm [Gil76] for the active copies dynamics, for the inactive copies dynamics, we used the tau-leap algorithm [Gil01] in the case of a pure accumulation process and a hybrid algorithm to simulate the dynamics of inactive copies with trans-complementation. An hybrid algorithm instead of a Gillespie was chosen to reduce the time of computation, it consists in the estimation of the expected number of inactive copies by ODE numerical integration to estimate the expected birth and death rates in that time interval, and use a tau-leap algorithm to generate the increment associated to that interval. Oracle comparison to the theoretically correct Gillespie algorithm was performed (Fig. 2.1).

We generated LINE ecosystems according to the model proposed starting from a single active copy for each Element. Competition between different elements has been also introduced according to the model previously described and compared with the neutral one. In all the simulations performed in red are shown the occupation states of the active copies and in green the expected number of inactive copies in the last time step before the extinction of active copies. If the green lines don't converge to the same value it means that equilibrium is not reached in the inactive copies dynamics. The cumulative of the distribution obtained have been qualitatively compared with the LINE one in H. sapiens.

We see that when $b_A << d_A$ (Fig. 2.2) we obtain an RSA compatible with a negative binomial as expected. In the other case we have $b_A \simeq d_A$ (Fig. 2.3) and deviations from negative binomial distribution are significant. Despite the asymptotic behavior is a negative binomial when the number of elements is low we may observe fluctuation that can be interpreted as a mixture.

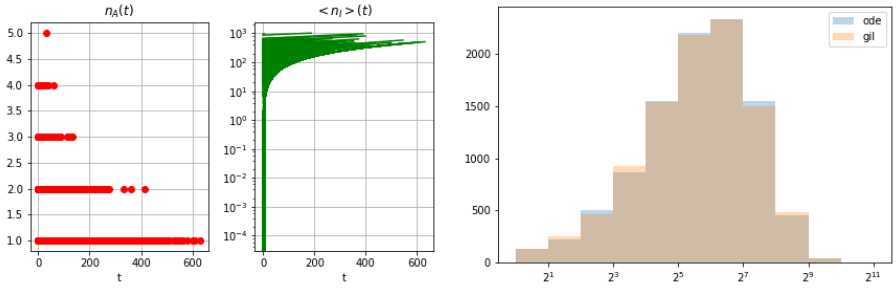

FIGURE 2.1: Example of oracle comparison of the hybrid algorithm with the Gillespiel algorithm.

In Fig. 2.4 we simulated the presence of competition for the promoter, a low probability of arising of competitors is used. At each time step of the simulation a competitor may arise with the defined probability, species that live longer have a higher probability to find a competitor but they result advantaged because with a higher probability their active copies are more than one.

The deviation introduced is not enough to generate bimodality if the competition starts in a symmetric configuration, we may consider that a competition of this kind can become relevant when the competing species arises while the other is already abundant, suppressing the rise of new elements.

2.5 Data sources

LINE abundances were calculated using RepeatMasker annotation (http://www.repeatmasker.org) for human genome build hg19 and 45 other Mammalian species. LINE consensus sequences were downloaded from RepBase [Jur00; Jur+05] (http://www.girinst.org). Only a subset of the LINE consensus sequences contain the 5′ UTR, noted in the RepBase associated report, which were selected for the analysis of LINE 5′ UTR sequence. Chronological ordering of LINEs in human, Chimpanzee, Rhesus Macaque, mouse and rat was derived from Giordano et al. [Gio+07]. Chromatin structure data available for mouse [Yue+14] and human [EK10] were used to assign genomic copies of LINEs to open and closed chromatin states by knowledge of their coordinates in the reference genome. In the cited references chromatin structure assignment was conducted using ENCODE chromatin models using the ChromHMM method [EK12].

2.6 ABC method implementation and discrimination between the models proposed

Approximate Bayesian Computation (ABC) is a useful method to find the distribution of optimized parameters of the tested model and to check directly the validity of a model respect to another. For each designed model M, priors are assigned to the parameters of the model. About 5 million simulations are performed in which a set of parameters $\widetilde{\theta}$ is extracted from the priors, defining $M(\widetilde{\theta})$. A dataset is then built from the defined distribution and compared with the real data, after that $\widetilde{\theta}$ is accepted or rejected. The ensemble of accepted $\widetilde{\theta}$ constitute the posterior of the parameters. Rejection is done fixing a maximum accepted distance ϵ between real data distribution and the simulated one.

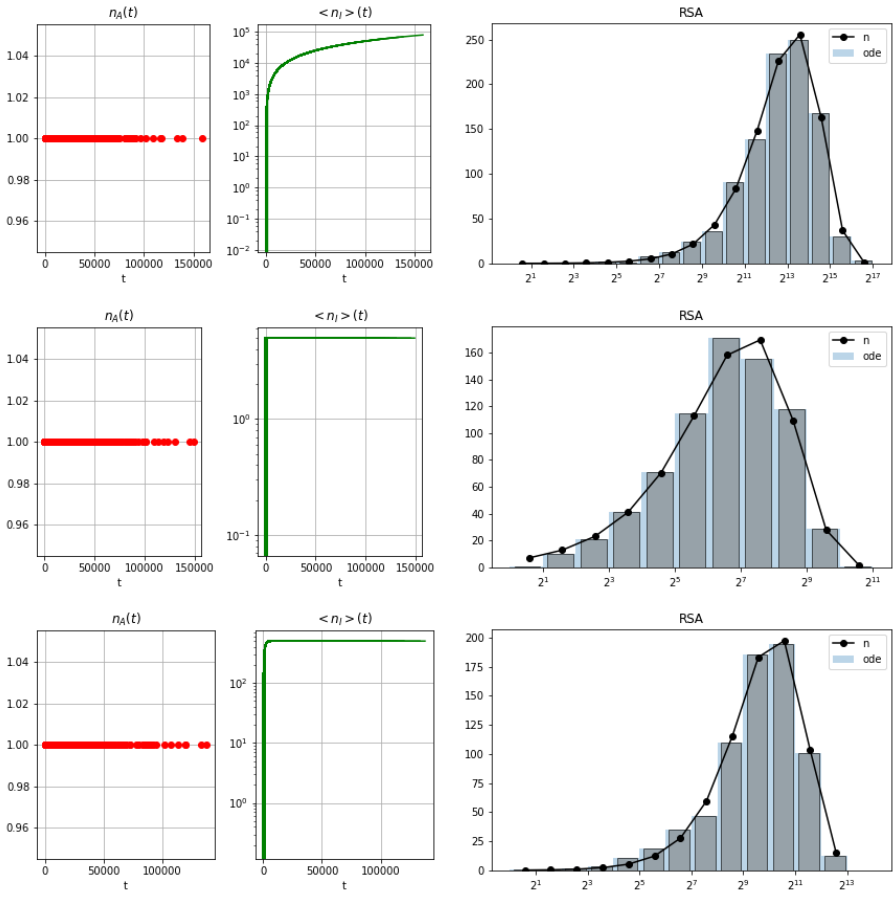

FIGURE 2.2: upper panels: $b_A, d_A, b_I, b_{AI}, d_I = 0.0, 0.00005, 0.5, 0, 0$; middle panels: $b_A, d_A, b_I, b_{AI}, d_I = 0.0, 0.00005, 0.5, 0.0, 0.1$; lower panel: $b_A, d_A, b_I, b_{AI}, d_I = 0.0, 0.00005, 0.5, 0.099, 0.1$.

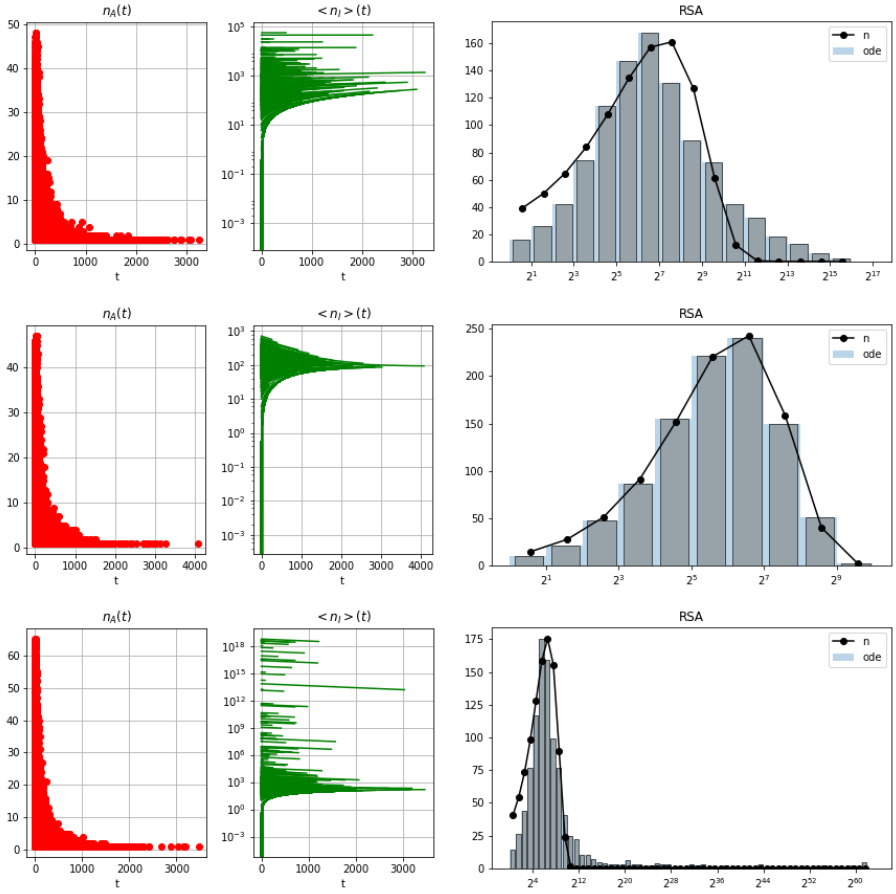

FIGURE 2.3: upper panels: $b_A, d_A, b_I, b_{AI}, d_I = 0.001, 0.0011, 0.09, 0, 0$; middle panels: $b_A, d_A, b_I, b_{AI}, d_I = 0.001, 0.0011, 0.09, 0, 0.001$; lower panels $b_A, d_A, b_I, b_{AI}, d_I = 0.001, 0.0011, 0.09, 0.0005, 0.001$.

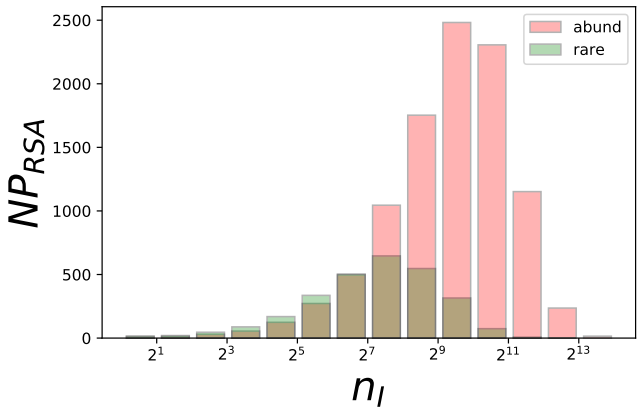

FIGURE 2.4: **Simulations of competition in accumulation model.** Simulation of RSA of size of 10^4 elements. Simulation of a process with $b = 0.0001$, $d = 0.0011$. At each step of the simulation with a Gillespiel algorithm a competitor may arise with $p = 0.25$. The resulting distribution have means around 10 and 8 log2 of the counts respectively.

To define efficient but general priors a preliminary procedure has been performed: very wide priors are defined and a new prior is built from the union of the posteriors of all the species tested. In this way the space of parameters is restricted to the most useful part, saving a lot of computation effort. Then a sampling is done and distribution of optimized parameters is defined for each sample tested. The rejection is defined by the value of two sample chi square significance , samples with $\chi_s > 0.5$ are rejected, and by a threshold (0.3) of the absolute difference between the number of elements in each class n with population 2^n. This additional score reduces the average of the chi square significance to lower values.

The probability of the model given the data is:

$$P(M|D) = P(D|M)P(M)\,, \tag{2.27}$$

in ABC $P(D|M)$ can be approximated by the ration of the number of successes (number of accepted $\widetilde{\theta}$) over the number of tentatives.

Then the ratio probability of a model respect to another is:

$$\frac{P(M_1|D)}{P(M_2|D)} = B_{1,2}\frac{P(M_1)}{P(M_1)}\,, \tag{2.28}$$

where $B_{1,2}$ is approximated by the ratio of positives over tentatives of each model.

The computational cost of ABC method could be relevant, in particular the sampling of the populations per each parameters. The strategy chosen is to collect and save a large sample of the populations ($\gtrsim 10^6$), each of them generated given a set of parameters from very wide priors. Then this sample is compared to each RSA of LINEs elements of the host genome under study, following the ABC procedure explained previously. The parameters from the priors that pass the test will form the posteriors. This procedure is repeated a second time generating a new prior from the union of the posteriors for all the RSA under study. This hierarchical procedure reduces the space of parameters to the most useful part, enhancing the probability to get good parameters in the second ABC step. It is interesting that 4 more species, i.e. Wallaby, Tasmania Devil, Opossum and Platypus, get zero parameters from the same priors used for the other 42 Mammals. These organisms have in fact a much lower number of LINE species in their genomes, which could affect the score results, furthermore they lie quite distant in the phylogenetic tree respect to all the others. For this reason they have been excluded in the rest of the analysis.

2.7 Results and discussion

The RSA of LINEs in 46 Mammals genomes have been fit by a negative binomial and a mixture of negative binomial through approximate Bayesian computation method (ABC) to test neutral and competition models respectively, mean parameters shown in Tables 2.1, 2.2 and 2.3. From the same priors we described 42 datasets, with the notable exception of Wallaby, Tasmania devil, Opossum and Platypus, which are the most isolated species of the group from an evolutive point of view.

sampleID	mean	n	p	chi_s	pos/tot
Alpaca	$7.59{\cdot}10^{3}$	0.47	$6.46{\cdot}10^{-5}$	$2.45{\cdot}10^{-2}$	0.0350
Armadillo	$8.19{\cdot}10^{3}$	0.37	$4.72{\cdot}10^{-5}$	$5.74{\cdot}10^{-2}$	0.0109
Baboon	$6.83{\cdot}10^{3}$	0.63	$9.53{\cdot}10^{-5}$	$9.01{\cdot}10^{-2}$	0.0538
Bushbaby	$6.02{\cdot}10^{3}$	0.49	$8.62{\cdot}10^{-5}$	$7.52{\cdot}10^{-2}$	0.1524
Cat	$8.18{\cdot}10^{3}$	0.44	$5.62{\cdot}10^{-5}$	$2.19{\cdot}10^{-2}$	0.1116
Chimpanzee	$7.41{\cdot}10^{3}$	0.65	$9.05{\cdot}10^{-5}$	$8.74{\cdot}10^{-2}$	0.0385
Cow	$7.30{\cdot}10^{3}$	0.47	$6.75{\cdot}10^{-5}$	$4.98{\cdot}10^{-2}$	0.0534
Dog	$7.91{\cdot}10^{3}$	0.44	$5.85{\cdot}10^{-5}$	$2.74{\cdot}10^{-2}$	0.0614
Dolphin	$7.41{\cdot}10^{3}$	0.65	$9.05{\cdot}10^{-5}$	$8.74{\cdot}10^{-2}$	0.0385
Elephant	$8.10{\cdot}10^{3}$	0.32	$4.20{\cdot}10^{-5}$	$1.90{\cdot}10^{-2}$	0.0056
Ferret	$7.28{\cdot}10^{3}$	0.44	$6.39{\cdot}10^{-5}$	$1.72{\cdot}10^{-2}$	0.0558
Gibbon	$6.86{\cdot}10^{3}$	0.62	$9.38{\cdot}10^{-5}$	$7.42{\cdot}10^{-2}$	0.0490
Gorilla	$7.03{\cdot}10^{3}$	0.65	$9.57{\cdot}10^{-5}$	$7.69{\cdot}10^{-2}$	0.0598
Guineapig	$8.35{\cdot}10^{3}$	0.59	$7.29{\cdot}10^{-5}$	$1.15{\cdot}10^{-2}$	0.0172
Hedgehog	$5.18{\cdot}10^{3}$	0.42	$8.80{\cdot}10^{-5}$	$5.11{\cdot}10^{-3}$	0.0272
Horse	$8.83{\cdot}10^{3}$	0.55	$6.42{\cdot}10^{-5}$	$1.34{\cdot}10^{-2}$	0.0006
Human	$7.70{\cdot}10^{3}$	0.59	$7.93{\cdot}10^{-5}$	$4.48{\cdot}10^{-2}$	0.0451
Kangaroorat	$4.71{\cdot}10^{3}$	0.46	$1.04{\cdot}10^{-4}$	$1.19{\cdot}10^{-2}$	0.0067
Manatee	$7.63{\cdot}10^{3}$	0.38	$5.26{\cdot}10^{-5}$	$3.03{\cdot}10^{-2}$	0.0053
Marmoset	$6.99{\cdot}10^{3}$	0.53	$7.95{\cdot}10^{-5}$	$7.03{\cdot}10^{-2}$	0.0562
Megabat	$7.01{\cdot}10^{3}$	0.46	$6.87{\cdot}10^{-5}$	$6.19{\cdot}10^{-3}$	0.0169
Microbat	$6.71{\cdot}10^{3}$	0.45	$7.11{\cdot}10^{-5}$	$1.42{\cdot}10^{-2}$	0.0177
Mouse	$6.25{\cdot}10^{3}$	0.45	$7.47{\cdot}10^{-5}$	$5.16{\cdot}10^{-2}$	0.0166
Mouselemur	$5.38{\cdot}10^{3}$	0.42	$8.23{\cdot}10^{-5}$	$8.78{\cdot}10^{-2}$	0.0028
Nakedmole-rat	$7.50{\cdot}10^{3}$	0.49	$6.84{\cdot}10^{-5}$	$2.54{\cdot}10^{-2}$	0.0110
Orangutan	$9.13{\cdot}10^{3}$	0.69	$7.73{\cdot}10^{-5}$	$2.81{\cdot}10^{-2}$	0.0143
Panda	$8.24{\cdot}10^{3}$	0.39	$4.94{\cdot}10^{-5}$	$1.26{\cdot}10^{-2}$	0.0016
Pig	$7.75{\cdot}10^{3}$	0.48	$6.49{\cdot}10^{-5}$	$3.31{\cdot}10^{-2}$	0.1152
Pika	$4.66{\cdot}10^{3}$	0.50	$1.13{\cdot}10^{-4}$	$5.33{\cdot}10^{-3}$	0.0077
Rabbit	$7.72{\cdot}10^{3}$	0.52	$7.02{\cdot}10^{-5}$	$6.91{\cdot}10^{-3}$	0.0309
Rat	$6.39{\cdot}10^{3}$	0.42	$6.78{\cdot}10^{-5}$	$5.73{\cdot}10^{-2}$	0.0042
RhesuSrheMac3	$6.82{\cdot}10^{3}$	0.64	$9.70{\cdot}10^{-5}$	$6.50{\cdot}10^{-2}$	0.0576
RockhyraXproCap1	$7.30{\cdot}10^{3}$	0.37	$5.36{\cdot}10^{-5}$	$9.19{\cdot}10^{-3}$	0.0093
SheePoviAri1	$4.60{\cdot}10^{3}$	0.43	$1.05{\cdot}10^{-4}$	$4.51{\cdot}10^{-2}$	0.0014
Shrew	$5.07{\cdot}10^{3}$	0.44	$9.34{\cdot}10^{-5}$	$6.83{\cdot}10^{-3}$	0.0103
Sloth	$8.21{\cdot}10^{3}$	0.30	$3.89{\cdot}10^{-5}$	$1.35{\cdot}10^{-2}$	0.0021
Squirrel	$7.16{\cdot}10^{3}$	0.52	$7.54{\cdot}10^{-5}$	$1.97{\cdot}10^{-2}$	0.1219
Squirrelmonkey	$6.15{\cdot}10^{3}$	0.58	$9.86{\cdot}10^{-5}$	$1.26{\cdot}10^{-1}$	0.0834
Tarsier	$7.34{\cdot}10^{3}$	0.51	$7.29{\cdot}10^{-5}$	$3.82{\cdot}10^{-2}$	0.0623
Tenrec	$5.11{\cdot}10^{3}$	0.40	$8.26{\cdot}10^{-5}$	$9.50{\cdot}10^{-3}$	0.0043
Treeshrew	$6.80{\cdot}10^{3}$	0.50	$7.80{\cdot}10^{-5}$	$8.26{\cdot}10^{-3}$	0.0173
Whiterhinocero	$7.60{\cdot}10^{3}$	0.53	$7.25{\cdot}10^{-5}$	$2.13{\cdot}10^{-2}$	0.0505

TABLE 2.1: Mean parameters obtained by the ABC method fitting the neutral model.

sampleID	mean1	n1	p1	mean2	n2	p2	a	chi_s	pos/tot
Alpaca	$8.89{\cdot}10^{3}$	0.69	$7.95{\cdot}10^{-5}$	$7.26{\cdot}10^{2}$	1.02	$2.26{\cdot}10^{-3}$	0.18	$2.60{\cdot}10^{-2}$	0.0203
Armadillo	$9.48{\cdot}10^{3}$	0.49	$5.31{\cdot}10^{-5}$	$6.57{\cdot}10^{2}$	0.92	$2.46{\cdot}10^{-3}$	0.17	$5.08{\cdot}10^{-2}$	0.0062
Baboon	$7.94{\cdot}10^{3}$	1.00	$1.28{\cdot}10^{-4}$	$8.08{\cdot}10^{2}$	1.05	$2.14{\cdot}10^{-3}$	0.17	$1.07{\cdot}10^{-1}$	0.0401
Bushbaby	$7.71{\cdot}10^{3}$	0.78	$1.04{\cdot}10^{-4}$	$7.30{\cdot}10^{2}$	1.00	$2.17{\cdot}10^{-3}$	0.22	$7.00{\cdot}10^{-2}$	0.0912
Cat	$9.37{\cdot}10^{3}$	0.63	$6.85{\cdot}10^{-5}$	$7.19{\cdot}10^{2}$	0.90	$2.27{\cdot}10^{-3}$	0.17	$2.58{\cdot}10^{-2}$	0.0521
Chimpanzee	$8.54{\cdot}10^{3}$	1.09	$1.30{\cdot}10^{-4}$	$7.32{\cdot}10^{2}$	1.21	$2.44{\cdot}10^{-3}$	0.17	$8.85{\cdot}10^{-2}$	0.0496
Cow	$8.63{\cdot}10^{3}$	0.67	$8.04{\cdot}10^{-5}$	$7.59{\cdot}10^{2}$	0.98	$2.12{\cdot}10^{-3}$	0.18	$5.36{\cdot}10^{-2}$	0.0295
Dog	$9.09{\cdot}10^{3}$	0.65	$7.32{\cdot}10^{-5}$	$6.94{\cdot}10^{2}$	0.89	$2.41{\cdot}10^{-3}$	0.17	$3.33{\cdot}10^{-2}$	0.0278
Dolphin	$8.54{\cdot}10^{3}$	1.09	$1.30{\cdot}10^{-4}$	$7.32{\cdot}10^{2}$	1.21	$2.44{\cdot}10^{-3}$	0.17	$8.85{\cdot}10^{-2}$	0.0496
Elephant	$9.54{\cdot}10^{3}$	0.46	$4.95{\cdot}10^{-5}$	$5.31{\cdot}10^{2}$	0.87	$3.22{\cdot}10^{-3}$	0.19	$1.78{\cdot}10^{-2}$	0.0042
Ferret	$8.65{\cdot}10^{3}$	0.63	$7.48{\cdot}10^{-5}$	$7.22{\cdot}10^{2}$	0.97	$2.32{\cdot}10^{-3}$	0.18	$1.94{\cdot}10^{-2}$	0.0299
Gibbon	$8.03{\cdot}10^{3}$	1.02	$1.29{\cdot}10^{-4}$	$7.65{\cdot}10^{2}$	1.15	$2.28{\cdot}10^{-3}$	0.18	$7.76{\cdot}10^{-2}$	0.0446
Gorilla	$8.20{\cdot}10^{3}$	1.10	$1.36{\cdot}10^{-4}$	$7.59{\cdot}10^{2}$	1.17	$2.31{\cdot}10^{-3}$	0.17	$8.19{\cdot}10^{-2}$	0.0691
Guineapig	$9.46{\cdot}10^{3}$	1.03	$1.10{\cdot}10^{-4}$	$6.64{\cdot}10^{2}$	1.29	$2.80{\cdot}10^{-3}$	0.17	$1.30{\cdot}10^{-2}$	0.0236
Hedgehog	$7.54{\cdot}10^{3}$	0.56	$7.81{\cdot}10^{-5}$	$8.54{\cdot}10^{2}$	0.87	$1.70{\cdot}10^{-3}$	0.27	$3.86{\cdot}10^{-3}$	0.0150
Horse	$9.80{\cdot}10^{3}$	1.00	$1.03{\cdot}10^{-4}$	$6.19{\cdot}10^{2}$	0.92	$2.35{\cdot}10^{-3}$	0.17	$1.89{\cdot}10^{-2}$	0.0005
Human	$8.98{\cdot}10^{3}$	1.00	$1.13{\cdot}10^{-4}$	$7.07{\cdot}10^{2}$	1.22	$2.40{\cdot}10^{-3}$	0.18	$4.31{\cdot}10^{-2}$	0.0560
Kangaroorat	$6.92{\cdot}10^{3}$	0.69	$1.04{\cdot}10^{-4}$	$7.24{\cdot}10^{2}$	1.03	$2.49{\cdot}10^{-3}$	0.28	$7.30{\cdot}10^{-3}$	0.0040
Manatee	$8.95{\cdot}10^{3}$	0.59	$6.77{\cdot}10^{-5}$	$5.97{\cdot}10^{2}$	0.78	$3.32{\cdot}10^{-3}$	0.19	$3.50{\cdot}10^{-2}$	0.0023
Marmoset	$8.47{\cdot}10^{3}$	0.90	$1.09{\cdot}10^{-4}$	$6.49{\cdot}10^{2}$	1.21	$2.68{\cdot}10^{-3}$	0.20	$5.65{\cdot}10^{-2}$	0.0621
Megabat	$8.50{\cdot}10^{3}$	0.75	$9.06{\cdot}10^{-5}$	$6.12{\cdot}10^{2}$	1.06	$2.66{\cdot}10^{-3}$	0.21	$8.46{\cdot}10^{-3}$	0.0134

TABLE 2.2: Mean parameters obtained by the ABC method fitting the mixture model part-1.

sampleID	mean1	n1	p1	mean2	n2	p2	a	chi_s	pos/tot
Microbat	$8.23{\cdot}10^{3}$	0.78	$9.91{\cdot}10^{-5}$	$5.64{\cdot}10^{2}$	1.07	$3.01{\cdot}10^{-3}$	0.23	$1.88{\cdot}10^{-2}$	0.0147
Mouse	$8.17{\cdot}10^{3}$	0.64	$8.09{\cdot}10^{-5}$	$6.50{\cdot}10^{2}$	1.16	$2.98{\cdot}10^{-3}$	0.23	$3.01{\cdot}10^{-2}$	0.0174
Mouselemur	$7.34{\cdot}10^{3}$	0.83	$1.15{\cdot}10^{-4}$	$5.09{\cdot}10^{2}$	0.82	$2.65{\cdot}10^{-3}$	0.27	$8.21{\cdot}10^{-2}$	0.0018
Nakedmole-rat	$8.90{\cdot}10^{3}$	0.77	$8.85{\cdot}10^{-5}$	$6.67{\cdot}10^{2}$	1.20	$2.64{\cdot}10^{-3}$	0.21	$2.27{\cdot}10^{-2}$	0.0119
Orangutan	$1.00{\cdot}10^{4}$	1.14	$1.15{\cdot}10^{-4}$	$7.65{\cdot}10^{2}$	1.37	$2.45{\cdot}10^{-3}$	0.15	$2.98{\cdot}10^{-2}$	0.0262
Panda	$9.54{\cdot}10^{3}$	0.61	$6.55{\cdot}10^{-5}$	$5.45{\cdot}10^{2}$	0.97	$2.74{\cdot}10^{-3}$	0.20	$1.84{\cdot}10^{-2}$	0.0011
Pig	$8.99{\cdot}10^{3}$	0.70	$8.00{\cdot}10^{-5}$	$7.62{\cdot}10^{2}$	1.00	$2.05{\cdot}10^{-3}$	0.17	$3.55{\cdot}10^{-2}$	0.0641
Pika	$6.77{\cdot}10^{3}$	0.74	$1.13{\cdot}10^{-4}$	$8.14{\cdot}10^{2}$	0.99	$1.99{\cdot}10^{-3}$	0.28	$4.21{\cdot}10^{-3}$	0.0035
Rabbit	$8.98{\cdot}10^{3}$	0.83	$9.45{\cdot}10^{-5}$	$6.78{\cdot}10^{2}$	1.19	$2.49{\cdot}10^{-3}$	0.19	$9.63{\cdot}10^{-3}$	0.0279
Rat	$8.36{\cdot}10^{3}$	0.60	$7.33{\cdot}10^{-5}$	$5.99{\cdot}10^{2}$	1.16	$3.18{\cdot}10^{-3}$	0.24	$3.16{\cdot}10^{-2}$	0.0060
RhesuSrheMac3	$8.00{\cdot}10^{3}$	1.07	$1.36{\cdot}10^{-4}$	$7.63{\cdot}10^{2}$	1.10	$2.26{\cdot}10^{-3}$	0.17	$8.15{\cdot}10^{-2}$	0.0563
RockhyraXproCap1	$8.96{\cdot}10^{3}$	0.50	$5.80{\cdot}10^{-5}$	$6.69{\cdot}10^{2}$	0.98	$2.20{\cdot}10^{-3}$	0.20	$1.12{\cdot}10^{-2}$	0.0075
SheePoviAri1	$8.00{\cdot}10^{3}$	0.49	$6.62{\cdot}10^{-5}$	$1.00{\cdot}10^{3}$	1.01	$1.27{\cdot}10^{-3}$	0.32	$2.90{\cdot}10^{-2}$	0.0018
Shrew	$7.19{\cdot}10^{3}$	0.66	$9.51{\cdot}10^{-5}$	$7.58{\cdot}10^{2}$	0.91	$2.09{\cdot}10^{-3}$	0.27	$6.99{\cdot}10^{-3}$	0.0054
Sloth	$9.47{\cdot}10^{3}$	0.38	$4.14{\cdot}10^{-5}$	$6.11{\cdot}10^{2}$	0.90	$2.82{\cdot}10^{-3}$	0.18	$1.30{\cdot}10^{-2}$	0.0016
Squirrel	$8.63{\cdot}10^{3}$	0.78	$9.32{\cdot}10^{-5}$	$7.77{\cdot}10^{2}$	1.14	$2.11{\cdot}10^{-3}$	0.20	$1.48{\cdot}10^{-2}$	0.0959
Squirrelmonkey	$7.53{\cdot}10^{3}$	0.94	$1.29{\cdot}10^{-4}$	$7.40{\cdot}10^{2}$	1.13	$2.41{\cdot}10^{-3}$	0.20	$1.05{\cdot}10^{-1}$	0.0580
Tarsier	$8.75{\cdot}10^{3}$	0.82	$9.56{\cdot}10^{-5}$	$7.42{\cdot}10^{2}$	1.14	$2.06{\cdot}10^{-3}$	0.20	$4.79{\cdot}10^{-2}$	0.0532
Tenrec	$7.64{\cdot}10^{3}$	0.54	$7.35{\cdot}10^{-5}$	$7.54{\cdot}10^{2}$	1.02	$1.77{\cdot}10^{-3}$	0.29	$9.84{\cdot}10^{-3}$	0.0048
Treeshrew	$8.14{\cdot}10^{3}$	0.80	$1.02{\cdot}10^{-4}$	$6.80{\cdot}10^{2}$	1.12	$2.51{\cdot}10^{-3}$	0.21	$1.10{\cdot}10^{-2}$	0.0122
Whiterhinocero	$8.68{\cdot}10^{3}$	0.87	$1.02{\cdot}10^{-4}$	$6.99{\cdot}10^{2}$	0.86	$2.34{\cdot}10^{-3}$	0.17	$2.88{\cdot}10^{-2}$	0.0304

TABLE 2.3: Mean parameters obtained by the ABC method fitting the mixture model part-2.

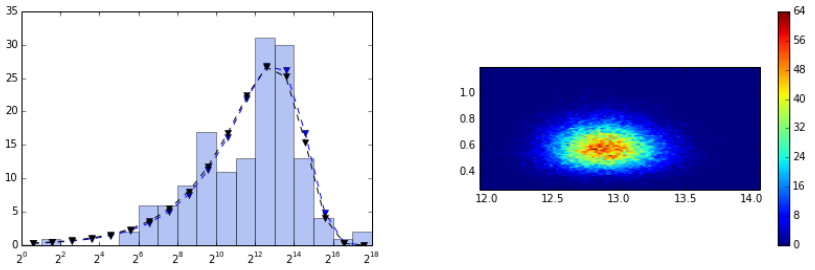

FIGURE 2.5: **ABC results of neutral model fit of Human RSA.**Comparison of the expected values from the mean parameters obtained with the ABC (black) to the Preston plot of the RSA and to the expected values obtained by non-linear least squares method (blue) on the same histogram(left panel);heat map of the posteriors of the parameters (right panel), log2 mean value (x axes), influx values (y axes).

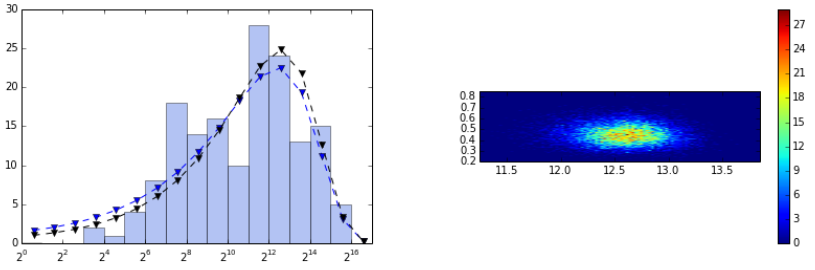

FIGURE 2.6: **ABC results of neutral model fit of Mouse RSA.**Comparison of the expected values from the mean parameters obtained with the ABC (black) to the Preston plot of the RSA and to the expected values obtained by non-linear least squares method (blue) on the same histogram(left panel);heat map of the posteriors of the parameters (right panel), log2 mean value (x axes), influx values (y axes).

Human and mouse results are shown in Fig. 2.5, 2.6, 2.7 and 2.8.

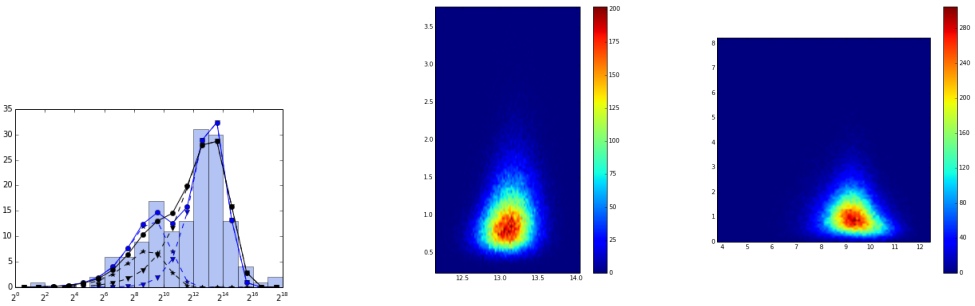

FIGURE 2.7: **ABC results of mixture model fit of Human RSA.**Comparison of the expected values from the mean parameters obtained with the ABC (black) to the Preston plot of the RSA and to the expected values obtained by non-linear least squares method (blue) on the same histogram(left panel);heat map of the posteriors of the parameters of the abund elements(center panel), heat map of the posteriors of the parameters of the rare elements(right panel), log2 mean value (x axes), influx values (y axes).

The probability of the model given the data is $P(M|D) = P(D|M)P(M)$. In ABC $P(D|M)$ can be approximated by the ration of the number of successes (number of accepted $\widetilde{\theta}$) over the number of tentatives for each model. The ABC model selection score is approximated to the ratio of fraction of successes for the two models because the a priori probability for the two models is defined equal. For the majority of the organisms under study the ABC model selection score was comparable for the two models tested.

Parameters associated to the mixture of two negative binomials allow to separate the Host Specie at level of their taxonomic Order, while the pure neutral model does not produce a good separation, as can be seen in Fig. 2.9. We show here only the most populated Orders ($n > 2$). ABC method already penalizes the model with higher number of parameters because phase-space is larger. Thus the mixture model is preferred respect the neutral one to describe data. In particular the couples of parameters Υ_1, Υ_2 and x_1, x_2 seems a good representation to discriminate the host organisms in different taxonomic Orders. Within our description such couples of parameters are related by the value of the disadvantage due to competition. The values of the Υ_1, Υ_2 parameters, all of order one, indicates that a pure accumulation process seems a more suitable model. Despite transcomplementation may take place, up to this description it is not a very relevant process in shaping the RSA of the community.

Transposons activity deeply contributes to shaping genomes, thus LINEs abundance could be a good indicator for phylogeny in Mammals where this family is particularly abundant. This can be noticed clustering in a heat map TEs abundances in different organisms (results not shown). For this reason it is not straightforward to determine if the better performance of a mixture model is due to the true existence of a competition or if a mixture of distribution better describe fluctuations in the RSA that have been inherited by common ancestors. If competition took place, when it happened and which Elements were involved? To answer this question data available from [Gio+07] have been used to order LINE Elements by their age of activity in human, Chimpanzee, Rhesus Macaque, mouse and rat genomes.

The list has been subdivided into time intervals (windows) containing a fixed

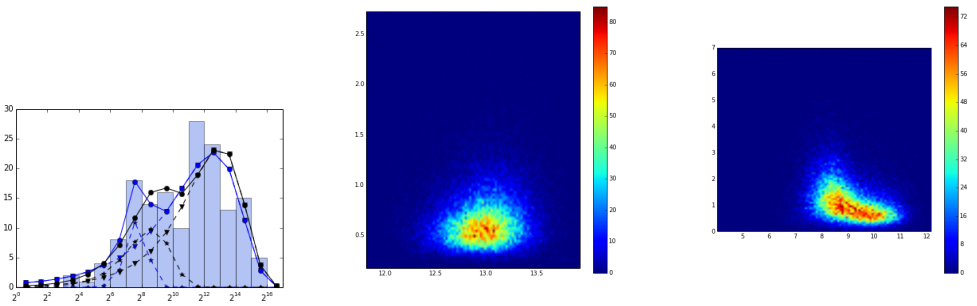

FIGURE 2.8: **ABC results of mixture model fit of Mouse RSA.**Comparison of the expected values from the mean parameters obtained with the ABC (black) to the Preston plot of the RSA and to the expected values obtained by non-linear least squares method (blue) on the same histogram(left panel);heat map of the posteriors of the parameters of the abund elements(center panel), heat map of the posteriors of the parameters of the rare elements(right panel), log2 mean value (x axes), influx values (y axes).

number of Elements. The number of Elements in the sliding window can be arbitrarily chosen, here we show results for $N = 15$, which is a compromise between a zoom in the action period and number of elements included in the ecosystem, for analysis purposes. Each window represents a picture of the RSA in a different evolution stage. This time dependent ecosystem has been tested respect to the neutral model and the mixture model with the same ABC method and same prior distributions. It results that in the primates under study neutrality is violated between the 40-65 interval of the rank, where the mixing coefficient of the mixture model is higher, and ABC model selection score suggest that a more complex description is suitable (Fig. 2.10). Similar results can be find for mouse and rat in Fig. 2.11, where a preference for the mixture model is maintained in all ancient LINEs. The trend for the birth-death ratio and the influx parameters are reported in Fig. 2.12,2.13.

A transition in time to different expectation values can be identified both for primates that for mouse and rat in Fig. 2.14, where the parameters of the negative binomial distributions describing the sliding windows RSA are shown. The reported couples of parameters are correlated by construction by the mean of the distribution, which represents the expectation value of LINE copy number in that time. In primates a transition to a lower average copy number is observed while in rat and mouse to a larger copy number. This transition can be observed in the pure neutral model as well as in the mixture model in the component describing the Elements with high copy number. The group associated to rare Elements results much more noisy, with parameters much less correlated. The transition is more evident in the mixture description, for this reason again a mixture model seems more suitable to describe the system.

To check if LINEs abundances follow phylogeny we clustered hierarchically the abundance of the Elements that should be involved in competition in human over the datasets considered (Fig. 2.15). The separation between rare and abundant Elements is maintained across all mammals included in this study, with the exception of the White Rhinoceros, which shows the opposite trend. Furthermore species results mostly clustered by their taxonomic orders.

To test the hypothesis that competition is caused by similarity in the promoter

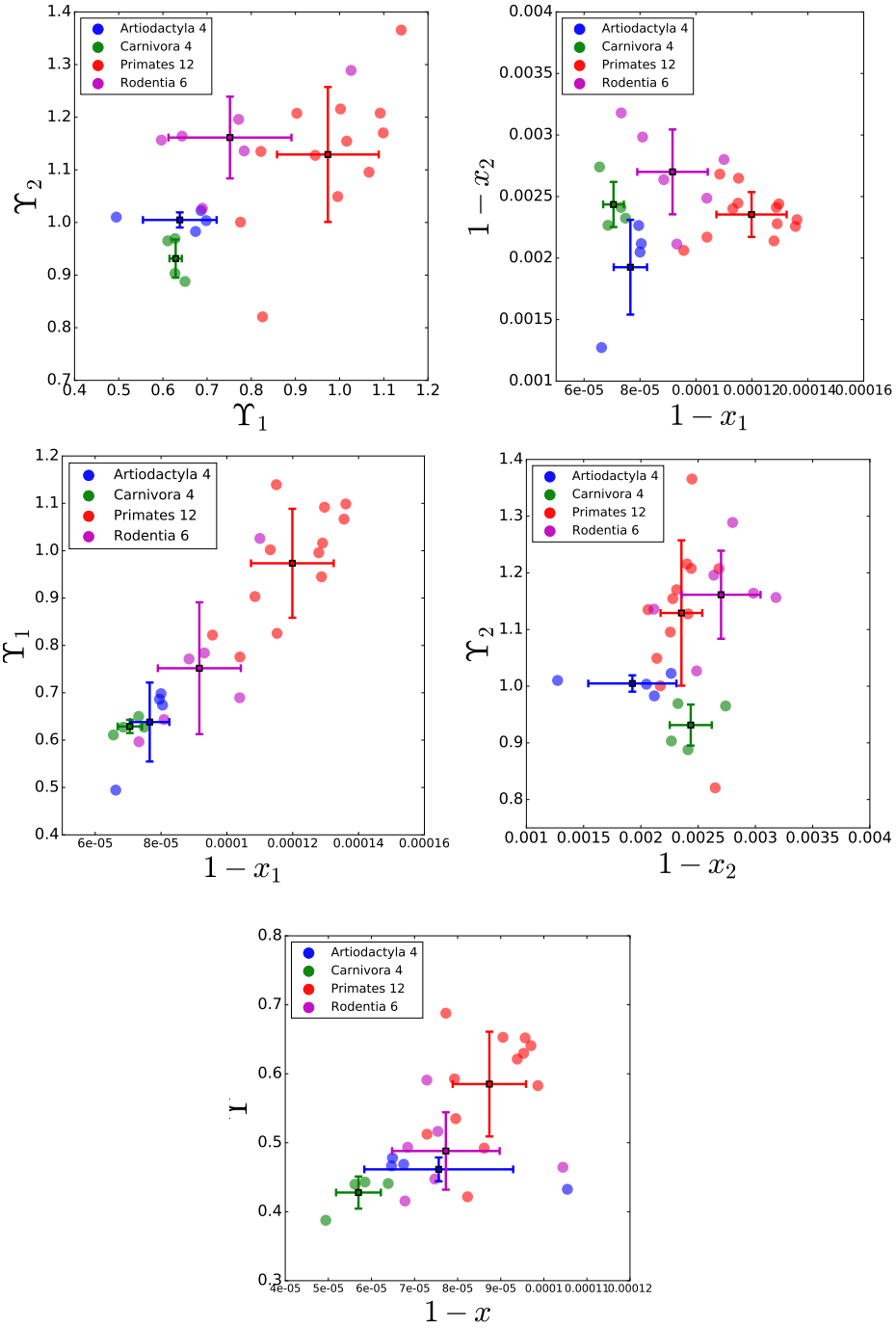

FIGURE 2.9: **Fit of the RSA of LINEs with a mixture model clusters different Mammalian Orders.** The set of optimized parameters obtained by fitting a mixture model of two negative binomials on 42 mammalian LINE RSAs are able to separate the most represented Taxonomic Orders: Υ_1 respect to Υ_2 and $1-x_1$ respect to $1-x_2$ respectively (upper panels). The couple of parameters for each of the two negative binomials (middle panels) allows as well a separation of different Orders, in particular in the case of the parameters describing the abundant group Υ_1, $1-x_1$. The parameters describing the pure neutral model (lower panel) do not clearly discriminate different Orders, however a trend similar to the couple of parameters of the abundant group of the mixture is observed.

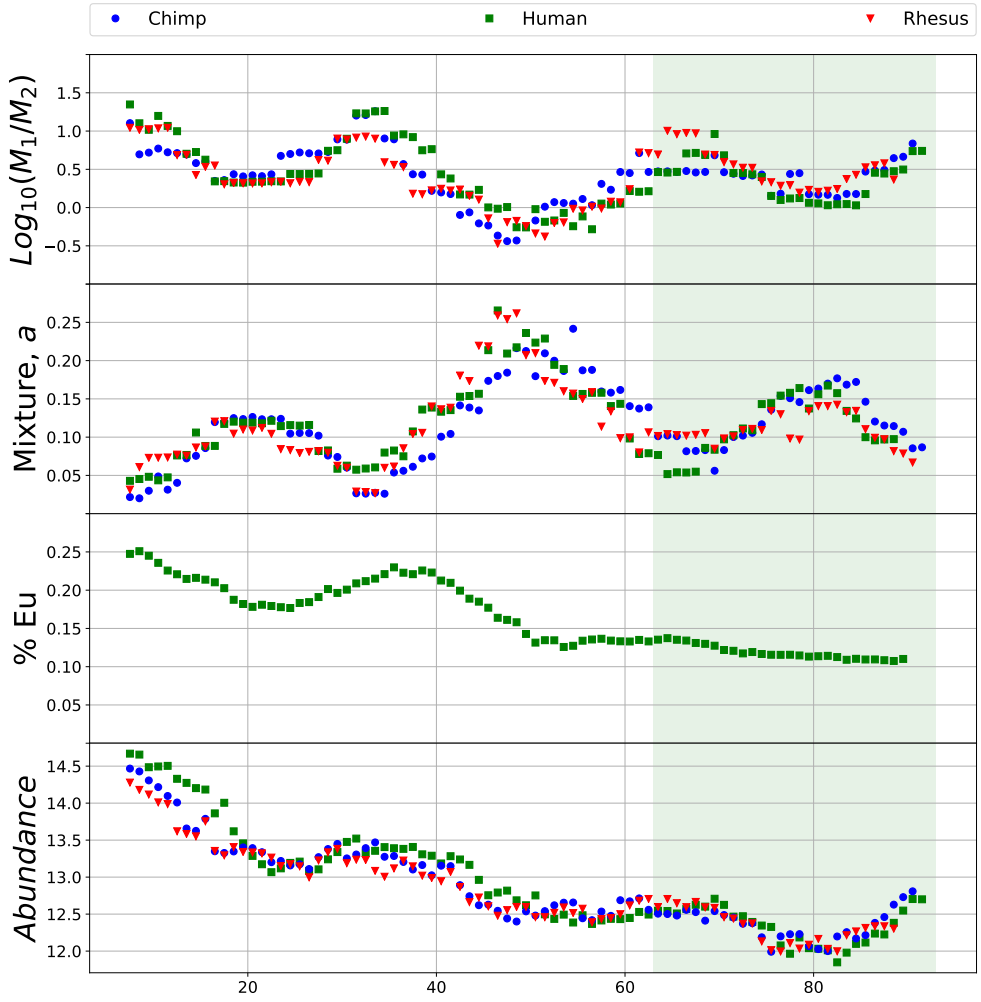

FIGURE 2.10: **Statistical comparison of a mixture model of two negative binomials with respect to the neutral model of [Vol+03] during the evolution of the genome ecosystem in three Primates.** Data available from [Gio+07] have been used to rank order LINE Elements by their age of activity in Homo Sapiens, Chimpanzee and Rhesus Macaque genomes. The rank has been subdivided into intervals containing a fixed number of contiguous Elements ($N = 15$), each interval has been used as a sample ecosystem to fit both the neutral model and the mixture model. (a) The ABC model selection approach was used to compare the goodness of fit between the two models. The ABC model selection approach shows that a mixture model provides a better fit between rank positions 40 and 65 of the time ordered age of LINEs ("non-neutral time interval"). (b) Estimation of the mixture coefficient a during evolution. This coefficient represents the proportion of species associated to the regime in the mixture model, described by the negative binomial with a lower mean (i.e. LINE subfamilies with fewer Elements in the genome). When a is higher, it indicates the presence of a non negligible group of rare LINE species. (c) The percentage of LINE copies inserted in euchromatic regions displays a decreasing trend with time ordered age in human. (d) The copy number of LINE (euchromatin and heterochromatin insertions) displays a decreasing trend with time order as well.

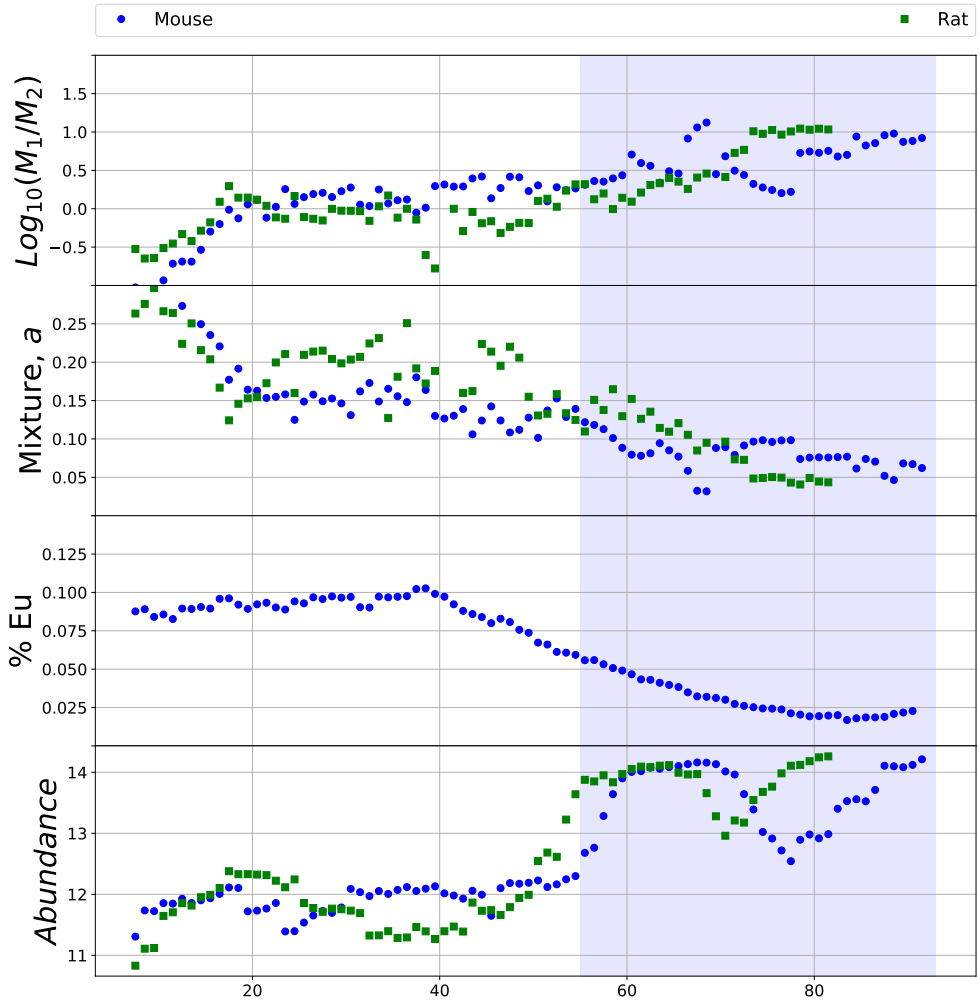

FIGURE 2.11: **Statistical comparison of a mixture model of two negative binomials with respect to the neutral model of [Vol+03] during the evolution of the genome ecosystem in mouse and rat.** Data available from [Gio+07] have been used to rank order LINE Elements by their age of activity in Mouse and Rat genomes. The rank has been subdivided into intervals containing a fixed number of contiguous Elements ($N = 15$), each interval has been used as a sample ecosystem to fit both the neutral model and the mixture model. (a) The ABC model selection approach was used to compare the goodness of fit between the two models. The ABC model selection approach shows that a mixture model provides a better fit in the ancient Elements of the time ordered age of LINEs ("non-neutral time interval"). (b) Estimation of the mixture coefficient a during evolution. This coefficient represents the proportion of species associated to the regime in the mixture model, described by the negative binomial with a lower mean (i.e. LINE subfamilies with fewer Elements in the genome). When a is higher, it indicates the presence of a non negligible group of rare LINE species. (c) The percentage of LINE copies inserted in euchromatic regions displays a decreasing trend with time ordered age in human. (d) The copy number of LINE (euchromatin and heterochromatin insertions) displays a decreasing trend with time order as well.

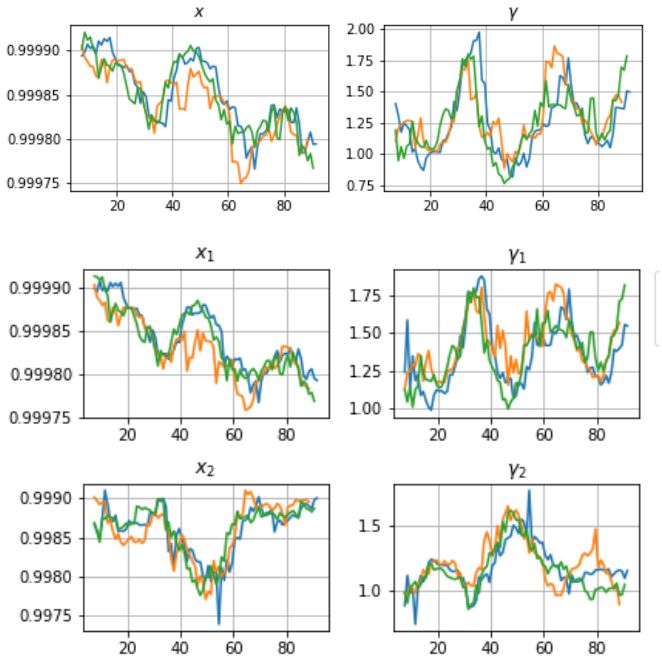

FIGURE 2.12: **Evolution of estimated parameters for the neutral [Vol+03] and the mixture model [TZ13] in three Primates (Homo S., Chimpanzee, Rhesus M.) using a ABC method method (https://pymc-devs.github.io/pymc/).** (a) Birth-death rate parameter evolution, $x = b/d$, obtained for the neutral model; (b) Constant influx parameter evolution, $\Upsilon = S/b$, obtained for the neutral model; (c) Birth-death rate parameter evolution, $x = b/d$, obtained for the rare species regime in the mixture model; (d) Constant influx parameter evolution, $\Upsilon = S/b$, obtained for the rare species regime in the mixture model; (e) Birth-death rate parameter evolution, $x = b/d$, obtained for the common species regime in the mixture model; (f) Constant influx parameter evolution, $\Upsilon = S/b$, obtained for the common species regime in the mixture model.

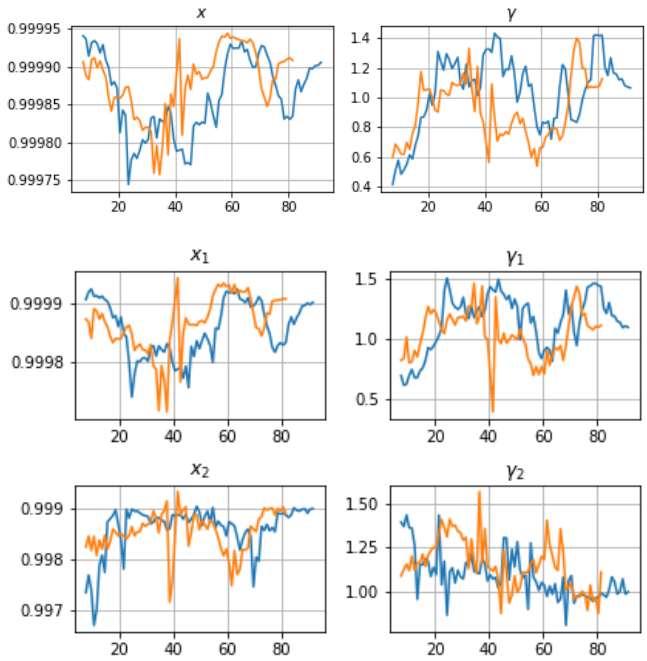

FIGURE 2.13: **Evolution of estimated parameters for the neutral [Vol+03] and the mixture model [TZ13] in mouse and rat using a ABC method.** The analogous analysis reported in Fig. 2.12 for Primates has been performed for two Rodents (mouse, rat), for which data were available. (a) Birth-death rate parameter evolution, $x = b/d$, obtained for the neutral model; (b) Constant influx parameter evolution, $\Upsilon = S/b$, obtained for the neutral model; (c) Birth-death rate parameter evolution, $x = b/d$, obtained for the rare species regime in the mixture model; (d) Constant influx parameter evolution, $\Upsilon = S/b$, obtained for the rare species regime in the mixture model; (e) Birth-death rate parameter evolution, $x = b/d$, obtained for the common species regime in the mixture model; (f) Constant influx parameter evolution, $\Upsilon = S/b$, obtained for the common species regime in the mixture model; (g) Estimation of the mixture coefficient a during evolution. This coefficient represents the proportion of species associated to the second regime in the mixture model, described by the second negative binomial (rare species); (h) Comparison of Bayesian Information Criteria (BIC) of the neutral and mixture models.

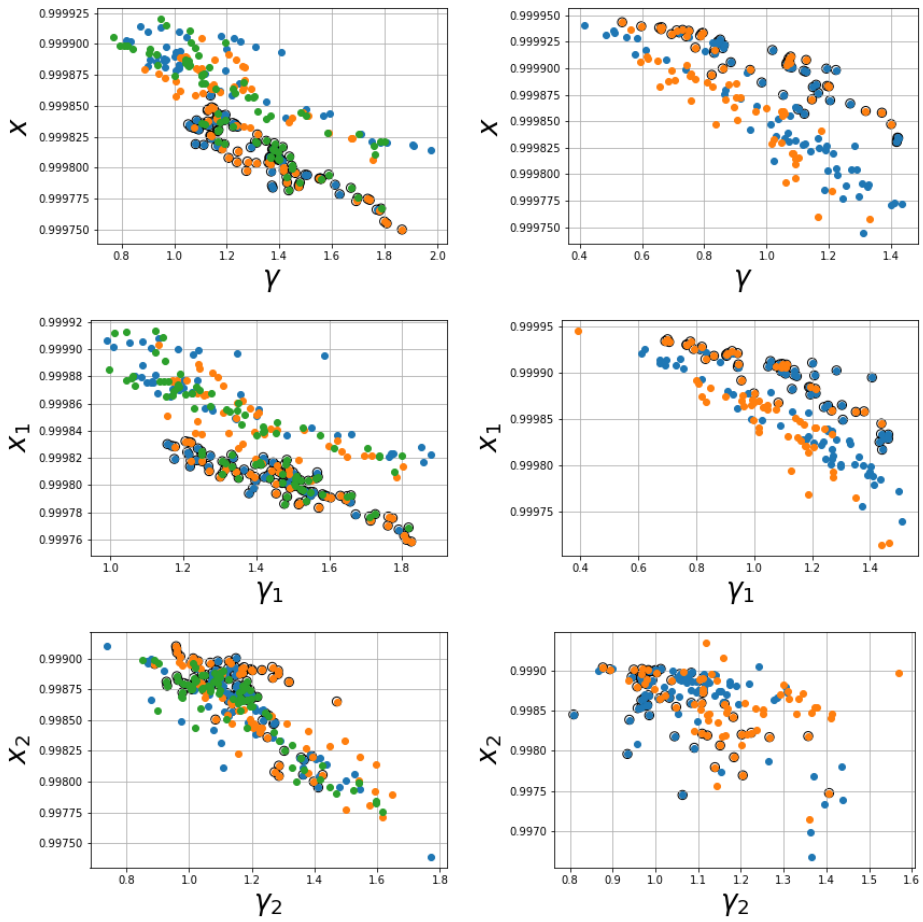

FIGURE 2.14: **Space of parameters of LINEs evolution in a primates and murinides shows evidences of radiation events.** The space of parameters describing sliding window ecosystem of LINEs in human, chimpanzee, rhesus macaque (left panels) and mouse, rat (right panels) is shown. x and Υ parameters are correlated by the expected value (mean) of the distribution. Upper panels refer to the neutral model, middle panels refer to the group of the mixture model with highest copy number, lower panels refer to the group of rare Elements of the mixture model. Black circles indicated the most recent Elements, associated to Primate and Murinide radiation. It can be noted that two groups can be distinguished and it leads to a lower average copy number for primates and to higher average copy number for murinides. This is reflected in the distribution of copies in chromatine states, but in the case of human this transition is not significant as for mouse. As for the discrimination of host species in different taxonomic Orders, a mixture model seems more efficient to isolate the different dynamics. The group of rare Elements behave as a noise for the ecosystem and then does not produce significant informations to distinguish the two dynamics.

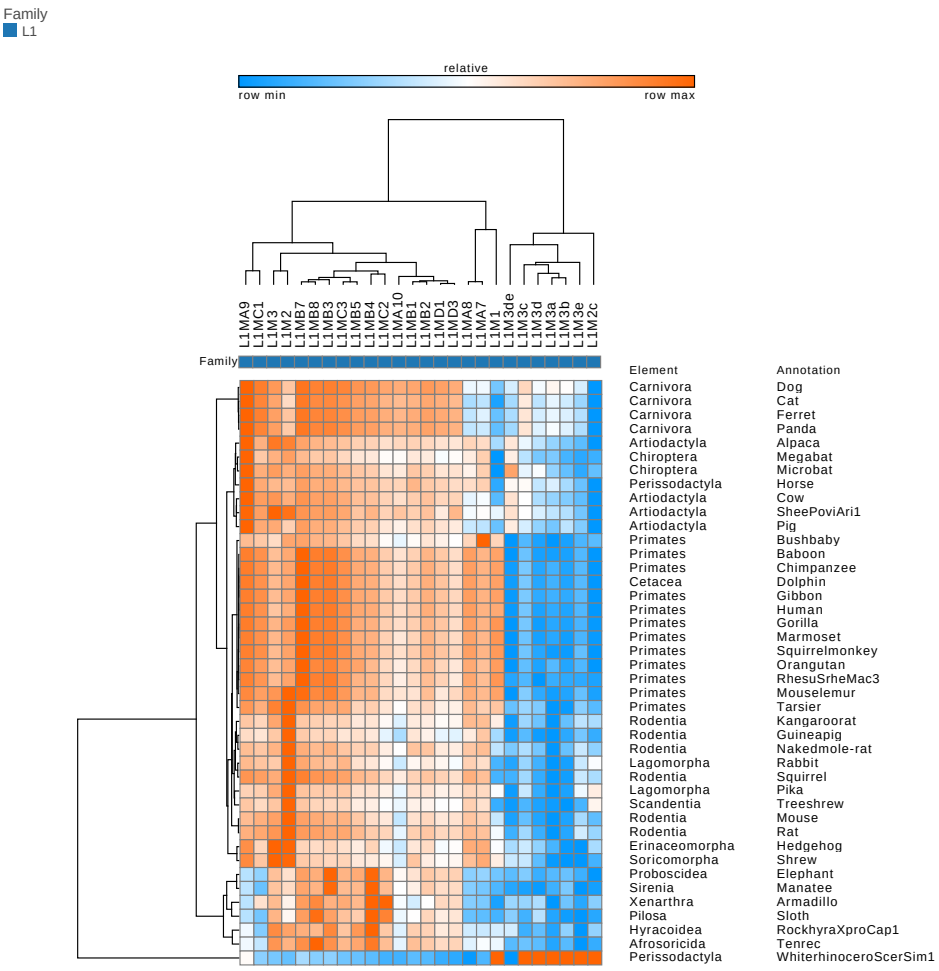

FIGURE 2.15: **Cluster species abundances in 46 mammalian reference genomes for LINEs in the non-neutral time interval.** The separation between abundant species (red) and rare species (blue) is maintained across all mammals included in this study, with the notable exception of the White Rhinoceros, which shows an opposite trend.

region, We aligned pairwise the available consensus sequences of the 5'UTR in RepBase using ClustalW2 and we calculated the distance between each couple. 5'UTR regions have been selected for a subset of Elements present in the human genome with a well characterized 5'UTR sequence, and compared to all the other sequences available. Significant similarity between 5'UTRs is observed for the following high and low copy numbers pairs: L1M2-L1M2c, L1MA9-L1M3b and L1M2-L1M3DE (Fig. 2.16).

Transposons activity should deeply affect the whole structure of the genome. This may be partially quantified by chromatine state occupation of LINEs copies. Using chromatin state assignments in human [EK10] and mouse [Yue+14] genomes, and the coordinates of the respective TEs insertions from Repbase , we assigned to each LINEs copy a chromatin state, distinguishing between insertions in open and closed chromatin states, currently known as euchromatin and heterochromatin respectively. New insertions fixate in the germ line, then We refer to the state assignment in embryonic stem cell. Multiply assignment of TEs insertions to chromatin states has been treated classifying the combination of states into open, weakly open and closed chromatin, depending if the states identified belong mainly to one of these groups. Weakly open chromatin population has been added one time to open chromatin and after to closed chromatin and we found that this choice did not change significantly our results. Unknown state has been included into closed chromatin group.

The average percentage of LINE copies inserted in euchromatic regions in the sliding window displays a decreasing trend with time ordered age in human (Fig. 2.10) and mouse (Fig. 2.11). However, in humans it also shows a clear peak within the neutral time interval. The average percentage of copies in euchromatic regions is bigger for the windows with higher average copy number respect to the one with low copy number in human and in ancient Elements in mouse. The presence of a higher fraction of rare species within the non-neutral time interval results then in agreement with the lower average percentage of insertions in euchromatin observed.

This relation is clarified considering the correlation between the number of insertions in euchromatin and the number of insertions in heterochromatin (Fig. 2.17). The linear correlation between the logarithm of the counts corresponds to a correlation of power law type between the raw counts (any age assignment is considered):

$$log2(N_{Eu}) = c \cdot log2(N_{Het}) + c_0 \pm \epsilon \,, \tag{2.29}$$

$$N_{Eu} = 2^{c_0 \pm \epsilon} N_{Het}^{c} \,, \tag{2.30}$$

in human we have $c = 1.18$, $c_0 = -4.58$ and $\epsilon = 0.035$, which correspond to the standard error in the estimate. The correlation coefficient is $r = 0.96$ and the p-value $p \sim 10^{-55}$. The superlinear correlation between the two quantity leads to the interesting result that an higher abundance, i.e. sum of euchromatin and heterochromatin contributions, is related to an higher percentage of insertion in euchromatin states. The average trend of Elements abundance in time result in fact slightly decreasing as well. We suggest that the decreasing trend can be caused by the host selection pressure which on average select less invasive transposons.

In Fig. 2.11 we observed in mouse genome a plateau in the percentage of euchromatin insertions followed by decreasing. However the average abundance drastically increases at a certain point. This correspond to a transition to a different value of the coefficient c_0 in the correlation plot for mouse as can be seen in Fig. 2.17.

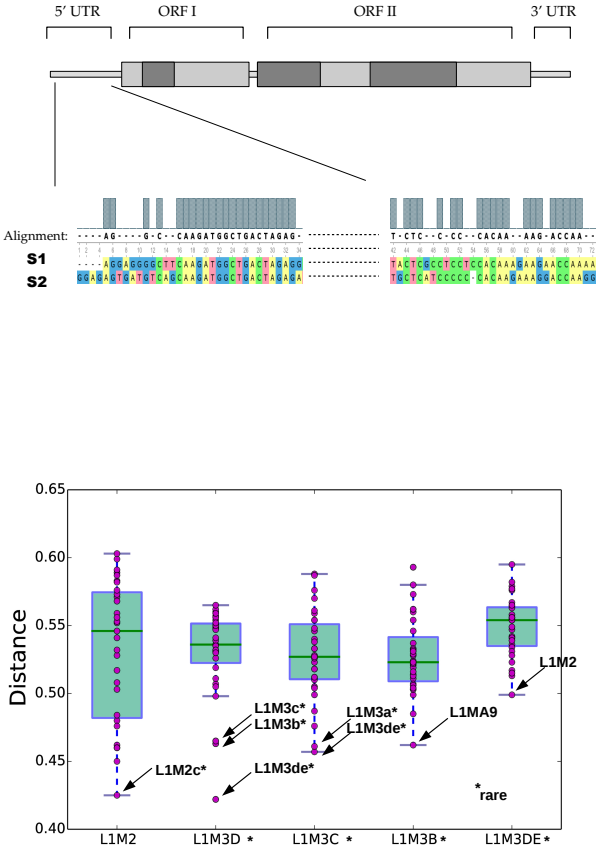

FIGURE 2.16: **5'UTR similarity between competing LINE retrotransposons in human.** In [KSB06] is suggested that different L1 Elements may coexist without competing if the 5'UTR is different, while a specie will overcome the others if the 5'UTR are very similar and might compete for the same factors. (a) The available consensus sequences of the 5'UTR of LINEs in the human genome have been aligned pairwise, with ClustalW2. (b) In several cases, the minimum distance is achieved between couple of Elements with similar ages and having high and low copy number respectively. 5'UTR regions have been selected for a subset of Elements with a well characterized 5'UTR sequence, and compared to all the other sequences available; the distance statistics and the label of the most similar sequences are reported in the plot. Significant similarity between 5'UTRs is observed for the following high and low copy numbers pairs: L1M2-L1M2c, L1MA9-L1M3b and L1M2-L1M3DE.

For ancient LINE species the correlation between the number of insertions in euchromatin and the number of insertions in heterochromatin is compatible with the one observed in human, for more recent Elements instead we observe a quite well separated cluster, which can be evidenced by PCA, associated to lower c_0 value. Given the same number of insertion in euchromatin states, a lower value for c_0 corresponds to a larger abundance and consequently to a lower percentage of insertions in euchromatin. The most present LINE family at the beginning of this transition is Lx, the amplification of Lx in the genomes is coincident with murine radiation according to [PVF] and [Fur+94], in fact the other Elements characterizing this group are mainly mouse and murine specific. This evidence supports the idea of evolutive adaptation of both the host and Elements.

It is interesting how this considerations are supported by observation of the time dependent RSA distribution. In mouse we expect to observe some transition in the RSA distribution, which indeed is observed in Fig. 2.14. Instead we recall that in primates a transition is observed but to lower average copy number. This could be the reason why chromatine states distribution in human is not affected significantly, since further silencing mechanisms where not necessary to preserve the host fitness.

Wewant to stress at this point that mechanism of competition proposed is independent of the chromatine state distribution of the Element copies, but act at the level of the Elements affecting their abundances. Instead chromatin state of the insertions should reflect the interaction of the Element with the host, by mechanisms of silencing and self regulation, affecting the whole dynamics of the Element. At least in the case of Murine family a change in transposon dynamics can be associated to evolutive radiation phenomena. Finally, the concept of neutrality we propose, that is that all Elements have the same birth death rates, can be relaxed rescaling the time by a constant the rates can properly adjusted if Elements are active mainly one by one. In conclusion we may consider that equilibrium between host and Elements does not hold and they possibly contribute as one of the driving forces to evolution processes. Furthermore competition with a stochastic origin introduces deviations from the neutrality proposed between the Elements and support the introduction of variation in Elements sequences. Our model approximation of the RSA allows to identify evolutive transition both in primates that in mouse and rat. Then it results a very interesting approach with promising applications in future studies.

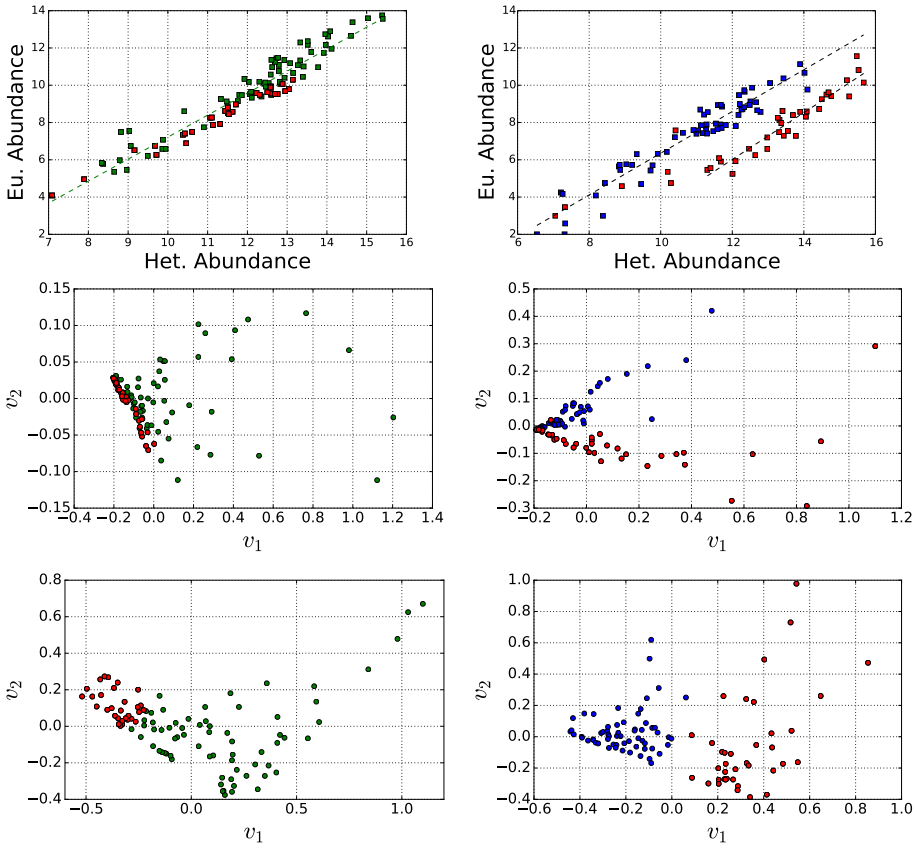

FIGURE 2.17: **Non linear correlation between number of insertion in euchromatin and heterochromatin states shows host genome adaptation mechanism.** Scatter plot in $log2$ scale of the number of insertions in euchromatin respect that in heterochromatin for each LINE specie in human (left panels) and mouse (right panels). The number of insertions in euchromatin and heterochromatin states result correlated by a power law with exponent ~ 1.2 (upper panels). The number of insertions in euchromatin and heterochromatin states result correlated by a power law with exponent. The group of most recent Elements (in red) is well separated from the others both in Human and Mouse by PCA (middle panels). This is much more evident when age variable is included in PCA (lower panels).

Chapter 3

Langevin approach to generate anomalous transport in complex environment

In this Chapter we will show that anomalous diffusion can be described in terms of a superposition of classical diffusion processes within a Langevin approach. In this representation the complexity of the system does not arise from the complexity of the process itself, but from the heterogeneity of the parameters characterizing the process. A suitable system where to apply this approach is a heterogeneous ensemble of Brownian particles that differ in their mass and radius. The stochastic dynamical equation of the center of mass of this ensemble is derived accordingly to three statistically equivalent approaches: the superposition of Langevin equations, the generalized grey Brownian motion, and a Langevin-type equation. The case of a test-particle immersed in a heterogeneous surround is studied for modeling anomalous diffusion in biological systems. We analytically and numerically demonstrated that with proper populations of masses and of frictions, fractional diffusion emerges.

3.1 Introduction

The very rich dynamics of biosystem movements have been attracting the interest of many researchers in the field of statistical physics and complexity for its inherent temporal and spatial multi-scale character. Further, new techniques allowed to track the motion of large biomolecule in the cell with great temporal and spatial accuracy, both *in vivo* and *in vitro* [HF13; Reg+13; MGP15]. Two main transport mechanisms were identified: (i) passive motion, determined by the cytoplasm crowding and (ii) active transport, given by the presence of molecular motors carrying biomolecules along filaments and microtubules (cytoskeleton) [TN+04; GC06; Jav+14; CGE00]. Diffusion processes have been used to describe many biological phenomena such as molecular motion through cellular membrane [Wei+11; Jav+13; Kra+16; MJC16], DNA motility within cellular nucleus [Jav+14], chromosome dynamics and motility on fractal DNA globules [Tam+15], motion of mRNA molecules in *Escherichia coli* bacteria [GC06] and of lipid granules in yeast cells [TN+04].

Standard or normal diffusive (Brownian) motion is uniquely described by the Wiener process [Ris89] and is associated with a Gaussian Probability Density Function (PDF) of displacements and linear time dependence of the Mean Square Displacement (MSD). However, biosystems' diffusion is often non-standard, with non-Gaussian PDF of displacements non-linear time dependence of MSD.

Anomalous diffusion behavior can be associated with the polydispersity of the system, when classical thermodynamics holds and space and time correlations "do not play a major role", or by long-range spatiotemporal correlations with even "anomalous" thermodynamics [GC04; BG90; MK00]. Several models and interpretations were proposed in the recent literature [Bur+11; HF13; Met+14; MGP15]. Widely investigated models of anomalous diffusion are Continuous Time Random Walk (CTRW) and Fractional Brownian Motion (FBM). Many authors compared these models with each other and with data, essentially finding some features to be satisfied by the CTRW (weak ergodicity breaking and aging) [He+08; Bur+11; Jeo+11] and other ones by the FBM (e.g., the p-variation index [Mag+09; KBG11; Bur+12]). Despite the efforts of many research groups, an exhaustive model explaining all the statistical features of experimental data does not yet exist and the research is recently focusing on alternative approaches, such as Heterogeneous Diffusivity Processes (HDPs) [CCM13; Mas+14; CS14; CM16].

In this framework, we here propose a modeling approach to anomalous diffusion based on the concept of grey Brownian motion introduced by Schneider [SW89; SAL92] [1] , and later by the generalized grey Brownian Motion (ggBM) [Mur11; MTM08; MP08b; MM09; PMM12; PMM13].

This is essentially equivalent to have a fluctuating diffusivity, thus giving a possible stochastic interpretation of HDPs. When the amplitude PDF is the Mainardi distribution [MMP10; Pag13; Pag14], the gBM-PDF $P(x,t)$ solves the Time Fractional Diffusion Equation (TFDE) [Gor+02a; Gor+02b; MLP01]. The ggBM generalizes grey noise by considering the FBM $B_H(t)$ as fundamental solution for constant amplitude [BMN68] and correspond to the stochastic solution of the Erdelyi-Kober Fractional Diffusion Equation (EKFDE) [Pag12]. A further extension of the ggBM is given by the process introduced by Pagnini and Paradisi in [PP16], where the amplitude distribution is generalized to a combination of Lévy distributions by imposing the ggBM-PDF to be compatible with the Space-Time Fractional Diffusion Equation (STFDE) [Gor+02a; Gor+02b; MLP01].

A crucial aspect of ggBM is that single trajectories are driven by a Gaussian process with stationary increments, then it is not necessarily related to standard or fractional Brownian motion, and that it is also suitable to describe nonstationary and aging behaviors. The potential applications of ggBM to biological transport were recently discussed in [MG+16], where the ggBM was investigated by means of several statistical indices commonly used in the analysis of particle tracking data, showing

[1] The Schneider grey noise is defined by the measure μ_α that satisfy in the space of tempered functions $S'(\mathbb{R})$ on $\mathbb{R}$ [SAL92]:

$$\int_{S'(\mathbb{R})} e^{i\langle\omega,\xi\rangle} d\mu_\alpha(\omega) = E_\alpha(-\|\xi\|_\alpha^2) \tag{3.1}$$

where $E_\alpha(\cdot) = \sum_{k=0}^{\infty} \frac{(\cdot)^k}{\Gamma(\alpha n+1)}$, with $0 < \alpha < 1$, is the Mittag Leffer function, representing the characteristic function of the variable defined by the measure. The Mittag Leffer function is the natural generalization of the exponential function, as well as the M-Wright distribution can be considered a generalization of the Gaussian distribution. In the limit $\alpha = 1$ the grey noise reduces to the white noise, losing the long time correlation, as the Mittag Leffer reduces to the exponential function. The generalized grey noise is defined by the expression [MP08b]:

$$\int_{S'(\mathbb{R})} e^{i\langle\omega,\xi\rangle} d\mu_{\alpha,\beta}(\omega) = E_\alpha(-\|\xi\|_\beta^2) \tag{3.2}$$

which describes grey noise for the particular combination of parameters $\alpha = \beta$. The diffusion process associated is the generalized grey Brownian motion introduced in Chapter 1, that corresponds to grey Brownian motion in the case $\alpha = \beta$.

that ggBM approach qualitatively accounts for the weak ergodicity breaking and aging (CTRW) and, at the same time, for the p-variation test (FBM) characterizing the dataset. However, the physical interpretation of ggBM approach based on the Gaussian noise is not completely clear. Further, potential applications to transport in a viscous fluid needs to include at least the effect of viscosity.

To this goal, in this Chapter, we describe the development of a model similar to the original ggBM, the complexity of the medium is described by proper random fluctuations of the parameters in the Langevin equation (friction or relaxation time, diffusivity or noise intensity), thus allowing to get anomalous diffusion from a superposition of simpler process, i.e. the Ornstein-Uhlenback (OU) process.

In the considered system, anomalous diffusion is caused by the heterogeneity of the mesoscopic surround, which is responsible for long-range correlations, and it is displayed during an intermediate asymptotic transient regime in the Barenblatt's sense [Bar79] requiring an underdamped (white noise) Langevin approach within the classical thermodynamics framework.

In Section 3.2 the basis and some useful details of the classical Langevin approach are recalled.

In Section 3.3 the motion of the center of mass of a system composed of non-identical Brownian particles that differ in their density (mass-to-volume ratio), let's call it *heterogeneous ensemble of Brownian particles*, is studied. The study of the center of mass allows for estimating the average concentration and the momentum of inertia of the ensemble by computing the mean and the mean square displacement, respectively.

In Section 3.4 is presented the diffusion of a test-particle coupled to the mesoscopic surround defined by this heterogeneous ensemble, which can be applied to understand and model anomalous diffusion phenomena.

In Section 3.5 we present the details of the randomized Langevin model for superdiffusion simply based on the free motion of Brownian particles in a complex viscous medium, and subdiffusion based on the insertion of a harmonic potential in the system.

In Section 3.6 we show a few examples of the numerical simulations that have been performed to check the validity of the approach.

3.2 The Langevin equation

The Langevin equation has been introduced by Langevin a short time after Einstein's work to describe the Brownian dynamics. The particle motion is defined in terms of stochastic equations on the basis of thermal energy considerations with the introduction of phenomenological stochastic forces. This description turns out to be "infinitely more simple" respect to other approaches, as considered by Einstein himself. Let's recall the bases of the Langevin approach, for which we refer to the classic textbooks [Kam81], [Gar90], [Ris89], that will result useful in the derivation of the model details later.

3.2.1 The standard Ornstein-Uhlenbeck process

A free Brownian particle in a viscous medium can be described by the classical Langevin equation, where the velocity evolution is driven by an Ornstein-Uhlenbeck

(OU) process.

$$m\frac{dV}{dt}(t) = -\gamma V(t) + \Gamma\xi(t)\ , \tag{3.3}$$

where $\xi(t)$ is a Gaussian white noise, Γ the noise intensity, m the particle inertial mass and $\gamma = 6\pi\nu r$ the friction coefficient given by the Stokes law, determined by fluid viscosity ν and particle radius r. Here we limit to the one-dimensional OU process, but the equations are easily generalized to the three-dimensional case. The white noise $\xi(t)$ is defined by the following conditions:

$$\langle\xi(t)\rangle = 0\ ; \quad \langle\xi(t)\xi(t')\rangle = \delta(t-t')\ . \tag{3.4}$$

being $\langle\cdot\rangle$ the notation for the Gibbs ensemble average. The Gaussianity of the process is defined by introducing the stochastic integral of the white noise $\xi(t)$ (see, e.g., [Kam81]):

$$W(t,\tau) = \int_t^{t+\tau} \xi(t')dt'\ . \tag{3.5}$$

$W(t,\tau)$ is defined to be a Wiener stochastic process when the increments $dW(t,dt) = W(t+dt) - W(t) = dW(dt)$ with mean and variance given by:

$$\langle dW(dt)\rangle = 0\ ; \quad \langle dW^2(dt)\rangle = dt\ . \tag{3.6}$$

As usual, the above equation for the acceleration is completed by the kinematic relation between velocity $V(t)$ and position $X(t)$. The particle system is then fully described by the random vector $(X(t), V(t))$:

$$\begin{cases} \frac{dX}{dt}(t) = V(t) \\ \\ \frac{dV}{dt}(t) = -\frac{V(t)}{\tau_c} + \sqrt{2D_{_V}}\ \xi(t) \end{cases} \tag{3.7}$$

The particle dynamics is characterized by the velocity diffusivity coefficient $D_{_V}$ and the relaxation time scale τ_c:

$$D_{_V} = \frac{\Gamma^2}{2m^2}\ ; \quad \tau_c = \frac{m}{\gamma}\ . \tag{3.8}$$

The formal solution for the OU process $V(t)$ in the Langevin equation (3.7) is given by:

$$V(t) = e^{-(t-t_0)/\tau_c}\left[v_0 + \sqrt{2D_{_V}}\int_{t_0}^{t} dt' e^{(t'-t_0)/\tau_c}\xi(t')\right]\ , \tag{3.9}$$

where $v_0 = V(t_0)$. The corresponding velocity autocorrelation function (VACF) reads:

$$C_v(t_1,t_2) = \langle V(t_1)V(t_2)\rangle = \left(\langle v_0^2\rangle - D_{_V}\tau_c\right)e^{-(t_1+t_2-2t_0)/\tau_c} + D_{_V}\tau_c e^{-|t_1-t_2|/\tau_c}\ . \tag{3.10}$$

The equilibrium state is reached in the long-time regime: $t_1, t_2 \gg \tau_c$. In this case, the exponential term with the sum of times t_1+t_2 becomes negligible, while the term with the time lag $t_1 - t_2$ remains different from zero for small time lags even when t_1 and t_2 become large. The equilibrium VACF then reduces to:

$$R_{\rm ou}(t = |t_1 - t_2|) = D_{_V}\tau_c e^{-t/\tau_c}\ . \tag{3.11}$$

From the above equations, and especially from Eq. (3.11), it is clear that, for the OU process, the relaxation time scale $\tau_c = m/\gamma$, related to both particle inertial mass and friction, also characterizes the exponential decay of the VACF.

3.2.2 The classical Langevin oscillator

The classical Langevin oscillator describes the motion of a Brownian particle which diffuses under the influence of an external harmonic force, $F(x) = -kx$. Where x represents the displacement of the particle respect to its equilibrium position and the constant k is the strength of the harmonic force. The associated Langevin equation reads

$$m\frac{dV}{dt}(t) = -\gamma V(t) - m\omega_0^2 X(t) + \Gamma\xi(t)\,, \tag{3.12}$$

where $V(t)$, $X(t)$ and m represent respectively the velocity, the displacement and the mass of the Brownian particle, $\gamma = m/\tau_c$ is the friction of the medium, $\omega_0 = \sqrt{k/m}$ is the characteristic frequency of the oscillation and Γ is the amplitude of the white Gaussian noise $\xi(t)$.

We rewrite the equation in term of the relaxation time τ_c, the coefficient of diffusion of velocity $D_V = \Gamma^2/(2m)$

$$\begin{aligned} \frac{dX}{dt}(t) &= V(t) \\ \frac{dV}{dt}(t) &= -\frac{1}{\tau_c}V(t) - \omega_0^2 X(t) + \sqrt{2D_V}\xi(t)\,, \end{aligned} \tag{3.13}$$

The formal solution of this equation can be obtained through its Fourier transform

$$\begin{aligned} -i\omega x(\omega) &= v(\omega) \\ -i\omega v(\omega) &= -\frac{1}{\tau_c}v(\omega) - \omega_0^2 x(\omega) + \sqrt{2D_V}\xi(\omega)\,, \end{aligned} \tag{3.14}$$

so we have that

$$x(\omega) = \frac{\sqrt{2D_V}\xi(\omega)}{\omega_0^2 - \omega^2 - i\omega/\tau_c}\,, \tag{3.15}$$

and

$$v(\omega) = \frac{-i\sqrt{2D_V}\omega\xi(\omega)}{\omega_0^2 - \omega^2 - i\omega/\tau_c}\,, \tag{3.16}$$

the time-dependent variables $X(t)$ and $V(t)$ can be obtained by Fourier inversion of these expressions. Analogously, thanks to the properties of Fourier transforms the correlation function of velocity at equilibrium , $R(\tau = |t - t'|)$, can be calculated as the inverse Fourier transform of its spectral density, defined as the modulus squared of the Fourier transformed variable, $|v(\omega)|^2 = S_v(\omega)$. So we have

$$
\begin{aligned}
R(t) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-i\omega t} S_v(\omega) d\omega \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-i\omega t}\left(\frac{-i\sqrt{2D_V}\omega\xi(\omega)}{\omega_0^2-\omega^2-i\omega/\tau_c}\right)\left(\frac{-i\sqrt{2D_V}\omega\xi(\omega)}{\omega_0^2-\omega^2-i\omega/\tau_c}\right)^* d\omega \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-i\omega t}\left(\frac{-i\sqrt{2D_V}\omega\xi(\omega)}{\omega_0^2-\omega^2-i\omega/\tau_c}\right)\left(\frac{i\sqrt{2D_V}\omega\xi(\omega)}{\omega_0^2-\omega^2+i\omega/\tau_c}\right) d\omega \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-i\omega t}\left(\frac{2D_V\omega^2 S_\xi(\omega)}{(\omega_0^2-\omega^2)^2+(\omega/\tau_c)^2}\right) d\omega\,,
\end{aligned}
\tag{3.17}
$$

which can be solved with the residue theorem considering the poles $\omega = \pm(\omega_1 + i/(2\tau_c))$, with $\omega_1 = \sqrt{\omega_0^2 - 1/(4\tau_c^2)}$. Finally, the correlation function reads

$$
R(t) = D_V\tau_c e^{-\frac{t}{2\tau_c}}\left[cos(\omega_1 t) - \frac{1}{2\omega_1\tau_c} sin(\omega_1 t)\right]\,, \tag{3.18}
$$

where two regimes must be distinguished. The underdamped regime is characterized by $\omega_0 > 1/(2\tau_c)$, corresponding to oscillating correlation function, the overdamped regime is characterized by $0 < \omega_0 < 1/(2\tau_c)$ and it is non-periodic. In the limit case $\omega_0 = 0$ we re-obtain the correlation function for a free particle.

3.2.3 Normal diffusion and Einstein-Smoluchowsky relations

By integrating the kinematic equation for the X variable, making the square and the ensemble average of both terms in the equality, we get the general expression:

$$
\sigma_x^2(t) = \langle (X(t) - X_0)^2\rangle = \int_{t_0}^{t} dt' \int_{t_0}^{t} dt'' C_v(t', t'')\,. \tag{3.19}
$$

When the system is in the stationary, equilibrium state, the above formula reduces to:

$$
\sigma_x^2(t) = \int_{t_0}^{t} dt' \int_{t_0}^{t} dt'' R(|t' - t''|) = 2\int_0^t (t-s)R(s)\, ds\,, \tag{3.20}
$$

or, equivalently:

$$
\frac{d\sigma_x^2(t)}{dt} = 2\int_0^t ds R(s)\,. \tag{3.21}
$$

These expressions were firstly studied by Taylor in 1921 [Tay22], which implicitly formulated the following theorem for the normal diffusion process (OU process).

Given the stationary correlation function $R(t)$ in Eq. 3.11, the correlation time scale is defined by:

$$
\tau = \int_0^\infty \frac{R(s)}{R(0)} ds\,, \quad R(0) = \sigma^2_{\mathrm{v,eq}}\,, \tag{3.22}
$$

where $\tau \equiv \tau_c$ in the case of the classic Langevin equation. Then, the following crucial assumption:

$$
0 \neq \tau < +\infty \tag{3.23}
$$

always determines the emergence of normal diffusion in the long-time regime:

$$
t \gg \tau \quad \Rightarrow \quad \sigma_x^2(t) = 2D_{\mathrm{X}}\, t\,; \quad D_{\mathrm{X}} := \lim_{t\to+\infty} \frac{d\sigma^2_{\mathrm{x}}}{dt}(t) = D_{\mathrm{V}}\tau_c^2 = \sigma^2_{\mathrm{v,eq}}\,\tau_c\,, \tag{3.24}
$$

independently from the details of the micro-dynamics.

D_X is the long-time spatial diffusivity of the Brownian particle:

$$D_X = \frac{\Gamma^2 \tau_c^2}{2m^2}\,, \tag{3.25}$$

When Maxwell-Boltzmann equilibrium holds we get the Einstein-Smoluchowsky relation:

$$D_X = \frac{kT}{m}\tau_c = \frac{kT}{6\pi\nu r}\,. \tag{3.26}$$

In the case of the harmonic oscillator Eq.(3.22) is equal to zero because the mean square displacement is confined in space by the harmonic potential and in the long-time limit it reaches a plateau

$$\sigma_x^2(t) = \frac{2kT}{m\omega_0^2}\,, \tag{3.27}$$

so that

$$D_X := \lim_{t\to+\infty} \frac{d\sigma_x^2}{dt}(t) = 0\,. \tag{3.28}$$

3.2.4 Fokker-Planck equation and probability distributions

The solution for the OU process $V(t)$ is described by means of the conditional probability density function (PDF) $G_{1|1}(x, v; t|x_0, v_0, t_0)$. This conditional PDF $G_{1|1}$ is also the fundamental solution of the following Fokker-Planck equation (or forward Kolmogorov equation) [Kam81]:

$$\frac{\partial G_{1|1}}{\partial t} - \frac{1}{\tau_c}\frac{\partial (v\, G_{1|1})}{\partial v} = D_V \frac{\partial^2 G_{1|1}}{\partial v^2}\,, \tag{3.29}$$

where τ_c and D_V are given by Eq. (3.8).

Given the Cauchy problem with initial condition:

$$G_{1|1}(v, t_0|v_0, t_0) = \delta(v - v_0)\,, \tag{3.30}$$

and boundary conditions $G_{1|1}(|v| \to \infty, t|v_0, t_0) = 0$, the fundamental solution of Eq. (3.29) is given by:

$$G_{1|1}(v, t|v_0, t_0) = \frac{1}{\sqrt{2\pi\sigma_v^2(t)}} \exp\left\{-\frac{(v - \langle v\rangle(t))^2}{2\sigma_v^2(t)}\right\}\,, \tag{3.31}$$

$$\langle v\rangle(t|v_0, t_0) = v_0 e^{-(t-t_0)/\tau_c}\,, \tag{3.32}$$

$$\sigma_v^2(t|v_0, t_0) = \langle (v - \langle v\rangle(t))\rangle(t|v_0, t_0) = D_V\tau_c\left(1 - e^{-2(t-t_0)/\tau_c}\right)\,, \tag{3.33}$$

It is easy to see that a long-time equilibrium distribution emerges after a initial transient, whose duration is of the order of some units of the relaxation time scale τ_c. The stationary, equilibrium solution for the OU process is rigorously defined by the limit $t \to +\infty$ of Eq. (3.33), and it becomes effective in the time range $t \gg \tau_c$:

$$G_{\text{eq}}(v) = \frac{1}{\sqrt{2\pi D_V \tau_c}} \exp\left\{-\frac{v^2}{2D_V\tau_c}\right\}\,. \tag{3.34}$$

We have the following relationships:

$$\sigma^2_{\mathrm{v,eq}} = \langle (v - m_v)^2 \rangle_{\mathrm{eq}} = D_V \tau_c = \frac{\Gamma^2 \tau_c}{2m^2} = \frac{\Gamma^2}{2m\gamma}\,, \tag{3.35}$$

being $m_v = \langle v \rangle_{\mathrm{eq}}$, where the subscript eq indicates that the average is computed with the equilibrium PDF. The average velocity m_v is zero if there are no external forcing determining a mean drift velocity.

The velocity distribution of a one-dimensional ideal gas in thermodynamical equilibrium is given by the well-known Maxwell-Boltzmann velocity distribution:

$$p_{\mathrm{eq}}(v) = \sqrt{\frac{m}{2\pi kT}} \exp\left\{ -\frac{mv^2}{2kT} \right\}, \tag{3.36}$$

being k and T the Boltzmann constant and temperature, respectively. Comparing (3.34) and (3.36) we get the particular case of the relations (3.35) for the Maxwell-Boltzmann case:

$$\sigma^2_{\mathrm{v,eq}} = \frac{kT}{m}\,. \tag{3.37}$$

This relationship makes evident that friction and velocity diffusivity are related to each other through the thermodynamical internal energy of the gas, i.e., the mean kinetic energy of the atoms or molecules in the gas, with important consequences over the process dynamics as the fluctuation-dissipation theorem:

$$\frac{\Gamma^2}{2m\gamma} = \frac{kT}{m}\,. \tag{3.38}$$

The amplitude of the noise is then related to the internal energy and to the drag by the relation $\Gamma^2 = 2kT\gamma$. Fluctuation dissipation theorem is treaten in details by Kubo [Kub66] for classic Langevin and generalized Langevin equation (GLE), physically this relationship arises because the frictional force and the random driving force of the motion have the same origin: the random impacts of the moving particle with the surround.

3.3 Heterogeneous ensemble of Brownian particles

Let $\kappa = 1, \ldots, N$ label the particles of the ensemble and let m^κ be the mass of the κ-particle, then position and velocity of the centre of mass are:

$$x_{\mathrm{CM}} = \sum \frac{m^\kappa}{M} x^\kappa\,, \quad v_{\mathrm{CM}} = \sum \frac{m^\kappa}{M} u^\kappa\,, \tag{3.39}$$

with $u^\kappa = dx^\kappa/dt$ and $M = \sum m^\kappa$, and the dynamics is given by

$$M \frac{dv_{\mathrm{CM}}}{dt} = \sum m^\kappa \frac{du^\kappa}{dt} = \sum F^\kappa\,. \tag{3.40}$$

Let us introduce capitol letters to denote the stochastic variables, then the stochastic dynamics of the centre of mass is

$$dX_{\mathrm{CM}} = V_{\mathrm{CM}} dt\,, \tag{3.41a}$$

$$dV_{\mathrm{CM}} = \sum \frac{m^\kappa}{M} dU^\kappa\,, \tag{3.41b}$$

and the motion of any Brownian κ-particle of the ensemble is governed by the Langevin equation

$$dX^\kappa = U^\kappa dt\,, \tag{3.42a}$$
$$dU^\kappa = -\frac{U^\kappa}{\tau^\kappa}dt + \sqrt{2\,D_{V,\kappa}}\,dW^\kappa\,, \tag{3.42b}$$

where dW^κ is a Wiener process with zero mean and variance dt, γ^κ the friction coefficient per unit mass corresponding to the κ-particle and $\sqrt{2\,D_{V,\kappa}}$ is the strength of the noise provided by the microscopic scales of the surround and then experienced by the mesoscale Brownian particle:

$$\frac{1}{\tau^\kappa} = \frac{6\pi\nu r_\kappa}{m^\kappa}\,, \quad D_{V,\kappa} = \frac{kT}{\tau^\kappa m^\kappa}\,. \tag{3.43}$$

We remind that the friction coefficient is given by the Stokes law $\gamma^\kappa = 6\pi\nu r^\kappa$ where ν is the viscosity of the medium (identically experienced by the mesoscale Brownian particles) and r^κ the radius of the κ-Brownian particle.

By integration of (3.41) through (3.42), stochastic position and velocity of the centre of mass are

$$X_{\rm CM} = \sum \frac{m^\kappa}{M}X^\kappa\,, \quad V_{\rm CM} = \sum \frac{m^\kappa}{M}U^\kappa\,. \tag{3.44}$$

Consider

$$X^\kappa = \sqrt{D_{V,\kappa}}\,X_0^\kappa\,, \quad U^\kappa = \sqrt{D_{V,\kappa}}\,U_0^\kappa\,, \tag{3.45a}$$

and from (3.42) it holds

$$dX_0^\kappa = U_0^\kappa dt\,, \tag{3.46a}$$
$$dU_0^\kappa = -\frac{U_0^\kappa}{\tau^\kappa}dt + \sqrt{2}\,dW^\kappa\,. \tag{3.46b}$$

Hence the motion of the centre of mass results to be statistically described also by the stochastic process

$$dX_{\rm CM} = V_{\rm CM}dt\,, \tag{3.47a}$$
$$dV_{\rm CM} = \sum \frac{m^\kappa}{M}\sqrt{D_{V,\kappa}}\,dU_0^\kappa = \sum \frac{\sqrt{kT\gamma^\kappa}}{M}dU_0^\kappa\,. \tag{3.47b}$$

Particle density is $3m/(4\pi r^3)$ for spherical particles, in general for any particles different densities translate into a population of masses $\rho(m)$ and a population of timescales $q(\tau)$. Since m^κ are independent identically distributed variables, M distribution is determined by $\rho(m)$. If the population of masses is an infinitely divisible density then the distribution of M is the same distribution of m. If all the particles experience the same friction and differ for their mass only $\gamma^\kappa = \gamma_0 = 6\pi\nu r_0\,, \forall\kappa$ and $\rho(m) = q(\tau)$. After integration of (3.47) through (3.46), stochastic position and velocity of the centre of mass are

$$X_{\rm CM} = \frac{\sqrt{D_{V,0}}}{M}\sum X_0^\kappa\,, \quad V_{\rm CM} = \frac{\sqrt{D_{V,0}}}{M}\sum U_0^\kappa\,, \tag{3.48}$$

where $D_{V,0} = kT\gamma_0$.

Consider the Central Limit Theorem (CLT) in the Lyapunov sense:

$$\sum Y^\kappa \to \sqrt{\sum \langle (Y^\kappa)^2 \rangle}\, Y_0\,, \quad Y_0 \sim \mathcal{N}(0,1)\,, \tag{3.49}$$

where Y^κ is a Gaussian variable with zero mean and density function $p_\kappa(y)$ such that $\langle (Y^\kappa)^2 \rangle = \int y^2 p_\kappa(y) dy$ and $\mathcal{N}(0,1)$ is the normal distribution with zero mean and unit variance. Then the processes

$$\sum X_0^\kappa = X_G\,, \quad \sum U_0^\kappa = V_G\,, \tag{3.50}$$

are two Gaussian processes, because X_0^κ and U_0^κ are Gaussian processes with classical scaling according to (3.46), but X_G and V_G may display anomalous scaling because of the population of timescales. In fact

$$\begin{aligned} \langle V_G(t) V_G(s) \rangle &= \langle \sum U_0^\kappa(t) \sum U_0^\kappa(s) \rangle \\ &= \sum \langle U_0^\kappa(t) U_0^\kappa(s) | \tau^\kappa \rangle \\ &= \int \langle U_0^\kappa(t) U_0^\kappa(s) | \tau \rangle q(\tau)\, d\tau\,, \end{aligned} \tag{3.51}$$

where the distribution $q(\tau)$ modifies the classical scaling displayed by $\langle U_0^\kappa(t) U_0^\kappa(s) | \tau \rangle$. In Section 3.5.1, it will be explicitly demonstrated that a properly chosen population of τ in the classic Langevin framework induces the appearance of a Hurst exponent $1/2 < H < 1$. The prefactor in (3.48) is a non-negative, time and space independent random variable, then, by setting $\sqrt{D_{V,0}}/M = \sqrt{\Lambda}$, process (3.48) can be re-written as

$$X_{\mathrm{CM}} = \sqrt{\Lambda}\, X_G\,, \quad V_{\mathrm{CM}} = \sqrt{\Lambda}\, V_G\,. \tag{3.52}$$

that is a stochastic process based on the same constructive approach adopted by Mura [Mur11] to built up the ggBm [Mur11; MP08b; MM09], i.e., a Gaussian process times a non-negative random variable.

Furthermore, by plugging (3.42) into (3.41) we can derive the following Langevin-type dynamics:

$$\begin{aligned} dV_{\mathrm{CM}} &= \frac{1}{M} \sum \left\{ -\gamma_0 U^\kappa dt + \sqrt{2 D_{V,0}}\, dW^\kappa \right\} \\ &= -\frac{\gamma_0}{M} \sum U^\kappa dt + \frac{\sqrt{2 D_{V,0}}}{M} \sum dW^\kappa\,, \end{aligned} \tag{3.53}$$

despite γ_0 is a constant, the sum over U^κ may display a time dependent timescale because of the heterogeneity of the mass, that is consistent with the idea that we cannot switch to the overdumped limit in transient anomalous diffusion regime. In

fact:

$$\begin{aligned}
\sum \gamma_0 U^\kappa &= \sum \gamma_0 \sqrt{D_{V,\kappa}}\, U_0^\kappa = \sqrt{D_{V,0}} \sum \frac{U_0^\kappa}{\tau^\kappa} \\
&\to \sqrt{D_{V,0}} \sqrt{\sum \left\langle \left(\frac{U_0^\kappa}{\tau^\kappa} \right)^2 \right\rangle}\, V_0 \,, \quad V_0 \sim \mathcal{N}(0,1) \\
&= \sqrt{D_{V,0}} \frac{\sqrt{\sum \left\langle \left(\frac{U_0^\kappa}{\tau^\kappa} \right)^2 \right\rangle}}{\sqrt{\sum \langle (U_0^\kappa)^2 \rangle}} \sqrt{\sum \langle (U_0^\kappa)^2 \rangle}\, V_0 \\
&= \frac{1}{\tau_{\text{eff}}} \sqrt{\sum \langle (\sqrt{D_{V,0}} U_0^\kappa)^2 \rangle}\, V_0 = \frac{1}{\tau_{\text{eff}}} \sqrt{\sum \langle (m^\kappa U^\kappa)^2 \rangle}\, V_0 \\
&\to \frac{1}{\tau_{\text{eff}}} \sum m^\kappa U^\kappa = \frac{1}{\tau_{\text{eff}}} M\, V_{\text{CM}} \,,
\end{aligned} \tag{3.54}$$

where

$$\frac{1}{\tau_{\text{eff}}} = \frac{\sqrt{\sum \left\langle \left(\frac{U_0^\kappa}{\tau^\kappa} \right)^2 \right\rangle}}{\sqrt{\sum \langle (U_0^\kappa)^2 \rangle}} = \left[\frac{\iint \frac{u_0^2}{\tau^2} p(u_0; t|\tau) q(\tau)\, d\tau}{\iint u_0^2 p(u_0; t|\tau) q(\tau)\, d\tau} \right]^{1/2} , \tag{3.55}$$

which is a function of time, i.e., $\tau_{\text{eff}} = \tau_{\text{eff}}(t)$, $p(u_0; t|\tau)$ are Gaussian densities (3.46), and

$$dW_{\text{CM}} = \sum dW^\kappa \,, \quad \langle (dW_{\text{CM}})^2 \rangle = N \langle (dW^\kappa)^2 \rangle \,. \tag{3.56}$$

Hence finally

$$dV_{\text{CM}} = -\frac{1}{\tau_{\text{eff}}} V_{\text{CM}} dt + \frac{\sqrt{2D_{V,0}}}{M} dW_{\text{CM}} \,. \tag{3.57}$$

In the derivation of (3.54), the first line follows from (3.45) and the second from the CLT (3.49) noting that for any κ-particle the parameter τ^κ is fixed and then the process U_0^κ is Gaussian, see (3.46). The third line contains the multiplication and division by $\sqrt{\sum \langle (U_0^\kappa)^2 \rangle}$ and in the fourth τ_{eff} is introduced, $D_{V,0}$ moved into the square root and the second equality follows from (3.45). Finally, the last line follows again from the CLT (3.49) and the last equality from (3.44).

The three processes (3.41), (3.47) and (3.57) are statistically equivalent as displayed in Fig. 3.1.

The time-dependence of the drift term $1/\tau_{\text{eff}}$ shows that anomalous diffusion emerges in the studied system during an *intermediate asymptotic regime* [Bar79] and consequently the need to adopt an *underdamped formulation*. These two features are consistent with the relation between them and anomalous diffusion already provided in the case of the underdamped scaled Brownian motion [Bod+16], and implicitly by the role of friction through a complex potential [San+04].

If the radii are kept random as well the calculation is less straightforward. If the population of particles with a given radius may feel the whole population of timescales, thus the random variables $\gamma^\kappa, \tau^\kappa$ remain independent and the same approach can be applied without any particular complication for the heterogeneous population of radii. The price to pay however is that the population of masses have a different mean value for each particle type, which depends on r^κ:

$$\rho_\kappa(m) = q(\tau) r^\kappa \,, \tag{3.58}$$

this can be understood intuitively in terms of density, given the same distribution of density, a larger radius is associated with a larger mean value of the masses.

As a concluding remark on this part, if we relax the fluctuation-dissipation theorem and assume that all the particles feel the same noise, both the populations of masses and radii contribute by means of the timescale $\tau = m/\gamma = m/(6\pi\nu r)$ to the emergence of the anomalous scaling as shown in (3.51), while only the population of masses contributes to the shape of the probability density functions by means of the prefactor $1/M = 1/\sum m^\kappa$.

The ggBm proposed by Mura is recovered from (3.52) in the case X_G is the fractional Brownian motion and $\Lambda \sim M_\beta(\lambda)$ where $M_\beta(\lambda)$, $0 < \beta < 1$, is the M-Wright/Mainardi function [MMP10; Pag13].

A physically sound choice of the PDF of m and τ can be done on the basis of experimental estimation of the PDF of the diffusion coefficient D, here denoted by $f(D)$. In particular, we consider the generalized Gamma distribution, i.e.,

$$f(D) = \frac{\eta}{D_* \Gamma(\nu/\eta)} D^{\nu-1} \, \mathrm{e}^{-(D/D_*)^\eta} , \tag{3.59}$$

that is based on experimental evidence in cases with random diffusivity. The generalized Gamma distribution includes as special cases both the Gamma and the exponential distributions, and allows for the stretched and compressed exponential distributions in agreement with some experimental data [Jav+13; Hap09; Man+15; Jeo+16].

We consider the case when the fluctuation-dissipation theorem holds, then $r^\kappa = r_0$ and from definitions we obtain:

$$\frac{1}{\tau^\kappa} = \frac{\gamma^\kappa}{m^\kappa} = \frac{6\pi\nu r_0}{m^\kappa} = \frac{\gamma_0}{m^\kappa} . \tag{3.60}$$

If $\rho(m)$ and $q(\tau)$ are the PDFs of m and τ, respectively, then from normalization condition we have that

$$\rho(m) = \frac{1}{\gamma_0} q\left(\frac{m}{\gamma_0}\right) , \quad q(\tau) = \gamma_0 \, \rho(\gamma_0 \, \tau) . \tag{3.61}$$

In this framework the diffusion coefficient is

$$D = \frac{k_\mathrm{B} T 6\pi\nu r_0}{(m^\kappa)^2} = \frac{k_\mathrm{B} T \gamma_0}{(m^\kappa)^2} , \tag{3.62}$$

and the corresponding distributions of m and τ in terms of $f(D)$ are

$$\rho(m) = \frac{1}{2m^3} f\left(\frac{k_\mathrm{B} T \, \gamma_0}{m^2}\right) , \quad q(\tau) = \frac{1}{2\gamma_0^2 \tau^3} f\left(\frac{k_\mathrm{B} T}{\gamma_0 \tau^2}\right) . \tag{3.63}$$

3.4 Test particle in heterogeneous ensemble

The formalism previously derived can be used also to describe the stochastic dynamics of a single mesoscopic test-particle that is immersed into a surround composed by the studied heterogeneous ensemble of Brownian particles. Let us consider a system with the following coupling between the surround and the test-particle

$$dU^\kappa = -\frac{1}{\tau^\kappa} U^\kappa dt - \alpha^\kappa (U^\kappa - V) dt + \sqrt{2D_{V,\kappa}} dW^\kappa , \tag{3.64a}$$

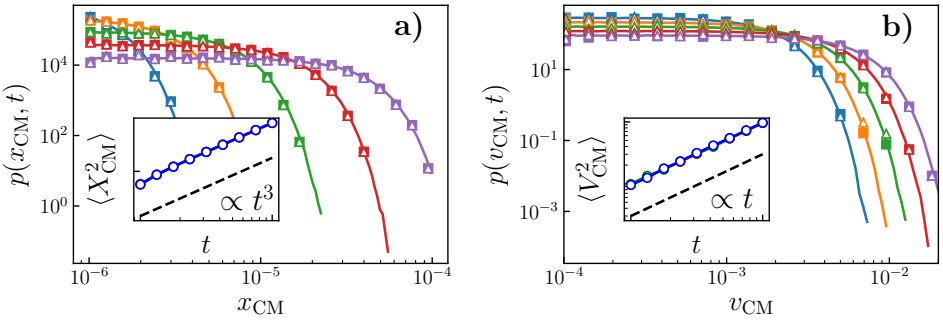

FIGURE 3.1: (color online) Statistical equivalence of the processes (3.41), (3.47) and (3.57). Panel a) shows the density function of the center of mass $p(x_{\rm CM};t)$ and in the inset the variance $\langle X^2_{\rm CM}\rangle$; and panel b) shows the density function $p(v_{\rm CM};t)$ and the variance $\langle V^2_{\rm CM}\rangle$. The PDFs of the process given by Eqs. (3.41) and (3.42) are presented via filled squares (■), Eqs. (3.46) and (3.47) are shown with empty triangles (△) and the process of Eq. (3.57) are the solid line. Different colors represent different times. Insets: green line refers to Eq. (3.41), blue line to Eq. (3.47) and red line refers to Eq. (3.57). Simulations have been performed with $\rho(\tau)$ and $p(m)$ given by Eqs. (3.63) and (3.59) with $\nu = 4/3\,, \eta = 3/4\,, \gamma_0 = 10^5$.

$$dV = \sum \alpha^\kappa (U^\kappa - V)dt\,, \tag{3.64b}$$

where U^κ and V are the velocities of the Brownian κ-particle of the ensemble and the test-particle, respectively. After setting $\sum \alpha^\kappa = A$, the velocity of the test-particle results to be

$$V = V_0\,\mathrm{e}^{-At} - \int_0^t \mathrm{e}^{-A(t-s)} \sum \alpha^\kappa U^\kappa ds\,. \tag{3.65}$$

In the limit $A \to \infty$, after multiplication and division by A of the integrand function and by using the following definition of the δ-function

$$\delta(r) = \lim_{\varepsilon\to 0} \frac{1}{\varepsilon}\mathrm{e}^{-r/\varepsilon}\,, \quad 0 \le r < \infty\,, \tag{3.66}$$

we have that

$$V \simeq -\frac{\sum \alpha^\kappa U^\kappa}{A}\,. \tag{3.67}$$

First by plugging (3.67) into (3.64a) and later by plugging this expression for dU^κ into (3.67), we obtain

$$\begin{aligned} dV &= -\frac{\sum \alpha^\kappa U^\kappa/\tau^\kappa_*}{A}\,dt \\ &\quad -\frac{\sum \alpha^\kappa \alpha^\kappa (\sum \alpha^\kappa U^\kappa)}{A^2}\,dt + \frac{\sum \alpha^\kappa \sqrt{2D_{V,\kappa}}dW^\kappa}{A}\,, \end{aligned} \tag{3.68}$$

with $1/\tau^\kappa_* = 1/\tau^\kappa + \alpha^\kappa$.

In the case $\alpha^\kappa = \alpha$ for all κ, we have that $A = \alpha N$ and $V \simeq -\sum U^\kappa/N$, hence

$$dV = -\alpha V dt + \frac{1}{N}\sum \left[-\frac{U^\kappa}{\tau^\kappa_*}dt + \sqrt{2D_{V,\kappa}}dW^\kappa\right]\,. \tag{3.69}$$

We can recognize that the second term in the RHS is the stochastic dynamics of

the centre of mass of an ensemble of Brownian particles with the same mass and a population of timescales τ_*^κ, i.e.,

$$dV = -\alpha V dt + dV^*_{\rm CM}\,, \tag{3.70}$$

where $dV^*_{\rm CM}$ is given by (3.57) with the proper changes, i.e., by setting in its derivation $m^\kappa = m$ for all κ and $M = mN$. Then the motion of the test-particle results to be forced by the ensemble of the Brownian particles and shows a drift due to the coupling with the surround.

We observe that in the passive tracer limit $\alpha \to 0$ and $\alpha N \to \infty$ (as for example the case $\alpha \propto 1/\sqrt{N}$ and $N \to \infty$), the motion of the tracer particle reduces to that of the centre of mass.

From statistical arguments about the symmetry of the density function of velocities V, by replacing V with $-V$ in (3.67) the stochastic dynamics of the test-particle results to be

$$dV = \frac{1}{A}\sum \alpha^\kappa\, dU^\kappa\,, \tag{3.71}$$

and if $\alpha^\kappa \propto m^\kappa$ then $A \propto M$ such that (3.71) is the analog of (3.47) after the replacement of α^κ with m^κ. Then, still in the case $\alpha^\kappa \propto m^\kappa$, from (3.71) the process analog of (3.52) follows by changing $U^\kappa = \sqrt{D_{V,\kappa}}U_0^\kappa$ (see (3.45)) and it holds

$$dV = \frac{1}{M}\sum m^\kappa \sqrt{D_{V,\kappa}}\, dU_0^\kappa = \frac{\sqrt{D_{V,0}}}{M}\sum dU_0^\kappa\,, \tag{3.72}$$

such that, by setting $\sqrt{D_{V,0}}/M = \sqrt{D}$, $X_G = \sum X_0^\kappa$ and $V_G = \sum U_0^\kappa$, we obtain the following ggBm-like representation

$$X = \sqrt{D}\, X_G\,, \quad V = \sqrt{D}\, V_G\,, \tag{3.73}$$

that suggests the application of model (3.73) to study biological systems, in analogy with the promising application of the ggBm [MG+16]. When $D \sim M_\beta(\lambda)$ the particle displacement density function is [Pag12]

$$\begin{aligned} p(x;t) &= \frac{1}{\sqrt{4\pi\lambda\, t^{2H}}}\int_0^\infty \exp\left\{-\frac{x^2}{4\lambda\, t^{2H}}\right\} M_\beta(\lambda)\, d\lambda \\ &= \frac{1}{2\, t^H}\, M_{\beta/2}\left(\frac{|x|}{t^H}\right)\,, \end{aligned} \tag{3.74}$$

and in general it displays anomalous and non-Gaussian diffusion and solves a fractional diffusion equation in the Erdélyi–Kober sense [Pag12]. Special cases are the time-fractional diffusion ($\beta = 2H$), the Brownian non-Gaussian motion ($2H = 1$), the Gaussian non-Brownian motion ($\beta = 1$) and the classical diffusion ($\beta = 2H = 1$).

Finally, still in the case $\alpha^\kappa \propto m^\kappa$ the process analog of (3.57) has the same equation, i.e.,

$$dV = -\frac{1}{\tau_{\rm eff}}\, V dt + \frac{\sqrt{2D_{V,0}}}{M}\, dW\,, \tag{3.75}$$

where $dW = \sum dW^\kappa$ and $\langle (dW)^2\rangle = N\langle (dW^\kappa)^2\rangle$.

As a concluding remark on this second part, we highlight that, in the case $\alpha^\kappa \propto m^\kappa$, the motion of the test-particle is analogous to the motion of the centre of mass of the ensemble of Brownian particles. It results to be affected by both the populations of mass and radii by means of the timescale $\tau = m/\gamma = m/(6\pi\nu r)$ for what

concerns the emergence of the anomalous scaling as shown in (3.51), while only the population of mass contributes to the shape of the probability density functions by means of the prefactor $1/M = 1/\sum m^{\kappa}$. If m^{κ} are independent identically stable distributed variables, then the population of mass $\rho(m)$ is an infinitely divisible density and M follows the same distribution $\rho(m)$.

3.5 Building anomalous diffusion regimes

We showed in Section 3.2 that, when the assumption of Eq. (3.23) is still valid, there exists a well-defined correlation time scale τ_c and the emergence of a normal diffusion scaling ($\langle X^2 \rangle \sim t$) in the long-time limit ($t \gg \tau_c$). As a consequence, the emergence of anomalous diffusion is strictly connected to the failure of the assumption (3.23). When the long time scaling of the MSD is anomalous, i.e.

$$\sigma_x^2(t) \propto t^{\phi}\,, \tag{3.76}$$

the relation between MSD and the VACF, which holds in general because the relation $dX(t)/dt = V(t)$ is maintained, leads to:

$$\lim_{t\to\infty} \frac{1}{2}\frac{d\sigma_x^2(t)}{dt} = \int_0^{\infty} ds R(s) \propto \lim_{t\to\infty} t^{(\phi-1)} ds\,. \tag{3.77}$$

For $\phi \neq 1$ it is straightforward to show that the assumption of finite and non-zero τ_c fails. Depending on the range of ϕ super-diffusive and sub-diffusive regimes can be distinguished on the bases of this result.

The super-diffusive regime $\phi > 1$ is consistent with an infinite value of this integral in the infinite time limit:

$$\int_0^{\infty} \frac{R(\tau)}{R(0)} d\tau = +\infty\,, \tag{3.78}$$

Instead sub-diffusive regime $\phi < 1$ is characterized by a vanishing integral as in the case of the Langevin oscillator:

$$\int_0^{\infty} \frac{R(\tau)}{R(0)} d\tau = 0\,, \tag{3.79}$$

This case occurs only when anti-correlation appears, i.e., there exist time lags τ such that $R(\tau) < 0$ (e.g., the anti-persistent Fractional Brownian Motion, with $H < 0.5$).

In the following, we use the fundamental results of the Langevin description to derive a model for anomalous diffusion. The basic idea is that the observed anomalous diffusion emerges as a linear superposition of independent contributions.

Each contribution is given by a single realization of a classic process associated with a particular value of the parameter, while the observed process feels the entire distribution characterizing the parameter. Starting from this consideration it results that also far from the equilibrium we can write the VACF of the complex process as:

$$C_v(t_1, t_2) = \int C_v(t_1, t_2; p_1, p_2, ...) P(p_1, p_2, ...) dp_1 dp_2 ...\,. \tag{3.80}$$

Apart from the physical origin of this mechanism, the superposition is driven by the randomness of some parameters $p_1, p_2, ...$ These parameters are here treated as an independent random variable with a given PDF. Here we assume that the

randomness lies in the parameters D_V and/or τ_c or, equivalently, in the parameters γ and/or Γ. Each parameter is defined statistically independent from the other ones. This is not in conflict with the FDT as far as we consider that the single components are not real but a mathematical counterpart of a complex problem, then FDT can be satisfied by the real process, defined as the superposition of simpler processes, but violated by the single components. Moreover in the special cases where FDT can be violated the single components can be interpreted as real processes. Within this Chapter, if other parameters characterize the process, for example, the characteristic frequency of the harmonic oscillator, they are kept not random.

Starting from the basic formulas of the considered stochastic process, the parameter PDF driving the linear superposition is here chosen in such a way to determine the emergence of global (effective) statistical features in agreement with fractional anomalous diffusion. We impose the emergence of the following global properties: correlation function with an asymptotic power-law decay, anomalous diffusion in the MSD (variance increasing with some power of the time), PDF $P(x,t)$ compatible with fractional diffusion. The first and second considerations are strictly connected to each other by Eq.(3.20), this relation will become clearer considering its Laplace transformation in the next section.

Laplace transformation mapping of global properties

Laplace transformations approach allows to simplify calculations and easily check limits of a function through initial and final value theorems. Given a function $f(t)$ and its Laplace transform $\mathcal{L}[f(t)](s)$ it holds:

$$\lim_{t\to+\infty} f(t) = \lim_{s\to 0} s\cdot\mathcal{L}[f(t)](s)\,, \tag{3.81}$$

and

$$f(0_+) = \lim_{s\to+\infty} s\cdot\mathcal{L}[f(t)](s)\,. \tag{3.82}$$

If the limit does not exist we can eventually find the asymptotic behavior for $t\to+\infty$. Consider the Laplace transform of a measure F, defined by the improper distribution function $F\{0,x\} = \int_0^x f(y)dy$, with density $f(x)$ (where improper means that it could be not normalized), given by

$$\omega(s) = \int_0^\infty e^{-sx}F\{dx\} = \mathcal{L}[f(t)](s)\,, \tag{3.83}$$

and introduce two positive variables t,τ, such that:

$$t\tau = 1\,, \tag{3.84}$$

so that $t\to+\infty$ when $\tau\to+0$, and apply the transformation $x=ty$. It results that $\omega(\tau s)$ is the Laplace transform of $F(ty)$ respect to y defined in Eq. (3.83). If such Laplace transform of F exists, it is unique for any t and its asymptotic behavior for $t\to+\infty$ is uniquely determined by the behavior of its Laplace transform near the origin, under reasonable conditions, defined by the following theorem [Fel71].

Each of the relation, for $\rho\geq 0$,

$$\frac{\omega(\tau s)}{\omega(\tau)} \to \frac{1}{s^\rho}\,, \quad \tau\to 0\,, \tag{3.85}$$

and

$$\frac{F(tx)}{F(t)} \to x^{\rho}\,, \quad t \to \infty\,, \tag{3.86}$$

implies the other as well as:

$$\omega(\tau) \sim F(t)\Gamma(\rho+1)\,, \tag{3.87}$$

for $\tau \to 0$ or, equivalently, $t \to \infty$.

Then considering the relation between F and f, for power law asymptotic behavior it follows that:

$$\tau\omega(\tau) \propto f(t)\,, \quad \tau = 1/t\,, \quad \tau \to 0\,. \tag{3.88}$$

Then any choice of correlation function and distribution for the parameters of the dynamics should satisfy the following global properties.

The scaling of the MSD in the long time limit is a power fraction of time:

$$s \cdot \mathcal{L}[\sigma_x^2(t)](s) \sim \frac{1}{s^{\phi}} \sim t^{\phi}\,, \quad s \to 0\,, \tag{3.89}$$

then:

$$\lim_{s\to 0} s^{\phi+1} \cdot \mathcal{L}[\sigma_x^2(t)](s) = 1\,. \tag{3.90}$$

The MSD at time zero is zero:

$$\lim_{s\to+\infty} s \cdot \mathcal{L}[\sigma_x^2(t)](s) = 0\,. \tag{3.91}$$

Laplace transformations of VACF and MSD are related by the following expression, resulting from Laplace transformation of Eq.(3.20)

$$\mathcal{L}\left[\sigma_x^2(t)\right](s) = \frac{2}{s^2}\mathcal{L}\left[R(t)\right](s)\,. \tag{3.92}$$

Then we may add some conditions on the VACF behavior, perhaps the scaling of the VACF in the long time limit should be a power law as well, with exponent $-\alpha = -(2-\phi)$:

$$\lim_{s\to 0} 2s^{\phi-1} \cdot \mathcal{L}[R(t)](s) = 1\,, \tag{3.93}$$

with $0 < \nu < 1$ or $1 < \nu < 2$ depending if we are describing super-diffusive or sub-diffusive regimes respectively. Then it still holds:

$$s \cdot \mathcal{L}[R(t)](s) \sim \frac{1}{s^{\phi-2}} = \frac{1}{s^{-\nu}} \sim t^{-\nu}\,, \quad s \to 0\,. \tag{3.94}$$

The VACF at time zero is $R(0) = \sigma^2_{v,eq.}$ finite positive number, which imply that $b(\tau_c)$, as well as $h(D_V)$, must have finite mean in order to describe a finite energy system:

$$\lim_{s\to+\infty} s \cdot \mathcal{L}[R(t)](s) = \langle D_V\rangle\langle\tau_c\rangle\,, \tag{3.95}$$

where the distribution functions of the parameters $b(\tau_c)$ and $h(D_V)$ are normalized to unity:

$$\mathcal{L}[b(\tau_c)](0) = 1, \quad \mathcal{L}[h(D_V)](0) = 1\,. \tag{3.96}$$

Normalization of the distribution function for the timescale can be obtained also from the VACF of the free particle Langevin equation considering that:

$$-\frac{dR(t)}{dt} = \int_0^\infty e^{-t/\tau_c} b(\tau_c) d\tau_c \,, \tag{3.97}$$

which leads to:

$$\lim_{s\to+\infty} s \cdot \mathcal{L}[-\frac{dR(t)}{dt}](s) = -\frac{dR(t)}{dt}|_{t=0} = 1 \,. \tag{3.98}$$

In the next Sections we will sometimes adopt, for sake of convenience, the improper notation for which the asymptotic behavior is written as the result of a limit. In those cases the equivalence wants to represent the asymptotic behavior of the function and not the value of the limit, which is indefinite.

3.5.1 Population of timescales τ_c and anomalous time scaling

In principle there exists an infinite number of distribution functions $b(\tau_c)$ that satisfy the conditions in Eqs. 3.93, 3.95 and 3.96. However it is not straightforward to determine a class of these functions. Here we propose a suitable function as reference example and we show explicitly that all the required properties are satisfied. We also comment how the macroscopic forces enter into the calculation modifying the diffusive regime maintaining the same parameters randomization.

Let's consider a distribution of time scales of the kind:

$$b(\tau_c) = \frac{\alpha}{\Gamma(1/\alpha)} \frac{1}{\tau_c} L_\alpha^{-\alpha}\left(\frac{\tau_c}{\tau_*}\right) \,, \tag{3.99}$$

where $L_\alpha^{-\alpha}(z)$ is the extremal Levy density, with $0 < \alpha < 1$ and $\tau_* = (\langle\tau_c\rangle \frac{\Gamma(1/\alpha)}{\alpha})$. The distribution is characterized by $\langle\tau_c\rangle$, the mean timescale of the process, which could be estimated experimentally. This distribution satisfies all the global properties requested by the theory.

The normalization constant $C = \frac{\alpha}{\Gamma(1/\alpha)}$ can be obtained imposing $\mathcal{L}[b(\tau_c)](0) = 1$, with $b(\tau_c) = C \cdot \frac{1}{\tau_c} L_\alpha^{-\alpha}(\tau_c)$:

$$\begin{aligned} \mathcal{L}[b(\tau_c)](s) &= C \cdot \int_s^\infty \mathcal{L}[L_\alpha^{-\alpha}(\tau_c)](\xi) d\xi \\ &= C \cdot \int_s^\infty e^{-\xi^\alpha} d\xi \\ x &= \xi^\alpha \\ &= C \cdot \int_s^\infty \frac{1}{\alpha} e^{-x} x^{1/\alpha - 1} dx \,, \end{aligned} \tag{3.100}$$

then:

$$\begin{aligned} \mathcal{L}[b(\tau_c)](0) &= C \cdot \int_0^\infty \frac{1}{\alpha} e^{-x} x^{1/\alpha - 1} dx \\ &= C \cdot \frac{\Gamma(1/\alpha)}{\alpha} = 1 \,. \end{aligned} \tag{3.101}$$

Within the new equilibrium condition, the VACF at time zero is $R(0) = \langle D_V\rangle\langle\tau_c\rangle$. In order to describe a real process $R(0)$ should have a positive and finite value because it represents the average kinetic energy associated with the process as well the temperature of the system. This condition is satisfied if $b(\tau_c)$ and $h(D_V)$ have

positive finite means. Concerning $b(\tau_c)$ we obtain the consistent relation:

$$\int_0^\infty \tau_c b(\tau_c) d\tau_c = \frac{\alpha}{\Gamma(1/\alpha)} \int_0^\infty L_\alpha^{-\alpha}\left(\frac{\tau_c}{\tau_*}\right) d\tau_c = \langle \tau_c \rangle \,. \tag{3.102}$$

Superdiffusive case

Anomalous super-diffusion is defined by the MSD growing with a nonlinear power of the time:

$$\sigma_x^2(t) \sim t^\phi \quad ; 1 < \phi < 2 \,. \tag{3.103}$$

From the subsection 3.5 we have that the condition given in Eq. (3.23) must fail going to infinity in the long-time limit. An asymptotic power-law behavior for the VACF is a typical condition violating Eq. (3.23). To build this asymptotic behavior, let us first rewrite the velocity VACF of the single OU process, as given by Eq. (3.10), making the variable dependence more explicit and avoiding the average over the initial velocities:

$$\langle V(t_1)V(t_2)|v_0, D_{_V}, \tau_c\rangle_\xi = \left(v_0^2 - D_{_V}\tau_c\right) e^{-(t_1+t_2-2t_0)/\tau_c} + D_{_V}\tau_c e^{-|t_1-t_2|/\tau_c} \,. \tag{3.104}$$

The dependence from the initial velocity v_0 means that the average is conditioned to the fictitious trajectories with initial velocity given by v_0. In the single OU process, the equilibrium condition is associated with the relationship (3.35): $\sigma^2_{\mathrm{v,eq}} = \langle v_0^2\rangle_{\mathrm{eq}} = D_{_V}\tau_c$. For the single OU process, this relationship follows the equilibrium condition given by the Gaussian distribution of Eq. (3.34), which is reached by the OU process in the long-time limit $t \gg \tau_c$. On the contrary, in the extended model the Gaussian law of Eq. (3.34) is no longer the equilibrium velocity PDF, so that Eq. (3.35) is not valid in general.

It is worth noting that Eq. (3.35) implies that the statistics of v_0 at equilibrium depends on τ_c. Then, even if the equilibrium of the single OU process fails, let us assume that, even in the extended global equilibrium condition, the initial velocity distribution depends on τ_c: $v_0 = F(\tau_c, ...)$. Let us now apply the averaging over the random parameters (D_V, τ_c) and over the initial velocity v_0 to Eq. (3.104):

$$\langle V(t_1)V(t_2)\rangle := \langle\langle V(t_1)V(t_2)|v_0, D_{_V}, \tau_c\rangle_\xi\rangle_{v_0,D_{_V},\tau_c} =$$

$$= \langle\left(v_0^2 - D_{_V}\tau_c\right) e^{-(t_1+t_2-2t_0)/\tau_c}\rangle_{v_0,D_{_V},\tau_c} + \langle D_{_V}\tau_c e^{-|t_1-t_2|/\tau_c}\rangle_{v_0,D_{_V},\tau_c} \,.$$

In order to get the global equilibrium condition, we must put to zero the first, non-stationary, term:

$$\langle v_0^2(\tau_c) e^{-(t_1+t_2-2t_0)/\tau_c}\rangle_{\tau_c} = \langle D_{_V}\rangle D_V \langle \tau_c e^{-(t_1+t_2-2t_0)/\tau_c}\rangle \tau_c \,,$$

where the dependence of v_0 on τ_c has been taken into account. This is the equality of two integrals and, thus, the choices on the integrand functions are infinite. However, the simplest and natural choice is given by the following one:

$$v_{0,eq}^2 = D_V\tau_c \,; \quad \langle v_{\mathrm{eq}}^2\rangle = \langle D_V\rangle\langle\tau_c\rangle \,, \tag{3.105}$$

where the equilibrium assumption on the initial velocity distribution has been taken into account: $\langle v_0^2\rangle_{\mathrm{eq}} = v_{\mathrm{eq}}^2$. This expression corresponds to the alternative definition

of FDT for the present model:

$$\frac{kT}{m} = \langle D_V \rangle \langle \tau_c \rangle \,, \tag{3.106}$$

where m is the mass of the real particle. All the derivation here presented can be performed also without imposing this equilibrium condition but calculation results more clear considering this special case without affecting the final result.

When stationarity is imposed we can write the following expression for the VACF:

$$\begin{aligned} R(\tau) &= \langle V(t_0+\tau)V(t_0)\rangle = \langle D_V \rangle \left\langle \tau_c e^{-\tau/\tau_c} \right\rangle \\ &= \int_0^\infty D_V f(D_V) dD_V \cdot \int_0^\infty \tau_c \, e^{-\tau/\tau_c} \, g(\tau_c) d\tau_c \,. \end{aligned} \tag{3.107}$$

As a consequence of the fundamental relationship between MSD and VACF, given by Eq. (3.19) and by Eqs. (3.20,3.21), we must guess a suitable choice for $R(t)$ in order to get anomalous diffusion as given by Eq. (3.103).

In particular we need that $R(t)$ scales in the long time limit as a powerlaw with negative exponent, $R(t) \sim t^{-\nu}$, with $0 < \nu < 1$, which leads to a fractional scaling of the MSD, $\sigma_x^2(t) \sim t^\phi$, with $1 < \phi < 2$ and $\phi = 2 - \nu$ as in Eq. 3.93. This property can be shown inserting the expression Eq. 3.99 in Eq. 3.107 thanks to the integral representation of the extremal Levy density distribution which characterize the fractional scaling of $b(\tau_c)$

$$L_\alpha^{-\alpha}(x) = \frac{1}{\alpha x} \frac{1}{2\pi i} \int_{\gamma - i\infty}^{\gamma + i\infty} \frac{\Gamma(s/\alpha)}{\Gamma(s)} x^s ds, \quad 0 < \alpha < 1 \,, \tag{3.108}$$

hence we have:

$$\begin{aligned} R(t) &= \langle D_V \rangle \frac{\alpha}{\Gamma(1/\alpha)} \int_0^\infty e^{-t/\tau_c} L_\alpha^{-\alpha}\left(\frac{\tau_c}{\tau_*}\right) d\tau_c \\ &= \langle D_V \rangle \frac{\alpha}{\Gamma(1/\alpha)} \int_0^\infty e^{-t/\tau_c} \left[\frac{1}{\alpha} \frac{1}{2\pi i} \int_{\gamma - i\infty}^{\gamma + i\infty} \frac{\Gamma(s/\alpha)}{\Gamma(s)} \left(\frac{\tau_c}{\tau_*}\right)^{(s-1)} ds \right] d\tau_c \\ &= \langle D_V \rangle \frac{\alpha}{\Gamma(1/\alpha)} \frac{1}{\alpha} \frac{1}{2\pi i} \int_{\gamma - i\infty}^{\gamma + i\infty} \frac{\Gamma(s/\alpha)}{\Gamma(s)} \left[\int_0^\infty e^{-t/\tau_c} \left(\frac{\tau_c}{\tau_*}\right)^{s-1} d\tau_c \right] ds \\ \xi &= t/\tau_c \\ &= \langle D_V \rangle \langle \tau_c \rangle \frac{1}{\alpha} \frac{1}{2\pi i} \int_{\gamma - i\infty}^{\gamma + i\infty} \frac{\Gamma(s/\alpha)}{\Gamma(s)} \left[\int_0^\infty e^{-\xi} \xi^{-1-s} \left(\frac{t}{\tau_*}\right)^s d\xi \right] ds \\ &= \langle D_V \rangle \langle \tau_c \rangle \frac{1}{\alpha} \frac{1}{2\pi i} \int_{\gamma - i\infty}^{\gamma + i\infty} \frac{\Gamma(s/\alpha)\Gamma(-s)}{\Gamma(s)} \left(\frac{t}{\tau_*}\right)^s ds \,, \end{aligned} \tag{3.109}$$

It is useful to rewrite the expression as:

$$\begin{aligned} R(t) &= \langle D_V \rangle \langle \tau_c \rangle \frac{1}{\alpha} \frac{1}{2\pi i} \int_{\gamma - i\infty}^{\gamma + i\infty} \frac{(\alpha/s)\Gamma(s/\alpha+1)\Gamma(-s)}{(1/s)\Gamma(s+1)} \left(\frac{t}{\tau_*}\right)^s ds \\ &= \langle D_V \rangle \langle \tau_c \rangle \frac{1}{2\pi i} \int_{\gamma - i\infty}^{\gamma + i\infty} \frac{\Gamma(s/\alpha+1)\Gamma(-s)}{\Gamma(s+1)} \left(\frac{t}{\tau_*}\right)^s ds \,, \end{aligned} \tag{3.110}$$

which can be solved through the residues theorem considering the poles $s/\alpha + 1 = -n$ or $s = n$, with $n = 0, 1, 2..\infty$.

In the first case it can be written as:

$$\begin{aligned} R(t) =& \langle D_V \rangle \langle \tau_c \rangle \sum_{n=0}^{\infty} \alpha \frac{(-1)^n}{n!} \frac{\Gamma(\alpha(n+1))}{\Gamma(1-\alpha(n+1))} \left(\frac{t}{\tau_*}\right)^{-\alpha(n+1)} \\ =& \langle D_V \rangle \langle \tau_c \rangle \sum_{n=1}^{\infty} \frac{(-1)^n}{n!} \frac{\Gamma(\alpha n)}{\Gamma(-\alpha n)} \left(\frac{t}{\tau_*}\right)^{-\alpha n} , \end{aligned} \tag{3.111}$$

where each term of the serie is obtained by the limit:

$$\begin{aligned} &\lim_{s\to-\alpha(n+1)} (s+\alpha(n+1)) \frac{\Gamma(s/\alpha+1)\Gamma(-s)}{\Gamma(s+1)} \left(\frac{t}{\tau_*}\right)^{s} \\ &\lim_{s\to-\alpha(n+1)} \alpha((s/\alpha+1)+n) \frac{\Gamma(s/\alpha+1)\Gamma(-s)}{\Gamma(s+1)} \left(\frac{t}{\tau_*}\right)^{s} \\ &\lim_{s\to-\alpha(n+1)} \frac{\alpha((s/\alpha+1)+n)}{(s/\alpha+1)_{n+1}} \frac{\Gamma(s/\alpha+n+2)\Gamma(-s)}{\Gamma(s+1)} \left(\frac{t}{\tau_*}\right)^{s} \\ &\lim_{s\to-\alpha(n+1)} \frac{\alpha(-1)^n}{n!} \frac{\Gamma(\alpha(n+1))}{\Gamma(1-\alpha(n+1))} \left(\frac{t}{\tau_*}\right)^{-\alpha(n+1)} , \end{aligned} \tag{3.112}$$

When $t \to \infty$ only the first term survives and we find:

$$R(t) = \langle D_V \rangle \langle \tau_c \rangle \frac{\Gamma(\alpha+1)}{\Gamma(1-\alpha)} \left(\frac{t}{\tau_*}\right)^{-\alpha} , \tag{3.113}$$

which is enough to obtain the desired scaling of the MSD $\sigma_x^2(t) \propto t^{\phi}$, with $\phi = 2-\alpha$.

Analogously, considering the poles in the other semi-plane, $s = n$ with $n = 0, 1, 2..\infty$, it can be find that:

$$R(t) = \langle D_V \rangle \langle \tau_c \rangle \frac{1}{\alpha} \sum_{n=0}^{\infty} \frac{(-1)^n}{n!} \frac{\Gamma(n/\alpha)}{\Gamma(n)} \left(\frac{t}{\tau_*}\right)^{n} , \tag{3.114}$$

converge to $R(0) = \langle D_V \rangle \langle \tau_c \rangle$, as already shown before.

In the special case $\alpha = 1/2$, the extremal Levy function corresponds to the Levy-Smirnov distribution, the whole exercise can be solved analytically and we may consider for simplicity $\langle \tau_c \rangle \frac{\Gamma(1/\alpha)}{\alpha} = 1$:

$$b(\tau_c) = \frac{1}{\sqrt{4\pi\tau_c^5}} e^{-1/(4\tau_c)} . \tag{3.115}$$

Solving the integral the analytical form of the VACF turns to be:

$$R(t) = \frac{\Gamma(1/2)}{\sqrt{4\pi}} \left(t+\frac{1}{4}\right)^{-1/2} , \tag{3.116}$$

which leads to the following MSD:

$$\sigma_x^2(t) = \frac{\Gamma(1/2)}{\sqrt{\pi}} \left[\frac{4}{3}\left(t+\frac{1}{4}\right)^{3/2} - t - \frac{1}{6}\right] , \tag{3.117}$$

that satisfies fractional long time scaling and $\sigma_x^2(0) = 0$ conditions.

In the limit in which $\alpha = 1$ we recover the classical VACF:

$$
\begin{aligned}
R(t) &= \langle D_V \rangle \frac{\alpha}{\Gamma(1/\alpha)} \int_0^\infty e^{-t/\tau_c} L_\alpha^{-\alpha}\left(\frac{\tau_c}{\tau_*}\right) d\tau_c \\
&= \langle D_V \rangle \frac{\alpha}{\Gamma(1/\alpha)} \int_0^\infty e^{-t/\tau_c} \left(\frac{\tau_*}{\tau_c}\right)^{\alpha+1} M_\alpha\left(\frac{\tau_*^\alpha}{\tau_c^\alpha}\right) d\tau_c \\
&= \langle D_V \rangle \int_0^\infty e^{-t/\tau_c} \left(\frac{\tau_*}{\tau_c}\right)^2 \delta\left(\frac{\tau_*}{\tau_c} - 1\right) d\tau_c \\
&= \langle D_V \rangle \int_0^\infty \tau_* e^{-(ty)/\tau_*} y^2 \delta\left(y-1\right) dy \\
&= \langle D_V \rangle \tau_* e^{-t/\tau_*} ,
\end{aligned}
\tag{3.118}
$$

where τ_* is a number and represents the relaxation time of the system.

Subdiffusive case

The most general condition leading to the subdiffusive case in our approach is:

$$
\int_0^\infty R(t) dt = 0 \,. \tag{3.119}
$$

This is satisfied if the VACF shows an oscillating behavior. This feature cannot be obtained in the free particle case because the exponential suppression and the distribution function of the timescale present in the VACF are both positive in the whole support. Then to describe subdiffusive processes the introduction of a further oscillating term is necessary.

To show the potentiality of this approach we consider the Langevin oscillator with random timescale τ_c and velocity diffusivity D_V as a reference case. We maintain the same parameter distributions of the free particle case. we will show that the presence of the harmonic potential, characterized in this case by a non vanishing characteristic frequency of oscillation, is enough to switch the system from a superdiffusive to a subdiffusive regime.

The free particle case must be recovered as in Eq.(3.18) when the frequency of the oscillator vanishes ($\omega_0 = 0$).

We consider time independent stochastic timescale, τ_c, and velocity diffusion coefficient, D_V, distributed in the same way as in the free particle case. The VACF becomes

$$
\begin{aligned}
R(t) &= \langle D_V \rangle \langle \tau_c e^{-\frac{t}{2\tau_c}} \left[cos(\omega_1 t) - \frac{1}{2\omega_1 \tau_c} sin(\omega_1 t) \right] \rangle \\
&= \int_0^\infty D_V h(D_V) dD_V \int_0^\infty \tau_c e^{-\frac{t}{2\tau_c}} \left[cos(\omega_1 t) - \frac{1}{2\omega_1 \tau_c} sin(\omega_1 t) \right] b(\tau_c) d\tau_c \,,
\end{aligned}
\tag{3.120}
$$

The condition leading to subdiffusion is $\int_0^\infty R(t) = 0$, so we have

$$\begin{aligned}\int_0^\infty R(t)dt &= \langle D_V \rangle \int_0^\infty \left\{ \int_0^\infty \tau_c e^{-\frac{t}{2\tau_c}} \left[cos(\omega_1 t) - \frac{1}{2\omega_1 \tau_c} sin(\omega_1 t) \right] b(\tau_c) d\tau_c \right\} dt \\ &= \langle D_V \rangle \int_0^\infty \tau_c \left\{ \int_0^\infty e^{-\frac{t}{2\tau_c}} \left[cos(\omega_1 t) - \frac{1}{2\omega_1 \tau_c} sin(\omega_1 t) \right] dt \right\} b(\tau_c) d\tau_c \\ &= \langle D_V \rangle \int_0^\infty \tau_c F(\tau_c) b(\tau_c) d\tau_c \\ &= \langle D_V \rangle \int_0^\infty 0 \cdot \tau_c b(\tau_c) d\tau_c = 0 \,,\end{aligned} \tag{3.121}$$

since $F(\tau_c) = 0$ for any value of τ_c if $\omega_0 \neq 0$.
This result can be proved solving explicitly the integral $F(\tau_c)$

$$F(\tau_c) = \int_0^\infty e^{-\frac{t}{2\tau_c}} \left[cos(\omega_1 t) - \frac{1}{2\omega_1 \tau_c} sin(\omega_1 t) \right] dt \,, \tag{3.122}$$

the integral can be solved as the sum of cosine and sine Fourier transforms of an exponential function [Bat54], respectively we have

$$\begin{aligned} g_1(y) &= \int_0^\infty f(x) cos(xy) dx \\ g_1(y) &= \frac{a}{a^2 + y^2} \,, \end{aligned} \tag{3.123}$$

and

$$\begin{aligned} g_2(y) &= \int_0^\infty f(x) sin(xy) dx \\ g_2(y) &= \frac{y}{a^2 + y^2} \,, \end{aligned} \tag{3.124}$$

where we consider for our purposes $f(x) = e^{-ax}$, $y = \omega_1 = f(\omega_0 | \tau_c)$, and $a = \frac{1}{2\tau_c}$. $F(\tau_c)$ then results:

$$F(\tau_c) = \frac{\frac{1}{2\tau_c}}{\left(\frac{1}{2\tau_c}\right)^2 + \omega_1^2} - \frac{\frac{1}{2\omega_1 \tau_c} \cdot \omega_1}{\left(\frac{1}{2\tau_c}\right)^2 + \omega_1^2} = 0, \quad \omega_0 \neq 0 \,, \tag{3.125}$$

for any value of τ_c.
This is the first condition to be satisfied in order to obtain fractional subdiffusion. Furthermore, we expect that in the long-time limit $R(t) \sim t^{-\nu}$ with $1 < \nu < 2$ since we have

$$\lim_{t \to \infty} \sigma_x^2(t) = \lim_{s \to 0} s \cdot \mathcal{L}[\sigma_x^2(t)](s) \propto \lim_{s \to 0} \frac{1}{s^2} s \cdot \mathcal{L}[R(t)](s) \,, \tag{3.126}$$

then

$$\lim_{t \to \infty} \sigma_x^2(t) \propto t^\phi = t^{2-\nu}, \quad 0 < \phi < 1, \quad 1 < \nu < 2 \,. \tag{3.127}$$

Similarly to the superdiffusive case, the integral in τ_c in the VACF $R(t)$ can be solved and written as a residues series, from which asymptotic behavior can be derived.

We have

$$R(t) = \langle D_V \rangle \int_0^\infty \tau_c e^{-\frac{t}{2\tau_c}} \left[cos(\omega_1 t) - \frac{1}{2\omega_1 \tau_c} sin(\omega_1 t) \right] b(\tau_c) d\tau_c \,, \tag{3.128}$$

to simplify the integral we first consider its Laplace transform respect to time, and solve it respect τ_c. The longtime limit can be then obtained thanks to the final value theorem of Laplace transformations.

We have

$$\mathcal{L}[e^{at} cos(bt)](s) = \frac{s-a}{(s-a)^2 + b^2} \,, \tag{3.129}$$

and

$$\mathcal{L}[e^{at} sin(bt)](s) = \frac{b}{(s-a)^2 + b^2} \,, \tag{3.130}$$

in both cases with $a = -\frac{1}{2\tau_c}$ and $b = \omega_1 = \sqrt{\omega_0^2 - \frac{1}{4\tau_c^2}}$. Then we have

$$\begin{aligned} \mathcal{L}[R(t)](s) &= \langle D_V \rangle \int_0^\infty \frac{s}{s^2 + s/\tau_c + \omega_0^2} \tau_c b(\tau_c) d\tau_c \\ &= \langle D_V \rangle \int_0^\infty \frac{\tau_c}{1 + \lambda \tau_c} \frac{\alpha}{\Gamma(1/\alpha)} L_\alpha^{-\alpha}(\tau_c/\tau_*) d\tau_c \\ &= \langle D_V \rangle \langle \tau_c \rangle \tau_* \int_0^\infty \frac{\tau_c'}{1 + (\lambda \tau_*)\tau_c'} L_\alpha^{-\alpha}(\tau_c') d\tau_c' \,, \end{aligned} \tag{3.131}$$

with $\lambda = s + \frac{\omega_0^2}{s}$ and $\tau_* = \langle \tau_c \rangle \Gamma(1/\alpha)/\alpha$

To solve the integral over the timescale we consider the following identity, which is a particular result of the method for integral evaluation extensively explained by [Mar84],[2]

[2] The method treated in the book allows solving as a series of residues many definite integrals that involve the product of elementary and special functions depending on a certain number of parameters. It does not need a specific value assignment of the parameters for the integral evaluation, which is instead necessary to numerical evaluation.

Let's consider the integral

$$\int_a^b \phi_1(cx^p, y_1, .., y_m) \phi_2(\gamma x^q, z_1, ..., z_n) dx = A(c, \gamma, y_1, ..., y_m, z_1, ..., z_n) \,, \tag{3.132}$$

with a, b, p, q real numbers and $0 \leq a < b \leq \infty$. By using the substitution $\gamma x^q = t$ and $y c^{q/p} = \gamma$ the integral reduces to the standard form of a Mellin convolution type integral

$$\int_0^\infty \mathcal{H}_1(y/t) \mathcal{H}_2(t) t^{-1} dt = \mathcal{H}(y), \quad y > 0 \,, \tag{3.133}$$

with $\mathcal{H}_1(\tau) = \phi_1(\tau^{-p/q}, y_1, ..., y_m)$, $\mathcal{H}_2(t) = t^{1/q} \phi_2(t, z_1, ..., z_n)$, $\mathcal{H}(y) = q\gamma^{1/q} A(c, \gamma, y_1, ..., y_m, z_1, ..., z_n)$.

Applying the Mellin transformation

$$\mathcal{H}^*(s) = \int_0^\infty \mathcal{H}(y) y^{s-1} dy \,, \tag{3.134}$$

$$\int_0^\infty g(xy)h(y)y^{z-1}dy = \frac{1}{2\pi i}\int_{\mathcal{L}} x^{-q}g^*(q)h^*(z-q)dq\,. \tag{3.138}$$

This expression results equivalent to Eq.(3.131) considering $x = \lambda\tau_*$, $y = \tau_c$, $z = 1$, $g(\tau_c) = \frac{1}{1+\tau_c}$, $h(\tau_c) = \tau_c L_\alpha^{-\alpha}(\tau_c)$, and $g^*(q)$, $h^*(q)$ are the Mellin transforms of the corresponding functions.

So we have

$$\begin{aligned}\mathcal{L}[R(t)](s) &= \langle D_V\rangle\langle\tau_c\rangle\tau_*\frac{1}{2\pi i}\int_{-i\infty}^{+i\infty}(\lambda\tau_*)^{-q}g^*(q)h^*(1-q)dq \\ &= \langle D_V\rangle\langle\tau_c\rangle\tau_*\frac{1}{2\pi i}\int_{-i\infty}^{+i\infty}(\lambda\tau_*)^{-q}\Gamma(q)\Gamma(1-q)\frac{1}{\alpha}\frac{\Gamma((q-1)/\alpha)}{\Gamma(q-1)}dq \\ &= \langle D_V\rangle\langle\tau_c\rangle\tau_*\frac{1}{2\pi i}\int_{-i\infty}^{+i\infty}(\lambda\tau_*)^{-q}\Gamma(1-q)\Gamma\left(\frac{(q-1)}{\alpha}+1\right)dq\,,\end{aligned} \tag{3.139}$$

where we recall that $\lambda = s + \omega_0^2/s$. This integral is then written as residues serie for the poles $q = 1-\alpha(n+1)$, in order to obtain the long time limit behavior of the VACF ($t\to\infty$, $s\to 0$), or for $q = 1+n$, to obtain its short time limit behavior ($t\to 0$, $s\to\infty$).

For the long time limit we obtain:

$$\begin{aligned}\mathcal{L}[R(t)](s) &= \langle D_V\rangle\langle\tau_c\rangle\tau_*\sum_{n=0}^{\infty}\alpha\frac{(-1)^n}{n!}\Gamma(\alpha(n+1))(\lambda\tau_*)^{-1+\alpha(n+1)} \\ &= \langle D_V\rangle\langle\tau_c\rangle\sum_{n=0}^{\infty}\alpha\frac{(-1)^n}{n!}\Gamma(\alpha(n+1))\left(s+\frac{\omega_0^2}{s}\right)^{-1+\alpha(n+1)}\tau_*^{\alpha(n+1)}\,.\end{aligned} \tag{3.140}$$

it can be proven that $\mathcal{H}^*(s) = \mathcal{H}_1^*(s)\mathcal{H}_2^*(s)$. When the functions $\mathcal{H}_1, \mathcal{H}_2$ belong to the class of function of the hypergeometric type their Mellin transform can be written in term of products and ratios of Gamma functions. Then

$$\begin{aligned}\mathcal{H}(y) &= \frac{1}{2\pi i}\int_{\mathcal{L}}\mathcal{H}_1^*(s)\mathcal{H}_2^*(s)y^{-s}ds \\ &= \sum_{k=0}^{m} Res_{s=a_k}\{\mathcal{H}_1^*(s)\mathcal{H}_2^*(s)\}y^{-a_k}\,,\end{aligned} \tag{3.135}$$

A particular case that can be reconducted to this general approach is $\int_0^\infty g(xy)h(y)y^{z-1}dy$

$$\begin{aligned}\int_0^\infty\left\{\int_0^\infty g(xy)h(y)y^{z-1}dy\right\}x^{s-1}dx &= \int_0^\infty\left\{\int_0^\infty g(xy)x^{s-1}dx\right\}h(y)y^{z-1}dy \\ &= \int_0^\infty g(\xi)\xi^{s-1}d\xi\int_0^\infty h(y)y^{z-s-1}dy \\ &= g^*(s)h^*(z-s)\,,\end{aligned} \tag{3.136}$$

inverting the Mellin transformation respect to x we obtain again an expression that could be solved as residues series

$$\int_0^\infty g(xy)h(y)y^{z-1}dy = \frac{1}{2\pi i}\int_{\mathcal{L}} x^{-q}g^*(q)h^*(z-q)dq\,, \tag{3.137}$$

Then we have the asymptotic behaviour:

$$\begin{aligned}\lim_{t\to\infty} R(t) &= \lim_{s\to 0} s\cdot\mathcal{L}[R(t)](s)\\ &= \lim_{s\to 0} s\cdot\langle D_V\rangle\langle\tau_c\rangle\sum_{n=0}^{\infty}\alpha\frac{(-1)^n}{n!}\Gamma(\alpha(n+1))\left(s+\frac{\omega_0^2}{s}\right)^{-1+\alpha(n+1)}\tau_*^{\alpha(n+1)}\\ &= \lim_{s\to 0}\langle D_V\rangle\langle\tau_c\rangle\sum_{n=0}^{\infty}\alpha\frac{(-1)^n}{n!}\Gamma(\alpha(n+1))\left(\omega_0^2\right)^{-1+\alpha(n+1)}s^{2-\alpha(n+1)}\tau_*^{\alpha(n+1)}\\ &= \langle D_V\rangle\langle\tau_c\rangle\alpha\Gamma(\alpha)\left(\omega_0^2\right)^{-1+\alpha}\tau_*^{\alpha}\left(\frac{1}{s}\right)^{\alpha-2}.\end{aligned} \tag{3.141}$$

Then $\lim_{t\to\infty} R(t)\propto t^{-\nu}$, $\nu=2-\alpha$, then the MSD scales as $\sigma_x^2(t)\sim t^{\phi}$, with $\phi=2-(2-\alpha)=\alpha$

Considering the poles $q=n+1$ we obtain the short time behavior, i.e. $R(0)=\langle D_V\rangle\langle\tau_c\rangle$.

$$\mathcal{L}[R(t)](s)=\langle D_V\rangle\langle\tau_c\rangle\tau_*\sum_{n=0}^{\infty}\frac{(-1)^n}{n!}\Gamma\left(\frac{n}{\alpha}+1\right)\left(s+\frac{\omega_0^2}{s}\right)^{-n-1}\tau_*^{-n-1}, \tag{3.142}$$

then we have

$$\begin{aligned}\lim_{t\to 0} R(t) &= \lim_{s\to\infty} s\cdot\mathcal{L}[R(t)](s)\\ &= \lim_{s\to\infty}\langle D_V\rangle\langle\tau_c\rangle\tau_* s\cdot\sum_{n=0}^{\infty}\frac{(-1)^n}{n!}\Gamma\left(\frac{n}{\alpha}+1\right)\left(s+\frac{\omega_0^2}{s}\right)^{-n-1}\tau_*^{-n-1}\\ &= \lim_{s\to\infty}\langle D_V\rangle\langle\tau_c\rangle\sum_{n=0}^{\infty}\frac{(-1)^n}{n!}\Gamma\left(\frac{n}{\alpha}+1\right)(s)^{-n}\tau_*^{-n}\\ &= \langle D_V\rangle\langle\tau_c\rangle,\end{aligned} \tag{3.143}$$

that corresponds to the variance of the velocity.

The shape of the PDF $P(x,t)$ in the long-time limit should be determined by the stochastic process $\xi(t)$, which is a Gaussian white noise, and by the random variable D_V as in the free particle case, with the fundamental difference that the tails of the distributions scale with a power law of time with exponent smaller than one, leading to a subdiffusive regime.

3.5.2 Population of velocity diffusion coefficient D_V and deviations from Gaussian distribution

We proposed an analytical expression for the VACF of the particle velocity $V(t)$ and we derived the variance of the diffusive variable $X(t)$.

Until the velocity coefficient of diffusion D_V is not random the resulting PDF is still a Gaussian density $G(x,\sigma_x^2(t))$, where the variance $\sigma_x^2(t)$ is the one derived from the VACF with anomalous time scaling, due to the randomness of the time scale τ_c (or the friction γ).

Including random velocity diffusivity D_V we may derive not Gaussian PDF.

We remind that the Gaussian fractional stochastic process is described by the following equation:

$$\begin{aligned} dX_\phi(t) &= V_\phi(t)dt \\ dV_\phi(t) &= -\frac{1}{\tau_c}V_\phi(t)dt - \omega_0^2 X_\phi(t) + \sqrt{2D_{v_\phi}}dW(dt), \quad \omega_0 \geq 0\,, \end{aligned} \tag{3.144}$$

Where the label ϕ is introduced to distinguish this process from the classical one and it represents the anomalous time scaling exponent. τ_c is the stochastic timescale of the process, distributed according to $b(\tau_c)$. D_{v_ϕ} is a constant non-random value that we will see it is proportional to the mean value of the random parameter D_V. The resulting PDF in the long time limit is:

$$P(x,t;\phi) = \frac{1}{\sqrt{4\pi C t^\phi}} e^{-\frac{x^2}{4Ct^\phi}}\,, \tag{3.145}$$

with $C = \frac{\Gamma(\alpha+1)}{\Gamma(3-\alpha)}\left(\frac{\Gamma(1/\alpha)}{\alpha}\right)^{(2-\phi)}\langle\tau_c\rangle^{(3-\phi)}D_{v_\phi}$ for the free particle case ($\omega_0 = 0$), while $C = \langle D_V\rangle\langle\tau_c\rangle^{\alpha+1}\left(\omega_0^2\right)^{\alpha-1}\alpha\Gamma(\alpha)\left(\frac{\Gamma(1/\alpha)}{\alpha}\right)^{\alpha}$ when $\omega_0 > 0$.

Let's consider a new stochastic variable which is the product of the Gaussian variable X_ϕ and an appropriate power of a random variable Λ not dependent on time:

$$X = \sqrt{\Lambda}X_\phi\,, \tag{3.146}$$

velocity changes as well:

$$V = \sqrt{\Lambda}V_\phi\,, \tag{3.147}$$

leading to the following stochastic process:

$$\begin{aligned} d\sqrt{\Lambda}X_\phi(t) &= \sqrt{\Lambda}V_\phi(t)dt \\ d\sqrt{\Lambda}V_\phi(t) &= -\frac{1}{\tau_c}\sqrt{\Lambda}V_\phi(t)dt - \omega_0^2\sqrt{\Lambda}X_\phi(t) + \sqrt{\Lambda}\sqrt{2D_{v_\phi}}dW(dt)\,, \end{aligned} \tag{3.148}$$

and we have

$$\begin{aligned} dX(t) &= V(t)dt \\ dV(t) &= -\frac{1}{\tau_c}V(t)dt - \omega_0^2 X(t) + \sqrt{2\Lambda D_{v_\phi}}dW(dt)\,, \end{aligned} \tag{3.149}$$

and $D_V = \Lambda D_{v_\phi}$.

Since the PDF $h(D_V)$ does not depend on time and the stochastic variable D_V is independent of τ_c, for each realization of D_V the Gaussian distribution is recovered and the diffusion coefficient D is proportional to the particular realization of D_V.

From the Lemma 3.1 in [PP16] we have that the PDF associated to a product of two independent random variables $Z = \lambda^\rho Z_1$ is:

$$p(z) = \int_0^\infty p_1(z/\lambda^\rho)p_\lambda(\lambda)\frac{d\lambda}{\lambda^\rho}\,. \tag{3.150}$$

Then if we consider $Z = Xt^{-\phi/2}$, $Z_1 = X_\phi t^{-\phi/2}$ and $\lambda^\rho = \sqrt{\Lambda}$ that is equivalent to $X = \sqrt{\Lambda} X_\phi$, with $\Lambda = D_V/D_{v_\phi}$ and we can write:

$$P_X(x/t^{\phi/2}) = \int_0^\infty P_{X_\phi}(x/(t^\phi D_V/D_{v_\phi})^{1/2}) h(D_V) \frac{dD_V}{(D_V/D_{v_\phi})^{1/2}}\,, \tag{3.151}$$

where $h(D_V) = p_2(D_V/D_{v_\phi})/D_{v_\phi}$.
Multiply both sides by $\frac{1}{t^{\phi/2}}$.

$$\frac{1}{t^{\phi/2}} P_X(x/t^{\phi/2}) = \int_0^\infty \frac{1}{(t^\phi D_V/D_{v_\phi})^{1/2}} P_{X_\phi}(x/(t^\phi D_V/D_{v_\phi})^{1/2}) h(D_V) dD_V \tag{3.152}$$

Let's recall the Gaussian distribution of X_ϕ in the long time limit:

$$\begin{aligned} P_{X_\phi}(x,t) = \frac{1}{\sqrt{t^\phi}} P_{X_\phi}(x/\sqrt{t^\phi}) =& \frac{1}{\sqrt{4\pi C \cdot D_{v_\phi} t^\phi}} e^{-\frac{x^2}{4C \cdot D_{v_\phi} t^\phi}} \\ =& \frac{1}{\sqrt{(4C \cdot D_{v_\phi} t^\phi)}} M_{1/2}\left(\frac{|x|}{\sqrt{(C \cdot D_{v_\phi} t^\phi)}}\right) \\ =& G_\phi(x, \sigma_x^2(t, D_{v_\phi})) \end{aligned} \tag{3.153}$$

Let's recall that P_X, P_{X_ϕ} are self-similar distribution. Then, we have the following general expression:

$$\begin{aligned} P(x,t) &= \int_0^\infty \frac{1}{\sqrt{(4C \cdot D_V t^\phi)}} M_{1/2}\left(\frac{|x|}{\sqrt{(C \cdot D_V t^\phi)}}\right) h(D_V) dD_V \\ &= \int_0^\infty G_\phi(x, \sigma_x^2(t, D_V)) h(D_V) dD_V\,, \end{aligned} \tag{3.154}$$

which means that the final probability is the conditional probability respect to the realization of the random variable D_V weighted over the distribution $h(D_V)$.

Assuming different distributions of the velocity diffusivity $h(D_V)$ is it possible to generate different PDFs as the Mainardi or Levy, which are related to the most known fractional processes in the literature.

In particular, if $p(\Lambda) = M_\beta(\Lambda)$, i.e. the special Wright function known as Mainardi function, we have $h(D_V) = M_\beta(D_V/D_{v_\phi})/D_{v_\phi}$, so that for $\beta = 1$ we obtain $h(D_V) = \delta(D_V - D_{v_\phi})$.

The mean diffusivity of velocity can be derived using the formula ([MLP01]):

$$\int_0^\infty r^\delta M_\nu(r) dr = \frac{\Gamma(\delta+1)}{\Gamma(\nu\delta+1)}, \quad \delta > -1 \tag{3.155}$$

$$\langle D_V \rangle = \int_0^\infty D_V \frac{M_\beta(D_V/D_{v_\phi})}{D_{v_\phi}} dD_V = D_{v_\phi} \frac{\Gamma(2)}{\Gamma(\beta+1)} \tag{3.156}$$

Analogously we have that $\langle D_V^2 \rangle = D_{v_\phi}^2 \frac{\Gamma(3)}{\Gamma(2\beta+1)}$. We may generalize this formula as:

$$\langle D_V^n \rangle = \left(\langle D_V \rangle \frac{\Gamma(\beta+1)}{\Gamma(2)} \right)^n \frac{\Gamma(n+1)}{\Gamma(n\beta+1)} \,. \tag{3.157}$$

The resulting PDF is the Mainardi space time distribution with a time stretching:

$$\begin{aligned} P_X(x,t) =& \int_0^\infty \frac{1}{2(C \cdot D_V t^\phi)^{1/2}} M_{1/2}(|x|/(C \cdot D_V t^\phi)^{1/2}) M_\beta(D_V/D_{v_\phi})/D_{v_\phi} dD_V \\ =& \frac{1}{2(C \cdot t^\phi)^{1/2}} M_{\beta/2} \left(\frac{|x|}{(C \cdot t^\phi)^{1/2}} \right) , \end{aligned} \tag{3.158}$$

If $\beta = \phi$ this is also the solution of the time fractional diffusion equation, obtained substituting the first-order time derivative in the standard diffusion equation with a Caputo derivative of order $\beta \in (0, 2]$. In this sense, we may cover only the range $\beta \in (0, 1]$, for which the PDF presents only a single peak, symmetric respect to X.

Considering $p(\Lambda) = L_{\alpha/2}^{-\alpha/2}(\Lambda)$, i.e. $h(D_V) = L_{\alpha/2}^{-\alpha/2}(D_V/D_{v_\phi})/D_{v_\phi}$, extremal Levy distribution, we obtain that the final PDF is the symmetrical Levy distribution:

$$\begin{aligned} P_X(x,t) =& \int_0^\infty \frac{1}{2(C \cdot D_V t^\phi)^{1/2}} M_{1/2}(|x|/(C \cdot D_V t^\phi)^{1/2}) L_{\alpha/2}^{-\alpha/2}(D_V/D_{v_\phi})/D_{v_\phi} dD_V \\ =& \frac{1}{2(C \cdot D_{v_\phi} t^\phi)^{1/2}} L_\alpha^0 \left(\frac{x}{(C \cdot D_{v_\phi} t^\phi)^{1/2}} \right) , \end{aligned} \tag{3.159}$$

In this last case, we notice that the diffusivity of velocity does not have a finite mean value, which means that the energy necessary to the system to generate such a process should be infinite. This is not realistic in nature but mathematically it is consistent with the fact that the final distribution $P(x,t)$ is a Levy density and has an infinite MSD.

Introducing a cut-off on $h(D_V)$, the mean value should become finite as well as the MSD in the final distribution and the energy involved in the process. However, the analyticity of the solution is lost and the final distribution can only be computed numerically. We expect this numerical solution to be similar, at least in the long-time limit, to the distribution related to Levy Flights. This part of the research goes beyond the scope of the present work and is being developed by a O. Sliusarenko, postdoc at BCAM. Preliminar results confirm our expectation and have been presented in [Sli+17].

3.6 Simulations of the model

Special effort is required in the simulation of the model because of the presence of the extremal Lévy distribution, $\alpha-$stable distributions, involved in the distribution of both the timescale τ and the random scale Λ. In fact also the Mainardi-Wright distribution can be written in terms of a extremal Levy distribution, $M_\beta(x) = \left[L_\beta^{-\beta}(x) \right]^{-\beta}$.

Generation of extremal Lévy distributed random numbers has been performed following the method of Chambers and Mallow [CMS76; JW94]:

$$L^{EXT}_{\alpha,CM} = \frac{\sin[\alpha(r_1+\pi/2)]}{(\cos r_1)^{1/\alpha}}\left\{\frac{\cos[r_1-\alpha(r_1+\pi/2)]}{-\ln r_2}\right\} \quad , \quad 0<\alpha<1\,, \tag{3.160}$$

where r_1 and r_2 are random variables uniformly distributed in $(-\pi/2,\pi/2)$ and $(0,1)$ respectively. Extreme ranges of the distribution can be described replacing the distorted samples by the asymptotic values. To ensure that semi-analytical representation is normalized, two normalization parameters $k_1(\alpha)$ and $k_2(\alpha)$ are introduced to have:

$$\int_0^\infty L^{EXT}_\alpha(x)dx = \int_0^{x_n} k_1(\alpha)x'^{-a}e^{-bx'^{-c}}dx' + \sum_{i=n}^{N} L^{EXT}_{\alpha,CM}(x_i)dx_i + \int_{x_N}^\infty \frac{k_2(\alpha)}{x'^{1+\alpha}}dx' = 1\,, \tag{3.161}$$

where $L^{EXT}_\alpha(x)$ is the constructed distribution, the second term on the right-hand is the histogram obtained by the Chambers-Mellow method and the first and last term are the asymptotic form respectively for $x\to 0^+$ and $x\to\infty$, with

$$a=\frac{2-\alpha}{2(1-\alpha)};\quad b=(1-\alpha)\alpha^{\alpha/(1-\alpha)};\quad c=\frac{\alpha}{1-\alpha}\,.$$

Normalization constant is needed only for the analytical forms because the numerical one is proved to be already normalized by Chambers[CMS76]. This expression is used to obtain a semi-analytical cumulative distribution and to generate the desired random variables by means of the inverse transformation method.

Simulations show that transient anomalous regime is obtained for both the super and subdiffusive case. This regime is asymptotic in the model construction, however in simulation becomes transient because of the finiteness of the maximum value of the random timescale extracted. The anomalous regime switch to normal diffusion regime for $t >> \tau_{max}$. The transient regime start as expected for $t > \langle\tau\rangle$ and variance of velocity reach asymptotically the expected value $\sigma_v = D_V\langle\tau\rangle$, as can be noted in Fig. 3.2.

3.7 Discussion

In this Chapter, we have shown a framework based on a Langevin approach in which fractional kinetics and anomalous diffusion can emerge from Gaussian-based stochastic processes. This corresponds in practice to a proper random modulation of the parameters in the Langevin equation.

We studied the center of mass motion of a heterogeneous ensemble of particles and the motion of a test particle coupled to this heterogeneous surround, demonstrating that it may display anomalous diffusion in an intermediate transient regime. We also show that its dynamics can be described by the superposition of OU processes, characterizing the ensemble. This corresponds more in general to consider a Langevin equation with properly distributed random timescale and velocity diffusivity parameters. We determined a suitable distribution of timescales, which can be attributed to mass density variability in the ensemble, of our extended Langevin process, and demonstrated that it allows obtaining a fractional kinetics with Gaussian PDF.

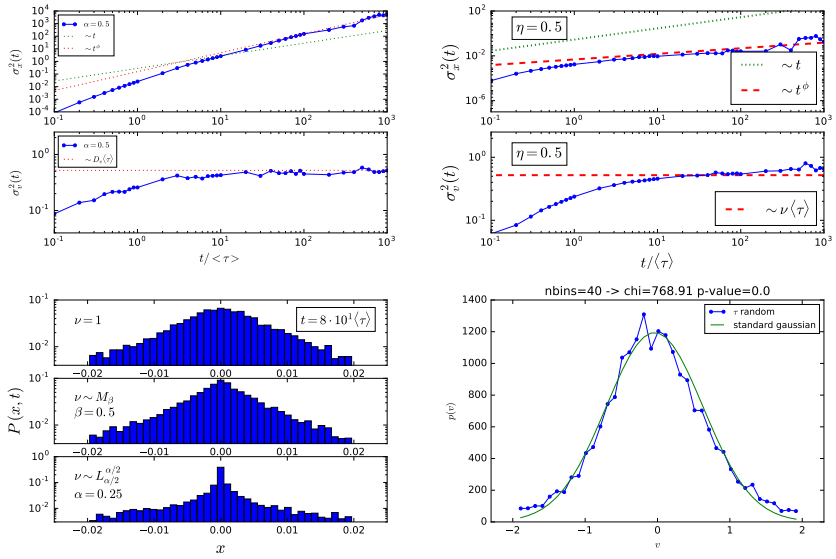

FIGURE 3.2: . Results variance behavior of simulations with constant Λ of superdiffusion (upper left), subdiffusion (upper right), comparison with a Gaussian of the PDF (lower right), results of simulated PDFs with $h(\Lambda)$ (lower left). Figures from [Spo16].

Considering a free particle in a viscous medium it is possible to describe anomalous super-diffusive regimes while introducing a confining term in the equation but maintaining the same parameter modulation, we can describe subdiffusive regimes. To this aim in the present work we considered the Langevin oscillator, that is obtained adding the Hook's restoring force to the free particle Langevin equation. This result is interesting because the Langevin harmonic oscillator converges to the free particle case in the limit of vanishing confining potential. This is an important result because the subdiffusive regime, generally associated with trapping phenomena, arises naturally by the introduction of a macroscopic confining potential and it is not related to changes in the mathematical description of the problem. Within our approach, finally, until the Langevin equation is linear we are able to easily recover non-Gaussian PDFs, in particular, the Mainardi and the Levy densities associated respectively to time and space-fractional diffusion, introducing a multiplicative random length scale in the process. This length scale corresponds to the velocity diffusion coefficient and can be related to the mass distribution of the ensemble.

Since the present approach may appear similar to the so-called superstatistics, i.e., the "statistics of statistics" [BC03], it is useful to dedicate a few words comparing these approaches. The superstatistics approach takes into account large-deviations of intensive quantities of systems in nonequilibrium stationary states [Bec01; BC03; ABC07] and it was motivated by some preliminary success obtained when fluctuations of parameters were considered [WW00; Bec02]. In general, superstatistics is successful to model turbulent dispersion considering energy dissipation fluctuations [Bec01; Rey03], renewal critical events in intermittent systems [PCG09], and for different distributions of the fluctuating intensive quantities different effective statistical mechanics can be derived [BC03], e.g., Tsallis statistics with χ^2-distribution.

The main idea of superstatistics is that a Brownian test particle experiences such fluctuations by moving from cell to cell [BC03]. Following this idea, the random value of the fluctuating parameter is generated at any change of cell. The main assumption behind this picture is that each cell is temporally in equilibrium, within

the cell, there are no fluctuations but different value assigned to each cell. The local value of the fluctuating parameter changes in the various cells on a long timescale that is much larger than the relaxation time that the single cells need to reach local equilibrium. This means that the fluctuating parameter follows a slow dynamics and then the integration over the fast variable is taken after that over the slow variable which is in opposition to what an adiabatic scheme requires [Abe14]. The order of integration could not affect the computation of the expected values but its role can be significant when entropy is considered [ABC07; Abe14]. This inconsistency is solved by considering a dynamical equation also for the slow fluctuating quantity [Abe14], such dynamical equation was already considered in [Rey03].

The present approach is clearly based on a different picture, even if the superposition of Langevin equations may suggest some analogies. Here the superposition generates the single particle trajectory because reproduces the effects of the ensemble heterogeneity. In fact, in the present approach, the fluctuations are not due to different values in different cells but to the distribution in the density of the particles of the surround related to $b(\tau)$. As a consequence of this, the present approach does not take into account slow and fast dynamics and then the issue concerning the order of integration does not arise.

Chapter 4

Anomalous diffusion in spiny dendrites

Time fractional extensions of the cable equation have been proposed in the literature to describe transmembrane potential in spiny dendrites concurrently with anomalous diffusion models for the ions after anomalous diffusion of tracers in this particular systems was observed. This anomalous behavior was directly related to the geometrical nature of the system, in particular the density of spines, by experiments and computer simulations. It was also described analytically in comb-like models. Following this idea in this Chapter we propose two stochastic processes that can be justified by the random geometry of the system and we derive the associated extension of the cable equation, which has a Caputo time fractional derivative instead of the first order time derivative. The fundamental solutions of the most common boundary problems are obtained by the application of the Efros theorem of Laplace transforms, and written in terms of Wright special functions. In literature, similar generalizations have been solved in terms of H-Fox functions. However, the present derivation allows writing explicitly the solution in terms of entire functions and in particular as an integral subordination form containing the solution of the classical cable equation. This is a direct consequence of the application of the Efros theorem, which can be interpreted as a generalization of convolution to subordination relations.

4.1 Introduction

Neurons are the fundamental structural units of the nervous system. These cells are specialized to communicate each other through electrical and chemical signals, specifically called neural signals. We refer for more extensive introductions to neurophysiology to the textbooks [DS94],[Wei96],[Tuc88] and briefly collect here the fundamental concepts devoted to the biological understanding of the problem.

Despite the incredible diversity existing between different neurons type, the basic mechanism to exchange electrical signals is the same as for other excitable cells and is driven by transmembrane ion currents, generating a variation in the transmembrane voltage. Depending on the properties of their membrane two kind of cells (or part of cells) can be distinguished: isopotential, which means that the transmembrane electric potential is the same overall the membrane, or non-isopotential, where this potential could be different in different points of the surface, as in the case of the axons and dendrites, the "electric wires" of the neurons. The principal ions involved are sodium (Na^+), potassium (K^+), calcium (Ca^{2+}) and chloride (Cl^-). Ions motion can be generated by a concentration gradient or by an electrical field.

The flux produced by diffusion in a chemical gradient is classically described by the Ficks law:

$$\overrightarrow{J}_{diff} = -D\overrightarrow{\nabla}[C]\,, \tag{4.1}$$

where D is the coefficient of diffusion and $[C]$ the concentration of the ionic specie that is diffusing.

Drift into an electric field is described by the Ohm law:

$$\overrightarrow{J}_{drift} = -\mu Z[C]\overrightarrow{\nabla}V\,, \tag{4.2}$$

where μ is called mobility and Z is the valence of the ion specie.

The frictional resistance exerted by the fluid medium is independent of the process driving the motion. The flux is an additive quantity, (4.1) and (4.2) can be then related by:

$$D = \frac{K_B T}{q}\mu\,. \tag{4.3}$$

Despite the presence of charged ions inside the cell all the parts of the cell are neutral except the membranes, through which a resting potential different from zero is maintained by an inside-outside chemical gradient. The maintenance of this gradient is fundamental for the correct working of both iso-potential and non-isopotential cells, but it is not trivial. It consists in a balance of passive and active mechanisms. Cell membranes are characterized by selective permeability to different ions going in and out following their concentration gradient, but ions can be also pumped against their chemical gradient by active transporters (specific proteins) that burn ATP (the fuel of the cell).

Let's consider the total flux:

$$J = J_{diff} + J_{drift} = -\mu\frac{K_B T}{q}\frac{\partial[C]}{\partial x} - \mu Z[C]\frac{\partial V}{\partial x}\,, \tag{4.4}$$

we may consider the molar flux $\mathcal{J} = J/N_A$

$$\mathcal{J} = \mathcal{J}_{diff} + \mathcal{J}_{drift} = -u\frac{RT}{F}\frac{\partial[C]}{\partial x} - uZ[C]\frac{\partial V}{\partial x}\,, \tag{4.5}$$

where $u = \mu/N_A$ and $F = qN_A$.

The current results to be the flux multiplied by the charge of the ion:

$$I = J\cdot(Zq) = -\mu Z k_B T\frac{\partial[C]}{\partial x} - \mu Z^2 q[C]\frac{\partial V}{\partial x}\,, \tag{4.6}$$

equivalently we may also consider the current I as the molar flux $\mathcal{J} = J/N_A$ multiplied by the total molar charge $ZqN_A = ZF$.

When $I = 0$ we have the following expression for the transmembrane potential:

$$\frac{\partial V}{\partial x} = -\frac{k_B T}{Zq}\frac{1}{[C]}\frac{\partial[C]}{\partial x}\,. \tag{4.7}$$

Integrating along the membrane thickness we obtain:

$$V_0 = V_{in} - V_{out} = \frac{RT}{ZF} ln\left(\frac{[C]_{out}}{[C]_{in}}\right)\,, \tag{4.8}$$

where $u = \mu/N_A$, $F = qN_A$ and V_0 represents the transmembrane potential at rest,

which is different from zero until a proper gradient of ions concentration through the membrane is maintained.

Neurons are non-isopotential cells, this condition is necessary to transmit the electrical signal, in terms of ions diffusion, in the dendrites and axons. They can be geometrically approximated by a cable, i.e. a cylinder in which the length is much larger than its radius, and eventually, transmit the excitation to another cell or to the soma of the cell. Ions diffusion along the longitudinal axes of the cylinder is fundamental in this type of cells, concurrently with the transmembrane potential previously defined.

The flux of the ions along the longitudinal direction, call it x, can be split again in a diffusion term plus a drift generated by a gradient o the potential. Thus we may describe the flux along longitudinal axes by the same $J(x)$ defined in Eq.(4.4), but calculated respect to this new definition of the variable x.

By applying the continuity equation

$$\frac{\partial [C]}{\partial t} = -\frac{\partial J}{\partial x}\,, \tag{4.9}$$

we may derive from the NPE the following diffusion equation for the ions:

$$\frac{\partial [C]}{\partial t} = D\frac{\partial^2 [C]}{\partial x^2} + \frac{\partial}{\partial x}\left(D\frac{ZF}{RT}[C]\frac{\partial V}{\partial x}\right)\,, \tag{4.10}$$

to which is added a loss/gain of particles quantified in terms of a transmembrane current $-\frac{4}{d}J_m$, that accounts for the passive escape of ions through the ionic channel due to the presence of a transmembrane potential higher then the prescribed one at rest.

The potential is defined as [Tuc88]:

$$V_m(x,t) = V_0 + \frac{Fd}{4c_m}\sum_i \left([C_i](x,t) - [C_i]_0(x,t)\right)\,. \tag{4.11}$$

It is assumed that axial concentration gradient is negligible while the potential variation is not due to the coefficient $\frac{Fd}{4c_m}$, and Eq.4.10 for potential reduces to:

$$c_m\frac{\partial V_m}{\partial t} = -D\left(\frac{d}{4r_L}\frac{\partial^2 V_m}{\partial x^2}\right) - i_m\,. \tag{4.12}$$

The ionic transmembrane current term i_m can be written in term of the Ohm law:

$$i_m = \frac{V_m - V_0}{r_m}\,. \tag{4.13}$$

This phenomenon is approximated by the electrical circuit as in Fig.4.1, consisting of a capacitor and a resistance in parallel. The cellular membrane, constituted by two lipid layers that isolate the internal region from the rest, is modeled by the capacitor. Inserted within the lipidic barrier, the specific proteins that act as channels permitting ions permeation are modeled by the resistance. The transmembrane current is then divided into capacitive and ionic current:

$$I_m = I_C + I_{ionic} = C_m\frac{dV_m}{dt} + \frac{V_m}{R_m}\,, \tag{4.14}$$

where

$$V_m = V - V_0 \,. \tag{4.15}$$

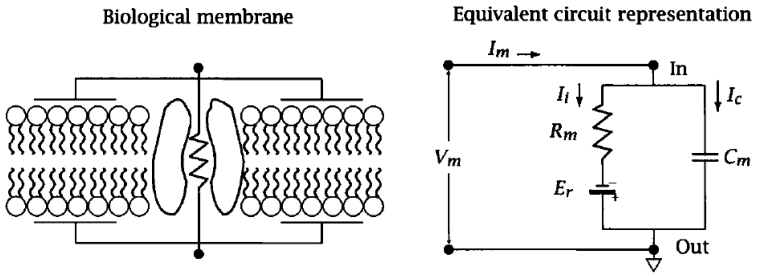

FIGURE 4.1: Figures taken from D. Johnston, S. Miao-Sin Wu. *Foundations of cellular neurophysiology*. MIT Press (1995). (Fig. 3.1).

Dendrites and axons are non-isopotential parts of the neuron in which the membrane can be geometrically approximated by a cylinder. The circuit structure of a cylindrical membrane is shown in Fig.4.2. The cell membrane is characterized by a linear density of RC modules, crossing the membrane, connected by an internal (and eventually external) resistance describing the ionic flow parallel to the membrane in the cytoplasm viscous medium inside the cell (or outside). External axial resistance could be eventually included. Because of this, we have a potential loss of the membrane from a site to another along the cylinder. In general, the external resistance is taken equal to zero ($r_0 = 0$). So we have:

$$\frac{\partial V_m(x,t)}{\partial x} = -r_i i_i \,. \tag{4.16}$$

Part of the current flowing in the cytoplasm may cross the membrane in the RC modules:

$$\frac{\partial i_i}{\partial x} = -i_m \,. \tag{4.17}$$

Combining (4.16) and (4.17) with (4.14) we obtain the cable model:

$$\frac{1}{r_i}\frac{\partial V_m(x,t)}{\partial x^2} = C_m \frac{\partial V_m(x,t)}{\partial t} + \frac{V_m(x,t)}{R_m} \,, \tag{4.18}$$

where coefficients correspond to the axial internal resistance r_i, the transmembrane capacitance c_m and the transmembrane resistance r_m.

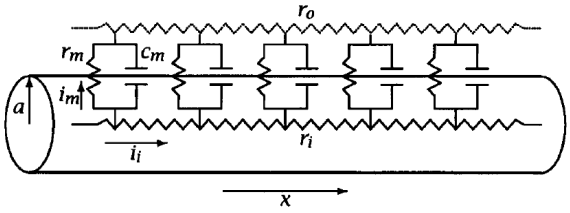

FIGURE 4.2: Figure taken from D. Johnston, S. Miao-Sin Wu. *Foundations of cellular neurophysiology*. MIT Press (1995). (Fig. 4.6.).

The resulting differential equation for the transmembrane potential takes the form of a standard diffusion equation with an extra term that accounts the decay

of the electric signal in time. For simplicity in the rest of this work we will use the dimensionless scaled variables $X = x/\lambda$ and $T = t/\kappa$, where $\lambda = \sqrt{r_m/r_i}$ and $\kappa = r_m c_m$ are the characteristic space and time scales of the process, determined by the values of the membrane resistance and capacitance per unit length of the system, following [Mag06]:

$$\frac{\partial^2 V_m(X,T)}{\partial X^2} - \frac{\partial V_m(X,T)}{\partial T} - V_m(X,T) = 0\,. \tag{4.19}$$

Interesting quantities to neurophysiology are connected to First kind boundary condition (the Signaling Problem) and Second Kind boundary condition problems, following the nomenclature of [Kev00]. Signaling Problem is interesting to understand how the system evolves when excited at one end with a specific potential profile, Second Kind boundary condition problem is interesting because it can be related to the profile of a current injected across the membrane.

In Signaling Problems the cable is considered of semi-infinite length ($0 \leq X < \infty$), initially quiescent for $T < 0$ and excited for $T \geq 0$ at the accessible end ($X = 0$) with a given input in membrane potential $V_m(0,T) = g(T)$. The solution can be derived via the Laplace Transform (LT) approach:

$$\frac{\partial^2 V_m(X,T)}{\partial X^2} = (s+1)V_m(X,T)\,, \tag{4.20}$$

and the LT of the solution results

$$\widetilde{V}_m(X,s) = g(s)e^{-\sqrt{s+1}X}\,. \tag{4.21}$$

Relevant cases are impulsive input $g(T) = \delta(T)$ and unit step input $g(T) = \theta(T)$ where $\delta(T)$ and $\theta(T)$ denote the Dirac and the Heaviside functions, respectively. The solutions corresponding to these inputs can be obtained by LT inversion [Mag06] and read in our notation:

$$\mathcal{G}_s(X,T) = \frac{X}{\sqrt{4\pi T^3}}\mathrm{e}^{-(\frac{\mathrm{X}^2}{4\mathrm{T}}+\mathrm{T})}\,, \tag{4.22}$$

and

$$\mathcal{H}_s(X,T) = \int_0^T \mathcal{G}_s(X,T')\,dT'\,. \tag{4.23}$$

We refer to $\mathcal{G}_s$ to as the fundamental solution or the Green function for the Signaling Problem of the (linear) cable equation in Eq.(4.19), whereas to $\mathcal{H}_s$ to as the step response. As known, the Green function is used in the time convolution integral to represent the solution corresponding to any given input $g(T)$ as follows

$$V_m(X,T) = \int_0^T g(T-T')\,\mathcal{G}_s(X,T')\,dT'\,. \tag{4.24}$$

The spatial variance associated to this model is known to evolve linearly in time.

If we consider an impulse or a step current injected at some point X the problem is subjected to the following boundary conditions, specifically

$$I = I_0\delta(T) = \frac{-1}{r_i\lambda}\frac{\partial V_m(X,T)}{\partial X}\,, \tag{4.25}$$

or

$$I = I_0\theta(T) = \frac{-1}{r_i\lambda}\frac{\partial V_m(X,T)}{\partial X}\,. \tag{4.26}$$

We consider the adimensional current $I = I_0 r_i \lambda$ and put it to unity for convenience. Applying the impulse in $X = 0$ the LT reduces to

$$\widetilde{V}_m(X,s) = \frac{1}{\sqrt{s+1}}\mathrm{e}^{-\sqrt{\mathrm{s+1}}\mathrm{X}}\,, \tag{4.27}$$

the Green function and the step response function (when a step current is applied in $X = 0$) reads, respectively,

$$\mathcal{G}_m(X,T) = \frac{1}{\sqrt{\pi T}}\mathrm{e}^{-(\frac{\mathrm{X}^2}{4\mathrm{T}}+\mathrm{T})}\,, \tag{4.28}$$

and

$$\mathcal{H}_m(X,T) = \int_0^T \frac{1}{\sqrt{\pi T'}}\mathrm{e}^{-(\frac{\mathrm{X}^2}{4\mathrm{T}'}+\mathrm{T}')}\,\mathrm{dT}'\,, \tag{4.29}$$

The Cauchy problem can be solved too in terms of Laplace transforms, the Green function in this special case is equivalent to Eq.(4.28) normalized to the real x axes and reads:

$$\mathcal{G}_c(X,T) = \frac{1}{\sqrt{4\pi T}}\mathrm{e}^{-(\frac{\mathrm{X}^2}{4\mathrm{T}}+\mathrm{T})}\,, \tag{4.30}$$

The motion of ions along the nerve cells is conditioned by this model, that predicts a mean square displacement of diffusing ions scaling linearly with time. By the way, significant deviations from linear behavior have been measured by experiments [JDS97][NSS02][San+06][Ion+17]. A relevant medical and biological example is the anomalous subdiffusion in neuronal dendritic spines [SM80] [Dua03]. Particularly appropriate systems are spiny Purkinje cell dendrites characterized by both spiny and not spiny branches. Spiny branches are in fact characterized by subdiffusive dynamics, while not spiny branches are not. The spatial variance of a diffusing inert tracer (concentration of) in spiny branches evolves as a sub-linear power law of the time, and the diffusion with smaller values of the power exponent is associated to higher spine density [San+06], as spines behave as a trap for the diffusing molecules.

In [HLW08] anomalous diffusion process at the level of ions diffusion has been included modifying the Nernst-Planck equation.

In these cases, the diffusion coefficient D is no more a constant as in the standard NPE but it becomes a parametrized time-dependent operator with scaling parameter $0 < \alpha \leq 1$. For FBM (I) and CTRW (II) respectively we have:

$$D(\alpha,t)_I = D(\alpha)_I D^*(\alpha,t)_I = D(\alpha)_{II}\alpha t^{\alpha-1}\,, \tag{4.31}$$

and

$$D(\alpha,t)_{II} = D(\alpha)_{II} D^*(\alpha,t)_{II} = D(\alpha)_{II}\frac{\partial^{1-\alpha}}{\partial t^{1-\alpha}}\,, \tag{4.32}$$

where $\frac{\partial^{1-\alpha}}{\partial t^{1-\alpha}}$ is the Riemann-Liouville fractional derivative operator.

Considering no external current the following this approach leads to the following differential equations for the two models:

$$\frac{\partial V_I}{\partial T} = \alpha T^{\alpha-1}\frac{\partial^2 V_I}{\partial X^2} - \mu^2\kappa T^{\kappa-1}V_I\,, \tag{4.33}$$

and

$$\frac{\partial V_{II}}{\partial T} = \frac{\partial^{\alpha-1}}{\partial T^{\alpha-1}} \frac{\partial^2 V_{II}}{\partial X^2} - \mu^2 \frac{\partial^{\kappa-1} V_{II}}{\partial T^{\kappa-1}} \,. \tag{4.34}$$

The introduction of a Caputo time derivative in the equation to model the voltage response in neurons has been instead proposed in [TMS14] to model spiking adaptation for a homogeneous membrane patch, where the space derivatives vanish, named fractional leaky integrate-and-fire model:

$$\frac{\partial^\alpha V(T)}{\partial T^\alpha} = -(V(T) - V_L) + I_{inj} \,, \tag{4.35}$$

where an external injected current I_{inj} is included.

Riemann-Liouville and Caputo fractional derivatives, respectively D^α and D^α_*, are related by the following relation [Mai97]:

$$D^\alpha f(t) = D^\alpha_* f(t) + \sum_{k=0}^{m-1} \frac{t^{k-\alpha}}{\Gamma(k-\alpha+1)} f^{(k)}(0^+) \,, \tag{4.36}$$

in terms of fractional integrals they are defined as:

$$D^\alpha = D^m J^{m-\alpha} \,, \tag{4.37}$$

and

$$D^\alpha_* = J^{m-\alpha} D^m \,. \tag{4.38}$$

4.2 Two stochastic models of anomalous diffusion

Experimental evidence of anomalous diffusion of an inert tracer in spiny branches of Purkinje cells [San+06] suggested that the origin of anomalous diffusion in this system is related to the geometry of the system, and that anomalous diffusion is related to the presence of spines more than to the presence of branches. Furthermore, spine density can change dynamically depending on neuronal activity. The idea of correlation between spines density and the anomalous time scaling exponent of the MSD was suggested also in [HLW08], because it was already known that spine density is an important feature for the correct working of several types of neurons, and low density of spines is associated to aging [JDS97; Dua03], neurological disorders [NSS02] and syndromes [SM80]. In [HLW08] it was observed that subdiffusive behavior enlarges the window of high potential at the soma, despite it lowers the maximum value of the peak. It was suggested that this effect could be in fact helpful to compensate time delay of postsynaptic potentials along dendrites and to reduce their longtime temporal attenuation. Then high spine density should have been related to more enhanced subdiffusive behavior. More recently several other experiments have been performed on Purkinje cells and pyramidal cells [San+11] and the correlation between spine density and anomalous diffusion exponent in these types of neurons was explicitly studied. The MSD was described in terms of an anomalous exponent d_ω by the introduction of a time-dependent diffusion coefficient:

$$\langle X^2(t) \rangle = 2D(t) \cdot t = \Gamma \cdot t^{2/d_w} \,, \tag{4.39}$$

and a linear correlation was found between the parameter d_ω and the measured density of spines in both the types of neurons studied. The geometric impact on

anomalous transport in dendrites was then modeled by exploiting the geometrical similarities between a comb structure and a spiny dendrite in terms of comb-like models of diffusion [IM13; MA13]. In this model was considered that particles may diffuse in both spines, fingers of the comb, and dendrite, the backbone of the comb. Where spines behave as traps for the moving particles, and the average survival time τ inside each spine is determined by its geometry. Markovian process was assumed inside each spine, i.e. exponential distribution of survival time $\Psi_M(t,\tau) = \int_t^\infty \frac{1}{\tau} e^{-t'/\tau} dt' = e^{-t/\tau}$, but the random size and shape of the spines [NSS02] entail that the final process is the sum of many independent Markovian processes averaged over the distribution of the timescale τ:

$$\Psi(t) = \int_0^\infty \Psi_M(t,\tau) f_S(\tau) d\tau \,. \tag{4.40}$$

When $f_S(\tau)$ is a power law $\Psi(t)$ shows a power law behavior as well, and subdiffusive diffusion appears.

The emergence of fractional kinetics in complex media in CTRW was introduced more explicitly as a general concept in [Pag14]. Analogously to the comb-like model presented, in that short note the special case of a survival probability of the Mittag-Leffer type was there derived in terms of a Markovian process with characteristic waiting time properly distributed:

$$\int_0^\infty \Psi_M(t,\tau) f_S(\tau) d\tau = E_\alpha(-t^\alpha) \,, \tag{4.41}$$

where $E_\alpha(\cdot)$ is the one parameter Mittag-Leffer function with $0 < \alpha < 1$ [Mai10]:

$$E_\alpha(z) := \sum_{n=0}^{\infty} \frac{z^n}{\Gamma[\alpha n + 1]} \,, \tag{4.42}$$

and $f_S(\tau) = \frac{1}{\tau^2} K_\alpha(1/\tau)$, with $K_\alpha = K_{\alpha,\alpha}^{-\alpha}$ fundamental solution of the neutral spacetime fractional diffusion equation [MLP01], i.e., when space and time fractional orders of derivation are equal, with extremal asymmetry parameter, i.e., defined on the positive real axes only. Within this approach $f_S(\tau)$ corresponds to the stationary distribution of these timescales. If instead we consider the non stationary case it holds for the general case:

$$\Psi(t) = \int_0^\infty \Psi_M(t,\tau) f(\tau,t) d\tau \,, \tag{4.43}$$

however in the non-stationary case the solution for $f(\tau,t)$ could be no unique given $\Psi(t)$.

In the present case the following identity holds:

$$\Psi(t) = \int_0^\infty \Psi_M(t,\tau) f(\tau,t) d\tau = \int_0^\infty e^{-q} \mathcal{H}(q,t) dq \,, \tag{4.44}$$

then for $\Psi(t) = E_\alpha(-t^\alpha)$ we may write:

$$\mathcal{H}(q,t) = \frac{1}{t^\alpha} M_\alpha(q/t^\alpha) \,, \tag{4.45}$$

or equivalently:

$$f(\tau,t) = \frac{1}{\tau^2} t^{1-\alpha} M_\alpha(t^{1-\alpha}/\tau)\,, \tag{4.46}$$

where $M_\alpha(z) = W_{-\alpha,1-\alpha}(z)$ is the M-Wright function, special case of the Wright function $W_{\lambda,\mu}(z)$ defined by the series [Mai10]:

$$W_{\lambda,\mu}(z) := \sum_{n=0}^{\infty} \frac{z^n}{n!\,\Gamma[\lambda n + \mu]}\,, \quad \lambda > -1,\ \mu \geq 0\,. \tag{4.47}$$

The relation in Eq.(4.45) is a consequence of the Laplace transform relation between the M-Wright and the Mittag-Leffer functions [Mai10]:

$$M_\alpha(r) \div E_\alpha(-s)\,, \quad r \in R^+\,, \tag{4.48}$$

thus:

$$\int_0^\infty e^{-rt^\alpha} M_\alpha(r) dr = E_\alpha(-t^\alpha)\,, \tag{4.49}$$

applying the change of variables $q = rt^\alpha$ we have:

$$\int_0^\infty e^{-q} \frac{1}{t^\alpha} M_\alpha(q/t^\alpha) dq = E_\alpha(-t^\alpha)\,. \tag{4.50}$$

Applying this idea to the most general solution for CTRW [Mon64; SGM04] is it possible to write it in term of a superposition of Markovian components, each characterized by the same jump PDF [Tul16]:

$$P(r,t) = \int_0^\infty P_M(r,t/\tau) f(\tau,t) d\tau = \int_0^\infty P_M(r,q)\mathcal{H}(q,t) dq\,. \tag{4.51}$$

The simplest diffusion process of molecules associated to the transmembrane potential solution of the classic cable equation is $P'_M(r,t) = P_M(r,q)e^{-q}$, with $P_M(r,q) = \frac{1}{\sqrt{4\pi q}} e^{-r^2/4q}$, standard diffusion process, multiplied by the exponential factor e^{-q} that accounts for the loss of particles in the system. Following the same superposition principle after turning on the exponential decay term, the transmembrane potential $P(r,t)$ corresponds to the integral of the solution of the classic cable equation averaged by the same $\mathcal{H}(q,t) = \frac{1}{t^\alpha} M_\alpha(q/t^\alpha)$, by considering that the decay is subjected to the same timescale of the diffusion process:

$$P(r,t) = \int_0^\infty \frac{1}{\sqrt{4\pi q}} e^{-(\frac{X^2}{4q}+q)} \frac{1}{t^\alpha} M_\alpha(\frac{q}{t^\alpha}) dq\,. \tag{4.52}$$

The classic problem was written in term of the adimensional variable $T = t/\tau$, with $\tau = c_m r_m$, and $X = x/\lambda$, with $\lambda = \sqrt{r_i/r_m}$ related to the circuit component of the membrane element. The solution of the fractional process can be written in term of a superposition of the classic solution weighted by the distribution of the circuit element parameters, thus we have:

$$V_\alpha(x,t) = \int_0^\infty \frac{1}{\sqrt{4\pi t/\tau}} e^{-\left(\lambda^2 \frac{x^2}{4t/\tau} + t/\tau\right)} \frac{t^{1-\alpha}}{\tau^2} M_\alpha(t^{1-\alpha}/\tau) d\tau\,. \tag{4.53}$$

Then in terms of circuit elements the system results characterized by a capacitance that varies between the elements of the circuit according to a certain time dependent distribution, considering r_i, r_m unitary constant for simplicity, in the present case it

correspond to:

$$f(c_m, t) = \frac{t^{1-\alpha}}{c_m^2} M_\alpha(t^{1-\alpha}/c_m)\,. \tag{4.54}$$

If there exists also a population of r_m, representing the transmembrane resistance, the time decay of the solution and diffusion processes are described by two different but correlated distributions, because the coefficient r_m disappears in the Gaussian factor.

In terms of comb-like model approach the timescale is the average sojourn time in the spine and can be related to the geometry of the spine, as example for spines with a head of volume V and cylindrical spine neck of length L and radius a, the mean life time is $\tau = (LV)/(\pi a 2D)$, where D is the diffusivity of the spine [IM13; MA13]. If this volume may change dynamically it makes sense to consider a time dependent distribution in this case as well.

Eq.(4.53) can be also interpreted within the ggBm approach [**Mura**]; rewriting the integral form as follows:

$$V_\alpha(X,T) = \int_0^\infty \frac{1}{\sqrt{4\pi\Lambda T^\alpha}} e^{-\left(\frac{X^2}{4\Lambda T^\alpha} + \Lambda T^\alpha\right)} M_\alpha(\Lambda) d\Lambda\,, \tag{4.55}$$

where inside the integral we recognize the fundamental solution of the fBm model defined in Eq.(4.33) for the particular case $\alpha = \kappa$.

The ggBm-like stochastic process can be defined by the product:

$$X'(t) = \sqrt{D} X(t)\,, \tag{4.56}$$

where $X(t)$ is a Gaussian process with unitary coefficient of diffusion, rescaled by the diffusion coefficient D distributed according to:

$$\rho(D,t) = \frac{1}{t^{\alpha-1}} M_\alpha(D/t^{\alpha-1})\,, \tag{4.57}$$

where comes natural the change of variables $D = \Lambda t^{\alpha-1}$, which is the fBm definition of the diffusion coefficient, thus $p(\Lambda) = M_\alpha(\Lambda)$ [DG+16]. The survival probability of each particle is conditioned to its diffusion coefficient D:

$$r(D,t) = e^{-Dt}\,. \tag{4.58}$$

The PDE for these processes can be derived by computing the Laplace-Fourier transform of the integral form in Eq.(4.52), that reads

$$\hat{\tilde{V}}_\alpha(s,k) = \frac{2s^{\alpha-1}}{s^\alpha + 1 + k^2}\,, \tag{4.59}$$

thus the transformed PDE is

$$2s^{\alpha-1} = s^\alpha \hat{\tilde{V}}_\alpha(s,k) + \hat{\tilde{V}}_\alpha(s,k) + k^2 \hat{\tilde{V}}_\alpha(s,k)\,, \tag{4.60}$$

which correspond to the time fractional cable equation described in [VM17] with $0 < \alpha < 1$ for Cauchy initial conditions:

$$\frac{\partial^\alpha V_\alpha(X,T)}{\partial T^\alpha} = \frac{\partial^2 V_\alpha(X,T)}{\partial X^2} - V_\alpha(X,T)\,, \tag{4.61}$$

where $\frac{\partial^\alpha}{\partial T^\alpha}$ is the Caputo time fractional derivative.

4.3 Time fractional cable model

In this Section, we consider the extension of the cable equation expressed in Eq.(4.19), which correspond the substitution of the integer time derivative with a fractional Caputo time derivative. This approach extends the model of [TMS14] to diffusing processes. The solution derived in here can be reconducted to a special case of the model presented in [HLW08], but the approach used to derive it, at least for us, is much simpler and allows to write explicitly the integral form of the solution in terms of entire functions. For this reason, different boundary conditions can be easily derived and computed. As for the model derived by [HLW08], our approach allows to reproduce at least qualitatively the main characteristics observed in experiments [NSS02], [JDS97],[Dua03],[SM80][San+06].

To model anomalous sub-diffusion we substitute the first-order time derivative in Eq.(4.19) with a fractional time derivative of Caputo type [GM97], [Pod99] of order $\alpha \in (0,1)$:

$$\frac{\partial^2 V_m(X,T)}{\partial X^2} - \frac{\partial^\alpha V_m(X,T)}{\partial T^\alpha} - V_m(X,T) = 0\,, \tag{4.62}$$

for $0 < \alpha < 1$.

4.3.1 Solution of the Signaling Problem via Laplace Transform

The solution of the Signaling Problem can be derived via LT [VM17], however, the inversion of the LT solution for Eq. (4.62) requires special efforts because of the term $V_m(X,T)$.

When this term is not present, the resulting equation is the well known time fractional diffusion equation:

$$\frac{\partial^2 V_m^*(X,T)}{\partial X^2} - \frac{\partial^\alpha V_m^*(X,T)}{\partial T^\alpha} = 0\,, \tag{4.63}$$

for which the solutions of the corresponding Cauchy and Signaling Problems have been derived in the 1990's in terms of two auxiliary Wright functions (of the second type) [Mai96; Mai97]. Specifically for the Signaling Problem, the general solution there provided in integral convolution form reads

$$V_m^*(X,T) = \int_0^T g(T-T')\,\mathcal{G}_{\alpha,s}^*(X,T')\,dT'\,, \quad \mathcal{G}_{\alpha,s}^*(X,T) = \frac{1}{T} W_{-\alpha/2,0}\left(-X/T^{\alpha/2}\right)\,, \tag{4.64}$$

where $\mathcal{G}_{\alpha,s}^*(X,T)$ denotes the Green function of the Signaling Problem of the fractional time diffusion equation (Eq.(4.63)) and $W_{-\alpha/2,0}(\cdot)$ is a particular case of the transcendental function known as Wright function defined in Eq.(4.47). This function, entire in the complex plane, is discussed extensively in [Mai10] where the interested reader can find the following relevant LT pairs, rigorously derived in [Sta]:

$$t^{\mu-1}\,W_{-\nu,\mu}\left(-x/t^\nu\right) \;\div\; s^{-\mu}\,\exp\left(-xs^\nu\right)\,, \quad 0 \le \nu < 1\,,\ \mu > 0\,. \tag{4.65}$$

Here we have adopted an obvious notation to denote the juxtaposition of a locally integrable function of time t with its LT in s with x a positive parameter. It is worth to recall the distinction of the Wright functions in first type ($\lambda \ge 0$) and second type ($-1 < \lambda \le 0$) and, among the latter ones, the relevance of the two auxiliary functions

introduced in [Mai96]:

$$F_\nu(z) = W_{-\nu,0}(-z)\,, \quad M_\nu(z) = W_{-\nu,1-\nu}(-z)\,, \quad 0 < \nu < 1\,, \tag{4.66}$$

inter-related as $F_\nu(z) = \nu z M_\nu(z)$. Indeed the relevance of both the Wright functions has been outlined by several authors in diffusion and stochastic processes. Particular attention is due to the M-Wright function (also referred to as the Mainardi function in [Pod99]) that, since for $\nu = 1/2$ reduces to $\exp(-z^2/4)/\sqrt{\pi}$, is considered a suitable generalization of the Gaussian density, see [Pag13] and references therein.

Then the Green function for the Signaling Problem of the time fractional diffusion equation (Eq.(4.63)) can be written in the original form provided in [Mai96] as

$$\mathcal{G}^*_{\alpha,s}(X,T) = \frac{1}{T} F_{\alpha/2}\left(X/T^{\alpha/2}\right) = \frac{\alpha}{2}\,\frac{X}{T^{\alpha/2+1}}\,M_{\alpha/2}\left(X/T^{\alpha/2}\right)\,, \tag{4.67}$$

where the superscript $*$ is added to distinguish the time fractional diffusion equation from our fractional cable equation, both depending on the order $\alpha \in (0,1)$.

Applying the LT to Eq.(4.62) with the boundary conditions required by the Signaling Problem, that is $V_m(X,0^+) = 0$, $V_m(0,T) = g(T)$, we have:

$$(s^\alpha + 1)\widetilde{V_m}(X,s) - \frac{\partial^2 \widetilde{V_m}(X,s)}{\partial X^2} = 0, \tag{4.68}$$

which is a second order equation in the variable X with solution:

$$\widetilde{V_m}(X,s) = \widetilde{g}(s) e^{-\sqrt{(s^\alpha+1)}\cdot X}. \tag{4.69}$$

Because of the shift constant in the square root of the LT in Eq.(4.69), the inversion is no longer straightforward with the Wright functions as it is in the time fractional diffusion equation (Eq.(4.63)). Consequently, we have overcome this difficulty recurring to the application of the Efros theorem [Gra04] that generalizes the well-known convolution theorem for LTs. For sake of convenience let us hereafter recall this theorem, usually not so well-known in the literature. The Efros theorem states that if we can write a LT $\widetilde{f}(s)$ as:

$$\widetilde{f}(s) = \phi(s) \cdot \widetilde{F}(\psi(s)), \tag{4.70}$$

where the function $\widetilde{F}(s)$ has a known inverse LT $F(T)$, the inverse LT can be written in the form:

$$f(T) = \int_0^\infty F(\tau) G(\tau, T) d\tau\,, \tag{4.71}$$

where:

$$G(\tau;T) \div \widetilde{G}(\tau,s) = \phi(s) e^{-\tau\psi(s)}\,. \tag{4.72}$$

In Eq.(4.69), LT solution of our Signaling Problem, we thus have:

$$\phi(s) = \widetilde{g}(s), \quad \psi(s) = s^\alpha\,, \tag{4.73}$$

and

$$\widetilde{F}(s)|_X = \mathrm{e}^{-\mathrm{X}\sqrt{\mathrm{s}+1}}\,. \tag{4.74}$$

Then, having $\widetilde{G}(\tau,s) = \widetilde{g}(s)\,\mathrm{e}^{-\tau \mathrm{s}^\alpha}$, thanks to the standard convolution theorem of LTs, we obtain:

$$G(\tau,T) = \int_0^T \frac{g(T-T')}{T'} W_{-\alpha,0}(-\tau/T'^\alpha) dT'\,, \tag{4.75}$$

where $W_{-\alpha,0}$ is the F-Wright function, and

$$F(T)|_X = \frac{X}{\sqrt{4\pi T^3}} e^{-(\frac{X^2}{4T}+T)}\,, \tag{4.76}$$

is the solution in Eq.(4.22) of the standard cable equation (Eq.(4.19)).

Then, the general solution for the Signaling Problem can be written in terms of known functions:

$$\begin{aligned} V_m(X,T) \quad &= \int_0^\infty \frac{X}{\sqrt{4\pi\tau^3}}\, \mathrm{e}^{-(\frac{\mathrm{X}^2}{4\tau}+\tau)} \left[\int_0^{\mathrm{T}} \frac{\mathrm{g(T-T')}}{\mathrm{T'}} \mathrm{W}_{-\alpha,0}(-\tau/\mathrm{T'}^{\alpha})\mathrm{dT'}\right] \mathrm{d}\tau \\ &= \int_0^T g(T-T') \left[\int_0^\infty \frac{X}{\sqrt{4\pi\tau^3}} \mathrm{e}^{-(\frac{X^2}{4\tau}+\tau)} \frac{1}{T'} F_\alpha(\frac{\tau}{T'^{\alpha}})\, d\tau\right] dT'\,. \end{aligned} \tag{4.77}$$

Substituting $g(T) = \delta(T)$ in the general solution in Eq.(4.77) we obtain the Green function for the fractional model (Eq.(4.62)), shown in Fig.(4.3):

$$\begin{aligned} V_m(X,T) := \mathcal{G}_{\alpha,s}(X,T) \quad &= \int_0^\infty \mathcal{G}_s(X,\tau)\frac{1}{T}F_\alpha(\frac{\tau}{T^\alpha})d\tau \\ &= \int_0^\infty \mathcal{G}_s(X,\tau)\mathcal{G}^*_{2\alpha,s}(\tau,T)d\tau \end{aligned} \tag{4.78}$$

When $g(T) = \theta(T)$ we obtain the step response of our fractional cable equation :

$$V_m(X,T) := \mathcal{H}_{\alpha,s}(X,T) \quad = \int_0^\infty \mathcal{G}_s(X,\tau)\left[\int_0^T \mathcal{G}^*_{2\alpha,s}(\tau,T')dT'\right] d\tau \tag{4.79}$$

$$= \int_0^\infty \mathcal{G}_s(X,\tau)\mathcal{H}^*_{2\alpha,s}(\tau,T)\, d\tau\,, \tag{4.80}$$

where $\mathcal{H}^*_{2\alpha,s}(\tau,T)$ is the step response function for the time fractional diffusion equation. After some manipulations including the change of variable $z = \tau/T'^{\alpha}$ and integrating by parts after using the recurrence relation of Wright functions:

$$\frac{dW_{\lambda,\mu}(z)}{dz} = W_{\lambda,\lambda+\mu}(z)\,, \tag{4.81}$$

and the relation between the auxiliary functions: $F_\nu(z) = \nu z\, M_\nu(z)$ we may rewrite the step-response solution as:

$$V_m(X,T) := \mathcal{H}_{\alpha,s}(X,T) = \int_0^\infty \mathcal{H}_s(X,\tau)\cdot\frac{1}{T^\alpha}M_\alpha(\frac{\tau}{T^\alpha})d\tau\,, \tag{4.82}$$

where $\mathcal{H}_{\alpha,s}(X,T)$ is the step response function for the standard cable model. The same expression can easily be derived by direct application of the Efros theorem and is plotted in Fig.4.4.

4.3.2 The Green function for the Cauchy Problem

Consider an infinite cable with boundary conditions $V_m(\pm\infty,T) = 0$ and initial condition $V_m(X,0) = f(X)$. The general solution of the Cauchy problem is related to the Green function $\mathcal{G}_{\alpha,c}(X,T)$ through the following relation:

$$V_m(X,T) = \int_{-\infty}^{+\infty} f(x-\xi)\mathcal{G}_{\alpha,c}(\xi,T)d\xi\,. \tag{4.83}$$

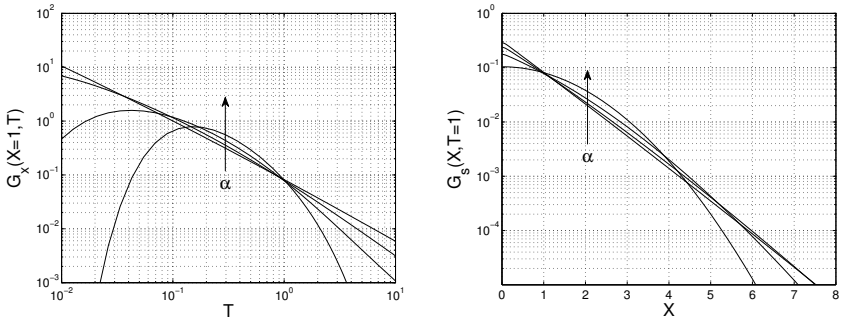

FIGURE 4.3: Green function for Signaling Problem is calculated and plotted for $X = 1$ as function of time T (left panel) and for $T = 1$ as function of X (right panel). Several values of parameter α are compared: $0.25, 0.5, 0.75, 1.$.

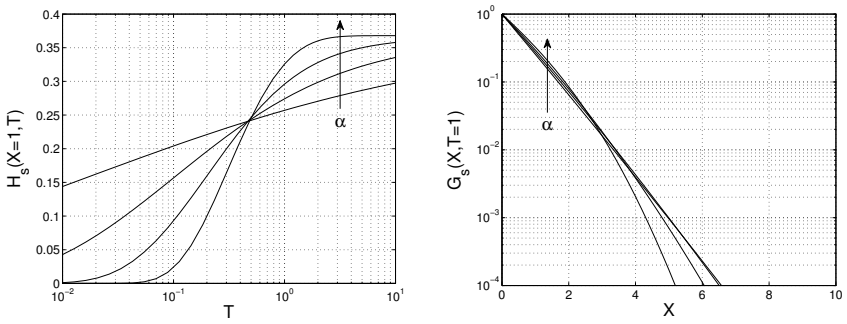

FIGURE 4.4: Step response function for Signaling Problem is calculated and plotted for $X = 1$ as function of time T (left panel) and for $T = 1$ as function of X (right panel). Several values of parameter α are compared: $0.25, 0.5, 0.75, 1.$.

$\mathcal{G}_{\alpha,c}(X,T)$ can be derived via Laplace Transform:

$$(s^\alpha + 1)\widetilde{\mathcal{G}}_{\alpha,c}(X,s) - \frac{\partial^2 \widetilde{\mathcal{G}}_{\alpha,c}}{\partial X^2} = \delta(X)s^{\alpha-1}\,, \tag{4.84}$$

boundary conditions imposes:

$$\widetilde{\mathcal{G}}_{\alpha,c}(X,s) = \begin{cases} c_1(s)e^{-X\sqrt{s^\alpha+1}}, & \text{if } X > 0 \\ c_2(s)e^{+X\sqrt{s^\alpha+1}}, & \text{if } X < 0 \end{cases}. \tag{4.85}$$

Imposing $\widetilde{\mathcal{G}}_{\alpha,c}(0^-,s) = \widetilde{\mathcal{G}}_{\alpha,c}(0^+,s)$ leads to $c_1(s) = c_2(s)$. Integrating Eq.(4.62) over X from 0^- to 0^+ we have:

$$\frac{\partial \widetilde{\mathcal{G}}_{\alpha,c}(0^+,s)}{\partial X} - \frac{\partial \widetilde{\mathcal{G}}_{\alpha,c}(0^-,s)}{\partial X} = -s^{\alpha-1} \tag{4.86}$$

the coefficients result:

$$c_1(s) = c_2(s) = \frac{1}{2s^{1-\alpha}\sqrt{s^\alpha+1}}\,, \tag{4.87}$$

the resulting LT of the Green function reads:

$$\widetilde{\mathcal{G}}_{\alpha,c}(X,s) = \frac{1}{2s^{1-\alpha}\sqrt{s^\alpha+1}} e^{-X\sqrt{s^\alpha+1}}\,. \tag{4.88}$$

The inversion can be easily performed for $X > 0$, thanks again to the Efros theorem, and extended by symmetry respect to the X-axes for $X < 0$.

Let's consider $\phi(s) = \frac{1}{s^{1-\alpha}}$, $\psi(s) = s^{\alpha}$, following the theorem we may set $G(\tau, s) = \frac{1}{s^{1-\alpha}} e^{-\tau s^{\alpha}}$ and $F(X, s) = \frac{1}{2\sqrt{s+1}} e^{-X\sqrt{s+1}}$, that have known inverse LT:

$$F(X,T) = \frac{1}{\sqrt{4\pi T}} e^{-\left(\frac{X^2}{4T} + T\right)}, \tag{4.89}$$

and

$$G(\tau,T) = \frac{1}{T^{\alpha}} W_{-\alpha,1-\alpha}(-\tau/T^{\alpha}) = \frac{1}{T^{\alpha}} M_{\alpha}(\tau/T^{\alpha}). \tag{4.90}$$

The inverse LT for the Green function is plotted in Fig.4.5 and reads:

$$\begin{aligned} \mathcal{G}_{\alpha,c}(X,T) &= \int_0^{\infty} \frac{1}{\sqrt{4\pi\tau}} e^{-\left(\frac{X^2}{4\tau} + \tau\right)} \frac{1}{T^{\alpha}} M_{\alpha}(\tau/T^{\alpha}) d\tau \\ &= \int_0^{\infty} \mathcal{G}_c(X,\tau) \mathcal{G}^*_{2\alpha,c}(\tau,T) d\tau \,. \end{aligned} \tag{4.91}$$

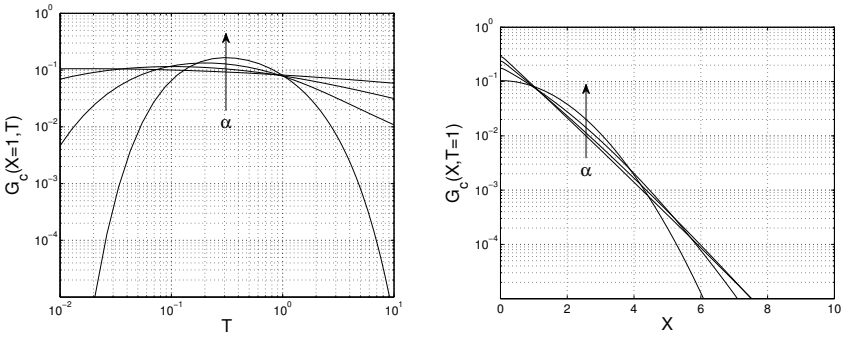

FIGURE 4.5: Green function for Cauchy Problem is calculated and plotted for $X = 1$ as function of time T (left panel) and for $T = 1$ as function of X (right panel). Several values of parameter α are compared: $0.25, 0.5, 0.75, 1$..

4.3.3 Response to injected current

An interesting biological problem is to consider an injected current in the system. Transmembrane potential is related to the transmembrane current through the relation $-I = \frac{\partial^2 V_m(X.T)}{\partial X^2}$, where the minus sign is due to the direction of the current, in this case flowing inside the cell. Let's consider a singular point injected current in $X = 0$, it takes the form $I(X,T) = I_0 \delta(X) f(T)$. Integrating from 0^- to 0^+ we obtain the relation

$$-I_0 f(T) = \frac{\partial V_m(X,T)}{\partial X}\Big|_{X=0^+} - \frac{\partial V_m(X,T)}{\partial X}\Big|_{X=0^-}. \tag{4.92}$$

We recall the LT for the semi-infinite cable for an initially undisturbed cable:

$$\widetilde{V}_m(X,s) = \widetilde{V}_m(0,s) e^{-X\sqrt{s^{\alpha}+1}}. \tag{4.93}$$

At the boundary condition we have:

$$I_0 \widetilde{f}(s) = -\frac{\partial \widetilde{V}_m(X,s)}{\partial X}\Big|_{X=0^+}, \tag{4.94}$$

if we consider an impulse injection of current in $X = 0$ we have $I_0\delta(T) = -\frac{\partial V_m(X,T)}{\partial X}\big|_{X=0^+}$. Applying this condition to the LT we obtain:

$$\widetilde{V}_m(0^+, s) = \frac{I_0}{\sqrt{s^\alpha + 1}}\,, \tag{4.95}$$

leading to the following Laplace Transformed solution:

$$\widetilde{\mathcal{G}}_{\alpha,m}(X, s) = \frac{I_0}{\sqrt{s^\alpha + 1}} e^{-X\sqrt{s^\alpha+1}}\,. \tag{4.96}$$

According to the previous derivations it is then straightforward that the inverse LT takes the form:

$$\begin{aligned}\mathcal{G}_{\alpha,m}(X, T) &= \int_0^\infty \frac{I_0}{\sqrt{\pi\tau}} e^{-\left(\frac{X^2}{4\tau}+\tau\right)} \frac{1}{T} W_{-\alpha,0}(-\tau/T^\alpha) d\tau \\ &= \int_0^\infty \mathcal{G}_m(X, \tau)\mathcal{G}^*_{2\alpha,s}(\tau, T) d\tau\,,\end{aligned} \tag{4.97}$$

represented in Fig.4.6.

For a generic boundary $I_0\widetilde{f}(s)$ we obtain:

$$\widetilde{V}_m(X, s) = \frac{I_0 f(s)}{\sqrt{s^\alpha + 1}} e^{-X\sqrt{s^\alpha+1}}\,. \tag{4.98}$$

The general solution becomes:

$$V_m(X, T) = \int_0^T f(T - T')\mathcal{G}_{\alpha,m}(X, T') dT'\,. \tag{4.99}$$

The solution is symmetric respect to X, the problem can be then extended to the infinite cable introducing a factor $1/2$: $\mathcal{G}^\infty_{\alpha,m}(X, T) = \frac{1}{2}\mathcal{G}_{\alpha,m}(X, T)$.

The extension to the infinite cable case admits also the following generalization, current injection in $X_0 \neq 0$ is equivalent to shift the cable of the same value X_0, then:

$$V^\infty_{X_0,m}(X, T) = \int_0^T f(T - T')\mathcal{G}^\infty_{\alpha,m}(X - X_0, T') dT'\,. \tag{4.100}$$

When the injected current is a step function we obtain the following LT solution:

$$\widetilde{\mathcal{H}}_{\alpha,m}(X, s) = \frac{I_0}{s\sqrt{s^\alpha + 1}} e^{-X\sqrt{s^\alpha+1}} = \frac{I_0}{s^{1-\alpha}s^\alpha\sqrt{s^\alpha + 1}} e^{-X\sqrt{s^\alpha+1}}\,, \tag{4.101}$$

considering $\phi(s) = \frac{1}{s^{1-\alpha}}$, $\psi(s) = s^\alpha$ we have $G(\tau, s) = \frac{1}{s^{1-\alpha}} e^{-\tau s^\alpha}$ and $F(X, s) = \frac{1}{s\sqrt{s+1}} e^{-X\sqrt{s+1}}$, that have known inverse LT Eq.(4.100) can be simplified to:

$$\begin{aligned}\mathcal{H}_{\alpha,m}(X, T) &= \int_0^\infty \mathcal{H}_m(X, \tau) \frac{1}{T^\alpha} M_\alpha(\tau/T^\alpha) d\tau \\ &= \int_0^\infty \mathcal{H}_m(X, \tau)\mathcal{G}^*_{2\alpha,c}(\tau, T) d\tau\,,\end{aligned} \tag{4.102}$$

which is shown in Fig.4.7.

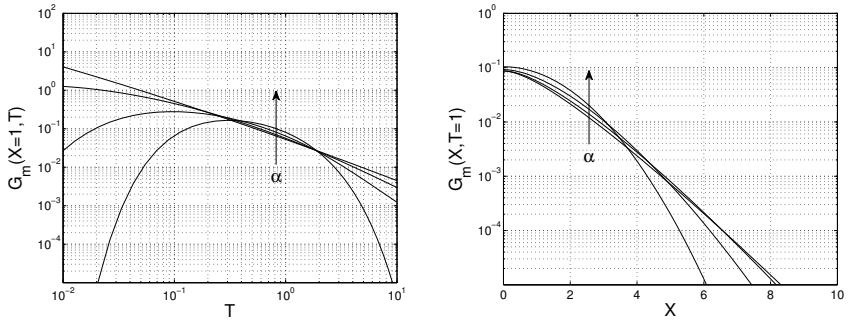

FIGURE 4.6: Green function for Second Kind Boundary Problem is calculated and plotted for $X = 1$ as function of time T (left panel) and for $T = 1$ as function of X (right panel). Several values of parameter α are compared: $0.25, 0.5, 0.75, 1.$.

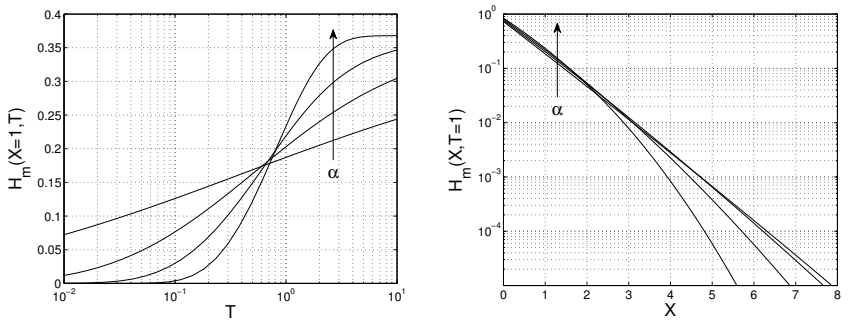

FIGURE 4.7: Step response function for Second Kind Boundary Problem is calculated and plotted for $X = 1$ as function of time T (left panel) and for $T = 1$ as function of X (right panel). Several values of parameter α are compared: $0.25, 0.5, 0.75, 1.$.

4.4 Discussion

Fractional calculus is often used to catch by parsimonious mathematical approach some underlying complex behavior. Caputo's fractional derivative is a non-local operator and for this reason, as pointed out in [TMS14], it could be introduced to explain emergent behaviors like the appearance of multiple timescale dynamics and memory effects, related to the complexity of the medium. In Section 4.2,we derived two possible stochastic processes for inert tracer diffusion in spiny dendrites that in principle give rise to the same partial differential equation for the transmembrane potential. The PDF evolution of both the processes is be described by the time fractional generalization of the cable equation presented in Eq.(4.62), that can be solved for the most common boundary and initial conditions by the application of the Efros theorem of Laplace transforms [VM17; SVM17] as shown in Section 4.3.

The first process is a CTRW built as a superposition of Markovian processes, each one subjected to a different timescale of the waiting time distribution, where the timescale is a non stationary distribution. This means that the whole system change in time modifying the profile of the timescale distribution. The second process is based on a ggBm-like approach in which a Brownian process with unitary diffusion coefficient is rescaled by a random scale that is non stationary distributed as well. This scale represents in fact the value of the diffusion coefficient and can be used to generate anomalous time-scaling of the mean square displacement in the final

variable:

$$\langle X'^2 \rangle = \langle D \rangle \cdot t \,. \tag{4.103}$$

If $r(D,t) = 1$ we obtain $\langle X'^2 \rangle \sim t^\alpha$. The exponential suppression has the effect that probability distribution collapse to zero in the infinite time, because all the particles disappear from the system.

Except for the exponential suppression that accounts for loss of particles, a similar non-stationary ggBm process has been proposed in [DG+16] as an alternative to CTRW to account the Ergodicity Breaking (EB) described by several experiments on diffusion of cellular components in living systems. Despite both ggBm and CTRW may account for EB, in [DG+16] it was shown that the p-variation test provides different values for the two alternative processes, and that values obtained for ggBm where compatible with the experimental dataset considered in their research, on the contrary of CTRW.

For this reasons it seems promising to characterize the present processes looking forward for single particle tracking data to be compared with the models. Moreover the two processes presented account the complexity of the phenomena directly from geometrical (waiting time timescales distribution) and/or electrophysiology (cell resistances and capacitance values distributions) properties of the system, that could be directly measured as it was done for spine density profiles in [San+11]. Finally the anomalous transport phenomena is generated by a proper superposition of classic processes, that is not ad-hoc but can be related to experimental observations, clearly simplifying also the computational efforts of the simulation procedures of the trajectories.

The cable model, fractional or linear, is used to describe subthreshold potentials or passive potentials, associated with dendritic processes in neurons. The traveling potential is summed up in the center of the cell, called soma, and an action potential is produced when a threshold is exceeded. Anomalous regimes of diffusion can then have a deep impact on the communication strength.

Diffusion results more anomalous, i.e. the fractional exponent α decreases, with increasing spine density [San+06]. Decreasing spine density is characteristic of aging [JDS97],[Dua03], pathologies as neurological disorders [NSS02] and Down's syndrome [SM80], then subdiffusive regimes are in some sense associated with a health condition. It has been suggested that increasing spine density should serve to compensate time delay of postsynaptic potentials along dendrites and to reduce their longtime temporal attenuation [HLW08].

Looking at our plotted solutions for the fractional cable equation when an impulsive potential is applied at the accessible end it can be noted from Fig.4.3 that peak high decreases more rapidly with decreasing α at early times, vice-versa is less suppressed at longer times, and the crossover time increases with decreasing α. Looking at the potential versus time it can also be noted that potential functions associated to lower α last for a longer time at an appreciable intensity and arrive faster at early times with respect to the normal diffusion case ($\alpha = 1$). By the way, when a constant potential is applied at the accessible end we note from Fig.4.4 that the exponential suppression of the potential along the dendrite is reduced for high X values with respect to normal diffusion. Instead, for small X the potential results just slightly more suppressed in the sub-diffusion process. These behaviors can be noticed also for the other cases in Fig.4.5 and Fig.4.6 - 4.7.

The presented fractional cable model satisfies the main biological features of the dendritic cell Signaling Problem. With respect to models solved for the Cauchy problem, our approach could include specific time-dependent boundary conditions,

which will allow reconstructing with accuracy the expected signal at the soma if the model will result capable to predict real data behavior. Furthermore the solutions can be computed directly, i.e. calculating the integral associated, as well as by Laplace Transform inversion [AW06] without any remarkable issue.

From a mathematical point of view, the Efros theorem extends the concept of convolution as an integral form that is consistent with a subordination-type integral. However such integral form does not necessary connote a subordinated process, as it has been shown for ggBM [MP08b], [DG+16], but could also be interpreted as a consequence of the random nature of the media in which particles are diffusing as shown in Section 4.2. This model can be also read as a generalization of time fractional diffusion processes where mass is not conserved due to leakage. This approach naturally recovers the solution for the time fractional case in the limit in which the leakage is put to zero in the integral forms.

Chapter 5

Conclusions

This thesis was centered on the modeling of biological stochastic processes, in particular, birth-death and diffusion processes. Three different mathematical tools were applied in the three chapters to model three different case studies: Master Equation (ME), stochastic differential equation (SDE) and partial differential equation (PDE). All the systems analyzed were characterized on the basis of their environment complexity. The case of LINE transposons ecosystem described in Chapter 2, in which the genomic environment is unique to each copy, directly affecting its dynamics parameters, was treated by the *competitive neutrality* assumption of the model proposed, in which stochastic mechanisms as demographic stochasticity, migration, and speciation are the most important forces shaping the community. Medium and/or ensemble complexity in the derivation of an alternative Langevin approach in Chapter 3 was described by a population of parameters characterizing the system, as the mass and radii of the particles composing the surround for the test particle motion, and almost equivalently by the introduction of a population of timescales and diffusivity parameters in the equations. Geometrical complexity, generated by the random shape and size of spines in dendrites, was modeled in Chapter 4 by a fractional generalization of the cable equation and a populations of scales in the stochastic processes derived.

Neutral assumption proposed in Chapter 2, is naturally broken by the introduction of a competition for the polymerase protein in the neutral model, which is necessary to start autonomous transposition. The model proposed was able to describe the observed distribution of the relative species abundances (RSA) of the LINE ecosystem in 42 Mammals genomes. Furthermore, it was able to catch some important features of the evolutionary history, and to point out the similarity between the RSA of genomes belonging to the same taxonomic order, and evidencing the presence of evolutive radiation phenomena within the same genome, when data were available, i.e., in two Murinides genomes (Mouse and Rat) and three primates (Homo sapiens, Rhesus macaque, Chimpanzee).

In Chapter 3 it was demonstrated that anomalous diffusion may arise in a heterogeneous ensemble of particles by a proper distribution of mass and radii within the Langevin approach. Furthermore suitable distributions of these parameters in terms of timescales and velocity diffusion coefficient was proposed in order to obtain the same PDFs associated to fractional Brownian motion and space and time fractional diffusion, observed in nature.

In Chapter 4 two different stochastic processes and the associated fractional generalization of the cable equation were introduced to describe anomalous diffusion

in spiny dendrites related to subthreshold potential propagation. The fundamental solutions for the Cauchy problem and first and second boundary problems were derived. These results were interpreted in comparison with other existing models in the literature. The possible advantage of this kind of models respect to real data was discussed.

Appendix A

A.1 Mean field approximation and reductiond to 1D problem

The system of LINEs can be considered as a two specie problem: the active copies and the inactive one. Such a system can be described by a two dimensional master equation as defined in Eq.(2.23).

We may define the mean field variables:

$$\begin{aligned} X &= \sum_{x,y} xP(x,y,t) =< x > \\ Y &= \sum_{x,y} yP(x,y,t) =< y > \end{aligned} \tag{A.1}$$

then

$$\begin{aligned} \dot{X} &= \sum_{x,y} x\dot{P}(x,y,t) \\ \dot{Y} &= \sum_{x,y} y\dot{P}(x,y,t) \end{aligned} \tag{A.2}$$

As an example we solve the significant pieces, the whole calculation is skipped.

$$\begin{aligned} \sum_{x,y} x(\Xi_x^- - 1)b_A xP(x,y,t) &= \sum_x x(\Xi_x^- - 1)b_A xP(x,t) \\ &= \sum_x b_A x(x-1)P(x-1,t) - \sum_x b_A x^2 P(x,t) \\ &= \sum_x b_A (x+1)xP(x,t) - \sum_x b_A x^2 P(x,t) \\ &= \sum_x b_A xP(x,t) \\ &= b_A X \end{aligned} \tag{A.3}$$

in which change of variable from n to n-1 has been done to collect P(x,y,t) from each addend. Analogously we may calculate:

$$\begin{aligned}\sum_{x,y} y(\Xi_y^- - 1)b_{AI}xyP(x,y,t) &= b_{AI}\sum_{x,y} y(x(y-1)P(x,y-1,t) - xyP(x,y,t)) \\ &= b_{AI}\sum_{x,y}(x(y+1)yP(x,y,t) - xy^2P(x,y,t)) \\ &= b_{AI}\sum_{x,y} xyP(x,y,t) \\ &\approx b_{AI}XY\end{aligned} \tag{A.4}$$

the last approximation is exact only if x,y are independent variables and the probability distribution is factorizable: $P(x,y,t) \approx P(x,t)P(y,t)$. In our case we keep also in mind that the coefficient b_{AI} is expected to be very small, hence the interaction should not affect dramatically the marginals.

With this approach we finally obtain that the 2D probabilistic master equation can be approximated by two coupled differential equations, one describing the active copies, the other describing the inactive copies. If we set $X = n_A$ and $Y = n_I$, finally we have:

$$\frac{dn_A}{dt} = b_A \cdot n_A - d_A \cdot n_A \tag{A.5}$$

$$\frac{dn_I}{dt} = (b_I + d_A) \cdot n_A + b_{AI} \cdot n_A \cdot n_I - d_I \cdot n_I \tag{A.6}$$

If we approximate n_A to a constant value, that is not exact but can be acceptable, the equation for the inactive copies recalls the equation for a birth death process with external influx:

$$\frac{dn_I}{dt} = b_{AI} \cdot \langle n_A \rangle \cdot n_I - d_I \cdot n_I + (b_I + d_A) \cdot \langle n_A \rangle \tag{A.7}$$

whom stationary solution in the probalistic form is a negative binomial, describing exactly the RSA of the ecosystem.

$$P_{RSA} = \langle \Phi_n \rangle = \Theta \frac{x^n}{n!}\Gamma(n + \Upsilon) \tag{A.8}$$

where $\Theta = \frac{S'}{[(1-x)^{-\Upsilon}-1]\Gamma(\Upsilon)}$.

A.2 Supplementary figures

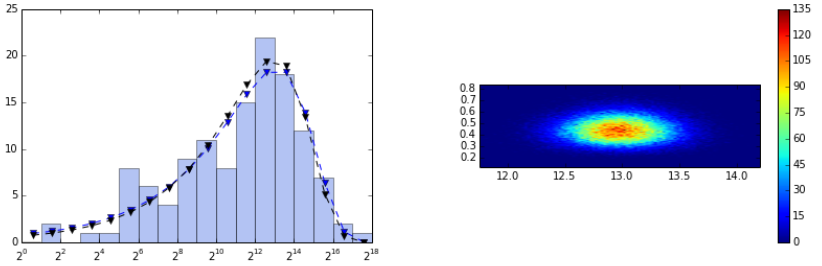

FIGURE A.1: **ABC results of neutral model fit of Cat RSA.**Comparison of the expected values from the mean parameters obtained with the ABC (black) to the Preston plot of the RSA and to the expected values obtained by non-linear least squares method (blue) on the same histogram(left panel);heat map of the posteriors of the parameters (right panel), log2 mean value (x axes), influx values (y axes).

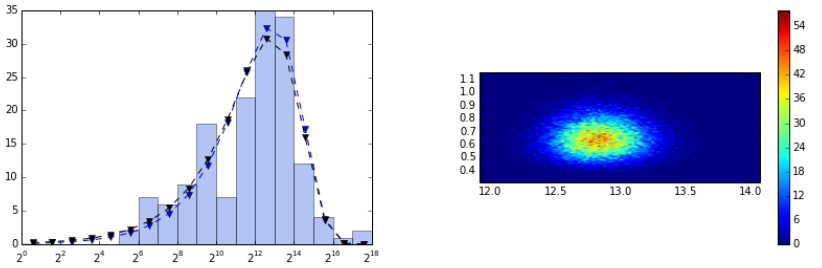

FIGURE A.2: **ABC results of neutral model fit of Chimpanzee RSA.**Comparison of the expected values from the mean parameters obtained with the ABC (black) to the Preston plot of the RSA and to the expected values obtained by non-linear least squares method (blue) on the same histogram(left panel);heat map of the posteriors of the parameters (right panel), log2 mean value (x axes), influx values (y axes).

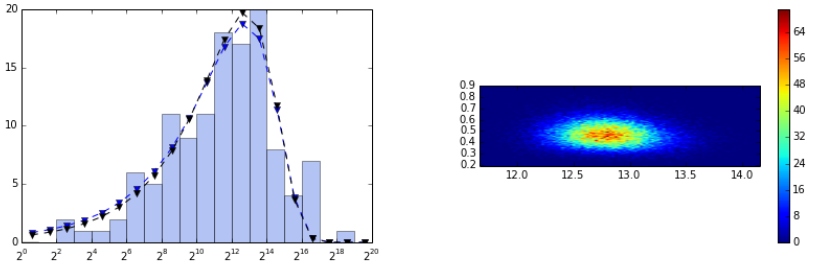

FIGURE A.3: **ABC results of neutral model fit of Cow RSA.**Comparison of the expected values from the mean parameters obtained with the ABC (black) to the Preston plot of the RSA and to the expected values obtained by non-linear least squares method (blue) on the same histogram(left panel);heat map of the posteriors of the parameters (right panel), log2 mean value (x axes), influx values (y axes).

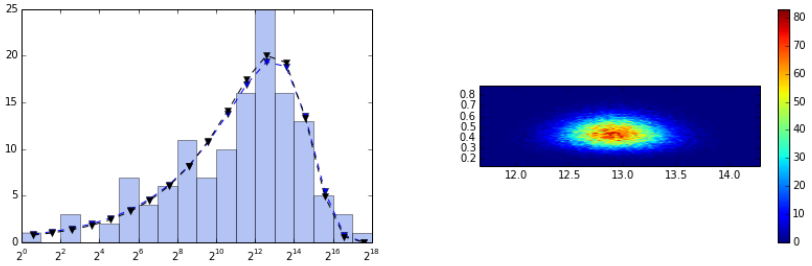

FIGURE A.4: **ABC results of neutral model fit of Dog RSA.**Comparison of the expected values from the mean parameters obtained with the ABC (black) to the Preston plot of the RSA and to the expected values obtained by non-linear least squares method (blue) on the same histogram(left panel);heat map of the posteriors of the parameters (right panel), log2 mean value (x axes), influx values (y axes).

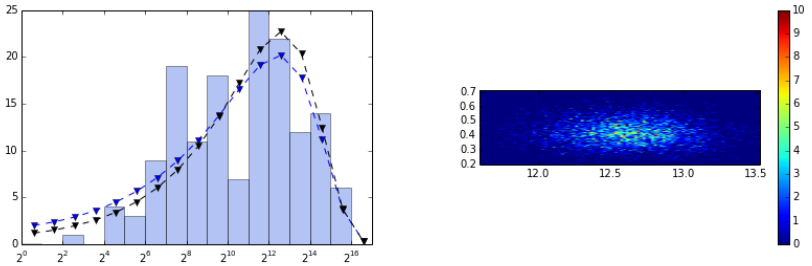

FIGURE A.5: **ABC results of neutral model fit of Rat RSA.**Comparison of the expected values from the mean parameters obtained with the ABC (black) to the Preston plot of the RSA and to the expected values obtained by non-linear least squares method (blue) on the same histogram(left panel);heat map of the posteriors of the parameters (right panel), log2 mean value (x axes), influx values (y axes).

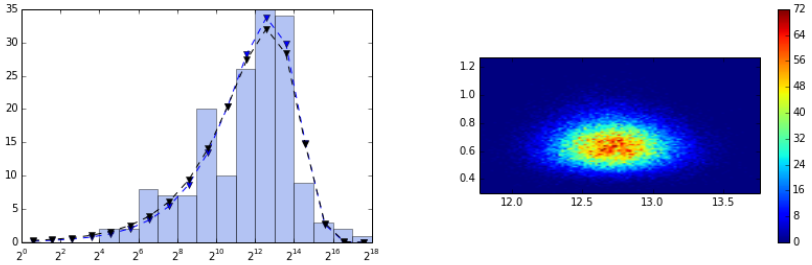

FIGURE A.6: **ABC results of neutral model fit of RhesuSrheMac3 RSA.**Comparison of the expected values from the mean parameters obtained with the ABC (black) to the Preston plot of the RSA and to the expected values obtained by non-linear least squares method (blue) on the same histogram(left panel);heat map of the posteriors of the parameters (right panel), log2 mean value (x axes), influx values (y axes).

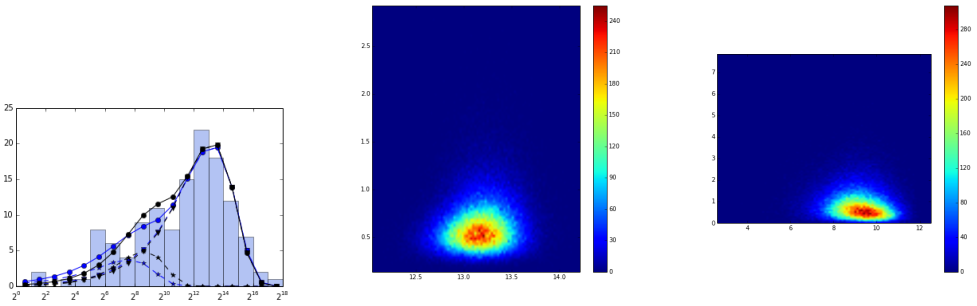

FIGURE A.7: **ABC results of mixture model fit of Cat RSA.**Comparison of the expected values from the mean parameters obtained with the ABC (black) to the Preston plot of the RSA and to the expected values obtained by non-linear least squares method (blue) on the same histogram(left panel);heat map of the posteriors of the parameters of the abund elements(center panel), heat map of the posteriors of the parameters of the rare elements(right panel), log2 mean value (x axes), influx values (y axes).

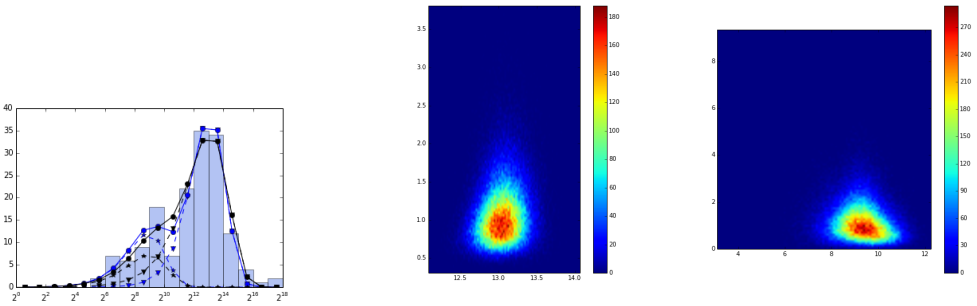

FIGURE A.8: **ABC results of mixture model fit of Chimpanzee RSA.**Comparison of the expected values from the mean parameters obtained with the ABC (black) to the Preston plot of the RSA and to the expected values obtained by non-linear least squares method (blue) on the same histogram(left panel);heat map of the posteriors of the parameters of the abund elements(center panel), heat map of the posteriors of the parameters of the rare elements(right panel), log2 mean value (x axes), influx values (y axes).

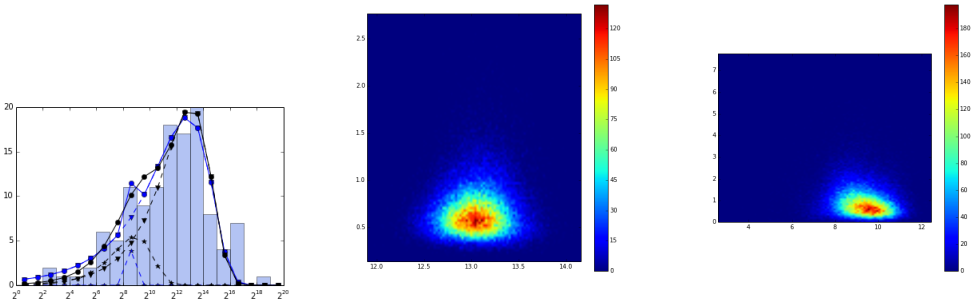

FIGURE A.9: **ABC results of mixture model fit of Cow RSA.**Comparison of the expected values from the mean parameters obtained with the ABC (black) to the Preston plot of the RSA and to the expected values obtained by non-linear least squares method (blue) on the same histogram(left panel);heat map of the posteriors of the parameters of the abund elements(center panel), heat map of the posteriors of the parameters of the rare elements(right panel), log2 mean value (x axes), influx values (y axes).

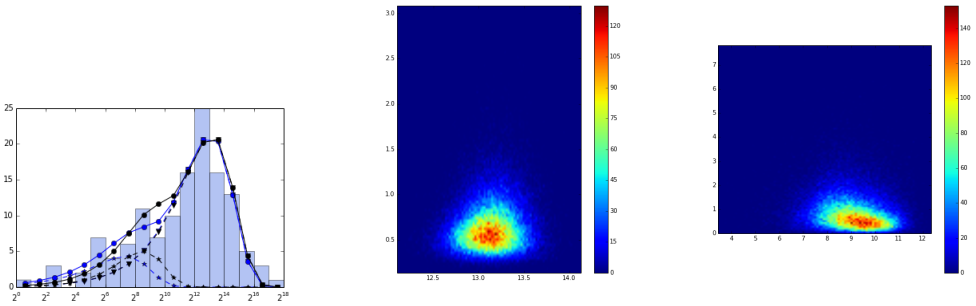

FIGURE A.10: **ABC results of mixture model fit of Dog RSA.**Comparison of the expected values from the mean parameters obtained with the ABC (black) to the Preston plot of the RSA and to the expected values obtained by non-linear least squares method (blue) on the same histogram(left panel);heat map of the posteriors of the parameters of the abund elements(center panel), heat map of the posteriors of the parameters of the rare elements(right panel), log2 mean value (x axes), influx values (y axes).

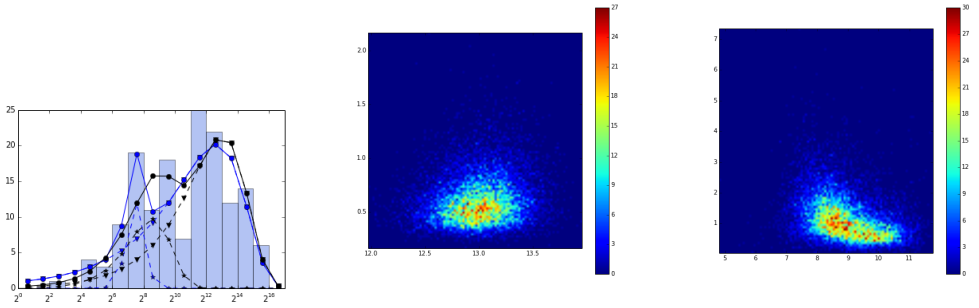

FIGURE A.11: **ABC results of mixture model fit of Rat RSA.**Comparison of the expected values from the mean parameters obtained with the ABC (black) to the Preston plot of the RSA and to the expected values obtained by non-linear least squares method (blue) on the same histogram(left panel);heat map of the posteriors of the parameters of the abund elements(center panel), heat map of the posteriors of the parameters of the rare elements(right panel), log2 mean value (x axes), influx values (y axes).

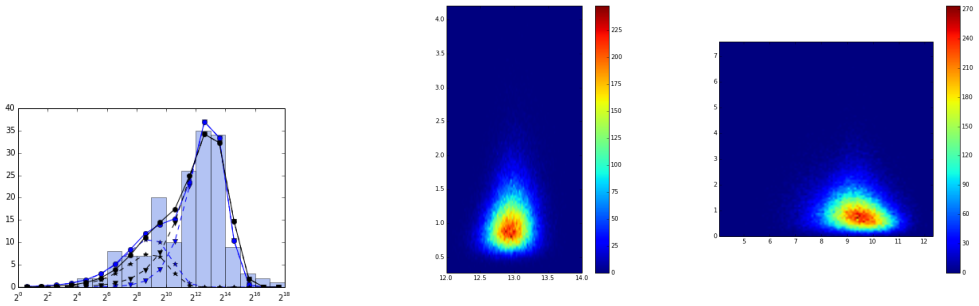

FIGURE A.12: **ABC results of mixture model fit of RhesuSrheMac3 RSA.**Comparison of the expected values from the mean parameters obtained with the ABC (black) to the Preston plot of the RSA and to the expected values obtained by non-linear least squares method (blue) on the same histogram(left panel);heat map of the posteriors of the parameters of the abund elements(center panel), heat map of the posteriors of the parameters of the rare elements(right panel), log2 mean value (x axes), influx values (y axes).

Appendix B

In this Appendix are reported some supplementary results and special cases of the model described in Chapter 3.

B.1 Examples of relations between $b(\tau)$ and $h(D_v)$

The motion of the center of mass of an ensemble of Brownian particles is determined by the distributions $b(\tau)$ and $h(D_v)$. When fluctuation dissipation theorem holds for each particle:

$$\tau = \frac{m}{6\pi\nu r}\,, \quad D_v = \frac{k_B T 6\pi\nu r}{m^2}\,. \tag{B.1}$$

The distributions $b(\tau)$ and $h(D_v)$ are determined by the mass and radius distributions of the heterogeneous ensemble. If the radius is a constant these distributions are correlated and determined by the same distribution of masses $\rho(m) \sim b(\tau)$, while $h(D_v) \sim h(1/m^2)$. Then we have:

$$h(D_v) = \frac{m^3}{2}\rho(m)|_{m=\frac{1}{\sqrt{D_v}}}\,. \tag{B.2}$$

Thanks to this relation several special cases can be derived.

Interesting distributions are the generalized gamma distribution:

$$\gamma^d_{\alpha,\beta}(D) = \frac{\beta}{d^\alpha \Gamma(\alpha/\beta)} D^{\alpha-1} e^{-(D/d)^\beta}\,, \tag{B.3}$$

and its inverse:

$$\tilde{\gamma}^d_{\alpha,\beta}(x) = \frac{\beta d^{\alpha\beta}}{\Gamma(\alpha)} D^{-\alpha\beta-1} e^{-(d/x)^\beta}\,. \tag{B.4}$$

If $h(D_v) = \gamma^{D_v^*}_{\alpha,\beta}(D_v)$ then:

$$\rho(m) = m^{\alpha'(\beta'-1)} \tilde{\gamma}^{m^*}_{\alpha',\beta'}(m)\,, \tag{B.5}$$

where $\alpha' = 2\alpha$, $\beta' = 2\beta$, $m^* = 1/\sqrt{(D_v^*)}$ and $m = 1/\sqrt{(D_v)}$.

In case of $\beta = 1/2$, $\rho(m)$ is an inverse gamma distribution with $\alpha' = 2\alpha$.

Another interesting case is the $h(D_v)$ associated to the PDF $b(\tau)$ suggested to obtain anomalous time scaling of the MSD, for the mass we have:

$$\rho(m) = \frac{\alpha}{\Gamma(1/\alpha)} \frac{1}{m} L_{\alpha}^{-\alpha}(m) \ , \tag{B.6}$$

the corresponding PDF for diffusivity is:

$$h(D_v) = \frac{\alpha}{\Gamma(1/\alpha)} \frac{1}{2D_v} L_{\alpha}^{-\alpha}\left(1/\sqrt{(}D_v)\right) \ . \tag{B.7}$$

When $h(D_v) = M_{\alpha}(D_v)$, Mainardi-Wright function, then:

$$\rho(m) = \frac{2}{\alpha} m^{2/\alpha - 1} L_{\alpha}^{-\alpha}(m^{\alpha/2}) \, . \tag{B.8}$$

B.2 Special case with harmonic Langevin oscillator

Subdiffusion is related to the presence of some negative part in the VACF. The identity $\int_0^{\infty} R(t) = 0$, is satisfied for an harmonic potential. It correspond to add a term $-m\omega_0^2 X$ in the Langevin equation.

$$m\frac{dV}{dt}(t) = -\frac{m}{\tau_c} V(t) - m\omega_0^2 X(t) + \Gamma\xi(t) \ , \tag{B.9}$$

The VACF becomes

$$R(t) = \int_0^{\infty} D_v h(D_v) dD_v \int_0^{\infty} \tau_c e^{-\frac{t}{2\tau_c}} \left[cos(\omega_1 t) - \frac{1}{2\omega_1 \tau_c} sin(\omega_1 t) \right] b(\tau_c) d\tau_c \tag{B.10}$$

where $\omega_1 = \sqrt{\omega_0^2 - \frac{1}{4\tau_c^2}}$

we considered an harmonic potential with constant characteristic frequency of oscillation ω_0, that could be related to the application of a constant external field to the system.

It results that in the long time limit the velocity correlatiom function scales with time as the power law: $\lim_{t\to\infty} R(t) \propto t^{-\nu}$, $\nu = 2 - \alpha$. Because the time scale parameter is stochastic, some components are subjected to a stronger potential then others. However since fractional behavior is associated to the tail of the distribution of the timescales, the oscillator term is always not negligible for the trajectories determining the dynamics.

$$\begin{cases} \tau_c > \frac{1}{2\omega_0}, & \text{oscillation (dumped)} \\ \tau_c < \frac{1}{2\omega_0}, & \text{over dumped} \end{cases} . \tag{B.11}$$

If the characteristic oscillator frequency is correlated to the timescale of the single process the behavior of the system could be much more complicated and lead to completely different behaviours. We may consider as example that the coefficient of the viscous drag and of the oscillator have the same order of magnitude: $\omega_0 = \frac{\sqrt{c}}{\tau_c}$, where c is a positive real constant of order 1. We obtain $\omega_1 = \frac{\sqrt{4c-1}}{2\tau_c} = \frac{a}{2\tau_c}$. We distinguish the corresponding underdumped behavior (oscillations) when $a \in \Re$ and overdamped behaviour (no oscillations) when $a \in \Im$, in term sof the constant c

we have:

$$\begin{cases} c > \frac{1}{4}, & \text{oscillation (dumped)} \\ c < \frac{1}{4}, & \text{over dumped} \end{cases}, \tag{B.12}$$

then depending on the value of the constant we should observe always a dumped or an over dumped regime.

Then we may rewrite:

$$R(t) = \int_0^\infty D_v h(D_v) dD_v \int_0^\infty \tau_c e^{-\frac{t}{2\tau_c}} \left[cos(\frac{a}{2\tau_c}t) - \frac{1}{a} sin(\frac{a}{2\tau_c}t) \right] b(\tau_c) d\tau_c \tag{B.13}$$

We may rewrite the trigonometric functions in their exponential forms ($cos(x) = \frac{e^{ix}+e^{-ix}}{2}$,$sin(x) = \frac{e^{ix}-e^{-ix}}{2i}$), then the integral result the sum of two parts:

$$R(t) = \int_0^\infty D_v h(D_v) dD_v \int_0^\infty \frac{1}{2ia} \left[(1+ia) e^{-\frac{t(1+ia)}{2\tau_c}} - (1-ia) e^{-\frac{t(1-ia)}{2\tau_c}} \right] \tau_c b(\tau_c) d\tau_c \tag{B.14}$$

until $a \neq 0$ it can be solved in the same way as the super diffusive case. so we have: thanks to the integral rapresentation of the extremal Levy density distribution which characterize the fractional scaling of $b(\tau_c)$

$$L_\alpha^{-\alpha}(x) = \frac{1}{\alpha x} \frac{1}{2\pi i} \int_{\gamma - i\infty}^{\gamma + i\infty} \frac{\Gamma(s/\alpha)}{\Gamma(s)} x^s ds, \quad 0 < \alpha < 1\,, \tag{B.15}$$

hence we have:

$$\begin{aligned}
R(t) &= \langle D_V \rangle \frac{\alpha}{\Gamma(1/\alpha)} \int_0^\infty \frac{1}{2ia} \left[(1+ia) e^{-\frac{t(1+ia)}{2\tau_c}} - (1-ia) e^{-\frac{t(1-ia)}{2\tau_c}} \right] L_\alpha^{-\alpha} \left(\frac{\tau_c}{\tau_*} \right) d\tau_c \\
&= \langle D_V \rangle \frac{\alpha}{\Gamma(1/\alpha)} \int_0^\infty \frac{1}{2ia} \left[(1+ia) e^{-\frac{t(1+ia)}{2\tau_c}} - (1-ia) e^{-\frac{t(1-ia)}{2\tau_c}} \right] \left[\frac{1}{\alpha} \frac{1}{2\pi i} \int_{\gamma - i\infty}^{\gamma + i\infty} \frac{\Gamma(s/\alpha)}{\Gamma(s)} \left(\frac{\tau_c}{\tau_*} \right)^{(s-1)} ds \right] d\tau_c \\
&= \langle D_V \rangle \frac{\alpha}{\Gamma(1/\alpha)} \frac{1}{\alpha} \frac{1}{2\pi i} \int_{\gamma - i\infty}^{\gamma + i\infty} \frac{\Gamma(s/\alpha)}{\Gamma(s)} \left[\int_0^\infty \frac{1}{2ia} \left[(1+ia) e^{-\frac{t(1+ia)}{2\tau_c}} - (1-ia) e^{-\frac{t(1-ia)}{2\tau_c}} \right] \left(\frac{\tau_c}{\tau_*} \right)^{s-1} d\tau_c \right] ds \\
&\xi_1 = \frac{t(1+ia)}{2\tau_c}, \xi_2 = \frac{t(1-ia)}{2\tau_c} \\
&= \langle D_V \rangle \langle \tau_c \rangle \frac{1}{\alpha} \frac{1}{2\pi i} \int_{\gamma - i\infty}^{\gamma + i\infty} \frac{\Gamma(s/\alpha)}{\Gamma(s)} \frac{1}{2ia} \cdot \\
&\quad \cdot \left[(1+ia) \int_0^\infty e^{-\xi_1} \xi_1^{-1-s} \left(\frac{t(1+ia)}{2\tau_*} \right)^s d\xi_1 - (1-ia) \int_0^\infty e^{-\xi_2} \xi_2^{-1-s} \left(\frac{t(1-ia)}{2\tau_*} \right)^s d\xi_2 \right] ds \\
&= \langle D_V \rangle \langle \tau_c \rangle \frac{1}{\alpha} \frac{1}{2\pi i} \frac{1}{2ia} \cdot \\
&\quad \cdot \left[(1+ia) \int_{\gamma - i\infty}^{\gamma + i\infty} \frac{\Gamma(s/\alpha)\Gamma(-s)}{\Gamma(s)} \left(\frac{t(1+ia)}{2\tau_*} \right)^s ds - (1-ia) \int_{\gamma - i\infty}^{\gamma + i\infty} \frac{\Gamma(s/\alpha)\Gamma(-s)}{\Gamma(s)} \left(\frac{t(1-ia)}{2\tau_*} \right)^s ds \right],
\end{aligned} \tag{B.16}$$

analogously to the superdiffusive case it can be solved through the residues theorem considering the poles $s/\alpha + 1 = -n$ or $s = n$, with $n = 0, 1, 2..\infty$.

Solving for the poles $s/\alpha + 1 = -n$ it can be written as:

$$R(t) = \langle D_V \rangle \langle \tau_c \rangle \sum_{n=1}^{\infty} \frac{(-1)^n}{n!} \frac{\Gamma(\alpha n)}{\Gamma(-\alpha n)} \left(\frac{t}{2\tau_*} \right)^{-\alpha n} \frac{1}{2ia} \left[(1+ia)^{-\alpha n+1} - (1-ia)^{-\alpha n+1} \right] . \tag{B.17}$$

Then we may distinguish 2 cases, the dumped and the overdumped.
When $a \in \Re$ we have:

$$R(t) = \langle D_V \rangle \langle \tau_c \rangle \sum_{n=1}^{\infty} \frac{(-1)^n}{n!} \frac{\Gamma(\alpha n)}{\Gamma(-\alpha n)} \left(\frac{t}{2\tau_*} \right)^{-\alpha n} \frac{1}{a} \left[\rho^{-\alpha n+1} sin(\varphi(1-\alpha n)) \right] , \tag{B.18}$$

where $\rho = \sqrt{1+a^2}$ and $\varphi = arctan(a)$
When $t \to \infty$ only the first term survives and we find:

$$R(t) = \langle D_V \rangle \langle \tau_c \rangle \frac{\Gamma(\alpha+1)}{\Gamma(1-\alpha)} \left(\frac{t}{\tau_*} \right)^{-\alpha} \frac{1}{a} \left[\rho^{1-\alpha} sin(\varphi(1-\alpha)) \right] , \tag{B.19}$$

which is enough to obtain the desired scaling of the MSD $\sigma_x^2(t) \propto t^{\phi}$, with $\phi = 2-\alpha$.
for the over dumped $a \in \Im$ and we obtain:

$$R(t) = \langle D_V \rangle \langle \tau_c \rangle \sum_{n=1}^{\infty} \frac{(-1)^n}{n!} \frac{\Gamma(\alpha n)}{\Gamma(-\alpha n)} \left(\frac{t}{2\tau_*} \right)^{-\alpha n} \frac{1}{2|a|} \left[(1+|a|)^{-\alpha n+1} - (1-|a|)^{-\alpha n+1} \right] , \tag{B.20}$$

the leading term for the limit $t \to \infty$ is:

$$R(t) = \langle D_V \rangle \langle \tau_c \rangle \left(\frac{t}{2\tau_*} \right)^{-\alpha} \frac{1}{2|a|} \left[(1+|a|)^{-\alpha+1} - (1-|a|)^{-\alpha+1} \right] \tag{B.21}$$

Considering the poles in the other semi-plane, $s = n$ with $n = 0, 1, 2..\infty$, for the short time limit we have:

$$R(t) = \langle D_V \rangle \langle \tau_c \rangle \frac{1}{\alpha} \sum_{n=0}^{\infty} \frac{(-1)^n}{n!} \frac{\Gamma(n/\alpha)}{\Gamma(n)} \left(\frac{t}{2\tau_*} \right)^{n} \cdot \frac{1}{2ia} \left[(1+ia)^{n+1} - (1-ia)^{n+1} \right] \tag{B.22}$$

converges to $R(0) = \langle D_V \rangle \langle \tau_c \rangle$.

A possible explanation to this behavior is that changing the timescale as well the oscillator frequency is equivalent to a time streching of each component determined by its characteristic timescale, i.e a large part of the components are so streched respect to the others that never reach negative values also if at their own time scale we should observe oscillation.

B.2.1 Solution via Laplace Transform

We can demonstrate the same result working with its LT.

The laplace transform can be applied only if the integral over τ_c has finite value, which can be demonstrated numerically or analitically.Then we have:

$$\begin{aligned}\mathcal{L}[R(t)](s) &= \langle D_v\rangle \int_0^\infty \frac{s}{s^2+s/\tau_c+c/\tau_c^2}\tau_c b(\tau_c)d\tau_c \\ &= \langle D_v\rangle \int_0^\infty \frac{s\tau_c^2}{s^2\tau_c^2+s\tau_c+c}\frac{\alpha}{\Gamma(1/\alpha)}L_\alpha^{-\alpha}(\tau_c/\tau_*)d\tau_c\end{aligned} \tag{B.23}$$

This integral takes the form:

$$\int_0^\infty g(xy)h(y)y^{z-1}dy = \frac{1}{2\pi i}\int_{\mathcal{L}} x^{-q}g^*(q)h^*(z-q)dq \tag{B.24}$$

considering $x = s\tau_*$, $y = \tau = \tau_c/\tau_*$, $z = 2$, $g(\tau) = \frac{1}{\tau^2-\tau+c}$, $h(\tau) = \tau L_\alpha^{-\alpha}(\tau)$. $g^*(q)$, $h^*(q)$ are the Mellin transforms of the correspective functions. $g(\tau)$ can be rewritten as :

$$\begin{aligned}g(\tau) &= \frac{1}{\tau^2-\tau+c} = \frac{1}{(\tau+1/2)^2-(1/4-c)} \\ &= \frac{1}{(\tau+1/2-\sqrt{1/4-c})(\tau+1/2+\sqrt{1/4-c})} \\ &= \frac{1}{(\tau+A)(\tau+B)}\,.\end{aligned} \tag{B.25}$$

The Mellin Transform of $g(\tau)$ is $g^*(q) = \pi csc(\pi q)\frac{A^{q-1}-B^{q-1}}{B-A} = \Gamma(q)\Gamma(1-q)\frac{A^{q-1}-B^{q-1}}{B-A}$, with $|arg(A)|, |arg(B)| < \pi$.

The Mellin Trasform of $h(\tau)$ is $h^*(q) = \frac{1}{\alpha}\frac{\Gamma(-q/\alpha)}{\Gamma(-q)}$.

The expressions for Laplace and Mellin trasforms have been taken from the Bateman project which collect a huge number of transforms for several type of functions.

So we have:

$$\begin{aligned}\mathcal{L}[R(t)](s) &= \langle D_v\rangle\frac{\alpha}{\Gamma(1/\alpha)}\tau_*^3 s\frac{1}{2\pi i}\int_{\mathcal{L}}(\tau_* s)^{-q}g^*(q)h^*(2-q)dq \\ &= \langle D_v\rangle\frac{\alpha}{\Gamma(1/\alpha)}\tau_*^2\frac{1}{2\pi i}\int_{\mathcal{L}}(\tau_* s)^{1-q}\Gamma(q)\Gamma(1-q)\frac{A^{q-1}-B^{q-1}}{B-A}\frac{1}{\alpha}\frac{\Gamma((-(2-q)/\alpha)}{\Gamma(q-2)}dq \\ &= \langle D_v\rangle\frac{\alpha}{\Gamma(1/\alpha)}\tau_*^2\frac{1}{2\pi i}\int_{\mathcal{L}}\left[\left(\frac{\tau_* s}{B}\right)^{1-q}-\left(\frac{\tau_* s}{A}\right)^{1-q}\right]\frac{1}{B-A}\Gamma(2-q)\Gamma((1-(2-q)/\alpha)dq\end{aligned} \tag{B.26}$$

This integral is then written as residues serie for the poles $q = 2-\alpha(n+1)$, in order to obtain the long time limit behaviour of the VACF ($t\to\infty$, $s\to 0$), or for $q = 1+n$, to obtain its short time limit behavior ($t\to 0$, $s\to\infty$).

Then for the long time limit we have:

$$\begin{aligned}\lim_{t\to\infty} R(t) &= \lim_{s\to 0} s\cdot\mathcal{L}[R(t)](s)\\ &= \lim_{s\to 0} s\cdot\langle D_v\rangle\langle\tau_c\rangle\tau_*\sum_{n=0}^{\infty}\alpha\frac{(-1)^n}{n!}\Gamma(\alpha(n+1))\left(\tau_* s\right)^{-1+\alpha(n+1)}\left[B^{+1-\alpha(n+1)}-A^{1-\alpha(n+1)}\right]\frac{1}{B-A}\\ &= \langle D_v\rangle\langle\tau_c\rangle\frac{\Gamma(\alpha+1)}{B-A}\left(\frac{1}{\tau_* s}\right)^{-\alpha}\left[B^{1-\alpha}-A^{1-\alpha}\right]\\ &\sim \langle D_v\rangle\langle\tau_c\rangle\frac{\Gamma(\alpha+1)}{B-A}\left(\frac{t}{\tau_*}\right)^{-\alpha}\left[B^{1-\alpha}-A^{1-\alpha}\right]\end{aligned} \tag{B.27}$$

where $A = 1/2 - \sqrt{1/4-c}$ and $B = 1/2 + \sqrt{1/4-c}$, that can be written in term of $a = \sqrt{4c-1}$: $A = \frac{1-ia}{2}$,$B = \frac{1+ia}{2}$. Then:

$$\begin{aligned}\lim_{t\to\infty} R(t) &\sim \langle D_v\rangle\langle\tau_c\rangle\frac{\Gamma(\alpha+1)}{B-A}\left(\frac{t}{\tau_*}\right)^{-\alpha}\left[B^{1-\alpha}-A^{1-\alpha}\right]\\ &\sim \langle D_v\rangle\langle\tau_c\rangle\frac{\Gamma(\alpha+1)}{2ia}\left(\frac{t}{2\tau_*}\right)^{-\alpha}\left[(1+ia)^{1-\alpha}-(1-ia)^{1-\alpha}\right]\end{aligned} \tag{B.28}$$

B.2.2 Failure of confinment

The failure of the confining behavior of an harmonic potential with timescale dependent characteristic oscillation frequency can be demonstrated by explicit calculation of the expression: $\int_0^\infty R(t)$, that is expected to be zero in case of subdiffusive regime and infinity in case of superdiffusive regime. This integral can be calculated thanks to the relation with Laplace Transforms: $\int_0^\infty R(t) = \mathcal{L}[R(t)](s)|_{s=0}$.

When the coefficient of the viscous drag and of the oscillator have the same order of magnitude, $\omega_0 = \frac{\sqrt{c}}{\tau_c}$, we have:

$$\mathcal{L}[R(t)](s) = \langle D_v\rangle\int_0^\infty \frac{s\tau_c^2}{s^2\tau_c^2+s\tau_c+c}\frac{\alpha}{\Gamma(1/\alpha)}L_\alpha^{-\alpha}(\tau_c/\tau_*)d\tau_c \tag{B.29}$$

that for $s\to 0$ leads to an indetermined form $0\cdot\infty$.

When the characteristic oscillation frequency is constant this does not happen because we have:

$$\begin{aligned}\mathcal{L}[R(t)](s) &= \langle D_v\rangle\int_0^\infty \frac{s\tau_c}{s^2\tau_c+s+\omega_0\tau_c}\frac{\alpha}{\Gamma(1/\alpha)}L_\alpha^{-\alpha}(\tau_c/\tau_*)d\tau_c\\ &= s\cdot\langle D_v\rangle\langle\tau_c\rangle/\omega_0\end{aligned} \tag{B.30}$$

Bibliography

[ABC07] S. Abe, C. Beck, and E. G. D. Cohen. "Superstatistics, thermodynamics, and fluctuations". In: *Phys. Rev. E* 76 (2007), p. 031102.

[Abe14] S. Abe. "Fokker–Planck theory of nonequilibrium systems governed by hierarchical dynamics". In: *Found. Phys.* 44 (2014), pp. 175–182.

[Ade76] S. A. Adelman. "Fokker–Planck equations for simple non-Markovian systems." In: *The Journal of Chemical Physics* 64 (1976), p. 124.

[AK06] G. Abrusán and H. J. Krambeck. "Competition may determine the diversity of transposable elements". In: *Theoretical Population Biology* 70.3 (2006), pp. 364–375.

[al.01] E. S. Lander et al. "International Human Genome Sequencing. Initial sequencing and analysis of the human genome". In: *Nature* 409.6822 (2001), pp. 860–921.

[ALC07] T. S. Boutin A. Le Rouzic and P. Capy. "Long-term evolution of transposable elements". In: *Proc Natl Acad Sci U S A* 104.49 (2007), pp. 19375–19380.

[AW06] J. Abate and W. Ward. "A unified framework for numerically inverting Laplace transforms". In: *INFORMS J. on Computing* 18.4 (Jan. 2006), pp. 408–421. ISSN: 1526-5528.

[Aza+06] S. Azaele et al. "Dynamical evolution of ecosystems". In: *Nature* 444.7121 (2006), pp. 926–928. ISSN: 0028-0836.

[Bal+12] D. Baleanu et al. *Fractional Calculus: Models and Numerical Methods*. Series on Complexity, Nonlinearity and Chaos, volume 3. New Jersey: World Scientific Publishers, 2012.

[Bar79] G. I. Barenblatt. *Similarity, Self–Similarity, and Intermediate Asymptotics*. New York: Consultants Bureau, 1979.

[Bat54] H. Bateman. *Tables of Integral Transforms [Volumes I]*. McGraw-Hill Book Company, 1954.

[BC03] C. Beck and E. G. D. Cohen. "Superstatistics, thermodynamics, and fluctuations". In: *Physica A* 322 (2003), pp. 267–275.

[Bec01] C. Beck. "Dynamical foundations of nonextensive statistical mechanics". In: *Phys. Rev. Lett.* 87 (2001), p. 180601.

[Bec02] C. Beck. "Non-additivity of Tsallis entropies and fluctuations of temperature". In: *Europhys. Lett.* 57 (2002), pp. 329–333.

[BF01] S. Boissinot and A. V. Furano. "Adaptive evolution in LINE-1 retrotransposons ". In: *Mol. Biol. Evol* 18 (2001), pp. 2186–2194.

[BG90] J.P. Bouchaud and A. Georges. "Anomalous diffusion in disordered media: Statistical mechanisms, models and physical applications". In: *Physics Reports* 195.4 (1990), pp. 127 –293. ISSN: 0370-1573.

[BIHS] T. A. M. Langlands "B. I. Henry and P. Straka". *Complex Physical, Biophysical and Econophysical Systems. An Introduction to Fractional Diffusion*.

[Bié10] C. Biémont. "A brief history of the status of transposable elements: From junk DNA to major players in evolution". In: *Genetics* 186.4 (2010), pp. 1085–1093.

[BMN68] Benoit B., Mandelbrot, and John W. Van Ness. "Fractional Brownian Motions, Fractional Noises and Applications". In: *SIAM Review* 10.4 (1968), pp. 422–437. ISSN: 00361445.

[Bod+16] A. S. Bodrova et al. "Underdamped scaled Brownian motion: (non-)existence of the overdamped limit in anomalous diffusion". In: *Sci. Rep.* 6 (2016), p. 30520.

[Bro+03] B. Brouha et al. " Hot L1s account for the bulk of retrotransposition in the human population." In: *Proc. Natl. Acad. Sci. U. S. A.* 100, 2003, pp. 5280–5285.

[Bur+11] S. Burov et al. "Single particle tracking in systems showing anomalous diffusion: the role of weak ergodicity breaking." In: *Phys. Chem. Chem. Phys.* 13.5 (2011), pp. 1800–1812.

[Bur+12] K. Burnecki et al. "Universal Algorithm for Identification of Fractional Brownian Motion. A Case of Telomere Subdiffusion." In: *Biophysical Journal* 103 (9 2012), pp. 1839–1847.

[CB09] R. Cordaux and M.A. Batzer. "The impact of retrotransposons on human genome evolution". In: *Nat Rev Genet* 10.10 (2009), pp. 691–703.

[CCM13] A.G. Cherstvy, A.V. Chechkin, and R. Metzler. "Anomalous diffusion and ergodicity breaking in heterogeneous diffusion processes". In: *New Journal of Physics* 15.8 (2013), p. 083039.

[CGE00] A. Caspi, R. Granek, and M. Elbaum. "Enhanced Diffusion in Active Intracellular Transport". In: *Phys. Rev. Lett.* 85 (26 2000), pp. 5655–5658.

[CL03] J. M. Chase and M. A. Leibold. *Ecological Niches: Linking Classical and Contemporary Approaches*. Chicago: University Press, 2003.

[CL89] B. Charlesworth and C. H. Langley. "The population genetics of Drosophila transposable elements." In: *Annual review of genetics* 23 (1989), pp. 251–287.

[CM16] A. G. Cherstvy and R. Metzler. "Anomalous diffusion in time-fluctuating non-stationary diffusivity landscapes". In: *Phys. Chem. Chem. Phys.* 18 (34 2016), pp. 23840–23852.

[CMS76] J. M. Chambers, C. L. Mallows, and B. W. Stuck. "A method for simulating skewed stable random variables". In: *J. Amer. Statist. Assoc.* 71 (1976), pp. 340–344.

[CO13] M. Cowley and R. J. Oakey. "Transposable Elements Re-Wire and Fine-Tune the Transcriptome". In: *PLoS Genetics* 9 (2013), p. 1.

[Cri+14] S. W. Criscione et al. "Transcriptional landscape of repetitive elements in normal and cancer human cells." In: *BMC genomics* 15 (2014), p. 583.

[CS14] M. V. Chubynsky and G. W. Slater. "Diffusing Diffusivity: A Model for Anomalous, yet Brownian, Diffusion". In: *Phys. Rev. Lett.* 113 (9 2014), p. 098302.

[DCB05] G. Deceliere, S. Charles, and C. Biémont. "The dynamics of transposable elements in structured populations". In: *Genetics* 169.1 (2005), pp. 467–474.

[DG+16] D.Molina-García et al. "Fractional kinetics emerging from ergodicity breaking in random media". In: *Physical Review E* 94 (5 Nov. 2016).

[DS94] D.Johnston and S.M.S.Wu. *Foundations of Cellular Neurophysiology (Bradford Books)*. 1st ed. Bradford Books. The MIT Press, 1994. ISBN: 0262100533, 9780262100533, 0262100533.

[Dua03] H. Duan. "Age-related Dendritic and Spine Changes in Corticocortically Projecting Neurons in Macaque Monkeys". In: *Cerebral Cortex* 13 (9 Sept. 2003).

[Ein05] A. Einstein. "Uber die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen." In: *Annalen der Physik* 322 (8 1905).

[EK10] J. Ernst and M. Kellis. "Discovery and characterization of chromatin states for systematic annotation of the human genome." In: *Nature biotechnology* 28.8 (2010), pp. 817–825.

[EK12] J. Ernst and M. Kellis. "ChromHMM: automating chromatin-state discovery and characterization". In: *Nature Methods* 9 (2012), pp. 215–216.

[Fel71] W. Feller. *An Introduction to Probability Theory and its Applications*. second. Vol. 2. New York: Wiley, 1971.

[Fes08] C. Feschotte. "The contribution of transposable elements ot the evolution of regulatory networks". In: *Nature Reviews Genetics* 9.5 (2008), pp. 397–405.

[Fur+94] A. V. Furano et al. "Amplification of the Ancient Murine Lx Family of Long Interspersed Repeated DNA Occurred During the Murine Radiation". In: *J. Mol. Evol.* 38 (1994), pp. 18–27.

[Gar90] C. W. Gardiner. *Handbook of Stochastic Methods for Physics, Chemistry and the Natural Sciences*. second. Springer-Verlag, 1990.

[GC04] S. Gheorghiu and M.-O. Coppens. "Heterogeneity explains features of "anomalous" thermodynamics and statistics". In: *Proc. Natl. Acad. Sci. USA* 101 (2004), pp. 15852–15856.

[GC06] I. Golding and E. C. Cox. "Physical nature of bacterial cytoplasm". In: *Phys. Rev. Lett.* 96 (2006), p. 098102.

[Gil01] D. T. Gillespie. "Approximate accelerated stochastic simulation of chemically reacting systems". In: *The Journal of Chemical Physics* 115.4 (2001), pp. 1716–1733.

[Gil76] D. T. Gillespie. "A general method for numerically simulating the stochastic time evolution of coupled chemical reactions". In: *Journal of Computational Physics* 22.4 (1976), pp. 403 –434. ISSN: 0021-9991.

[Gio+07] J. Giordano et al. "Evolutionary history of mammalian transposons determined by genome-wide defragmentation". In: *PLoS Computational Biology* 3.7 (2007), pp. 1321–1334.

[GM97] R. Gorenflo and F. Mainardi. "Fractional Calculus: Integral and Differential Equations of Fractional Order". In: *Fractals and Fractional Calculus in Continuum Mechanics*. Ed. by A. Carpinteri and F. Mainardi. CISM Courses and Lecture Notes, Vol. 378. Springer-Verlag, Wien, 1997, pp. 223–276. ISBN: 3-211-82913-X.

[Gor+02a] R. Gorenflo et al. "Discrete random walk models for space-time fractional diffusion". In: *Chem. Phys.* 284 (2002), pp. 521–541.

[Gor+02b] R. Gorenflo et al. "Fractional diffusion: probability distributions and random walk models". In: *Physica A* 305.1–2 (2002), pp. 106–112.

[GP12] J. Gonzalez and D. A. Petrov. "Evolution of Genome Content: Population Dynamics of Transposable Elements in Flies and Humans". In: (2012), pp. 361–383.

[Gra04] U. Graf. *Applied Laplace Transforms and z-Transforms for Scientists and Engineers*. Springer, 2004.

[HA04] S. P. Hubbell and L. B. de Água. "The unified neutral theory of biodiversity and biogeography: reply." In: *Ecology* 85.11 (2004), pp. 3175–3178.

[Hap09] Crawford J. W. & Young I. M. Hapca S. "Anomalous Diffusion of Heterogeneous Populations Characterized by Normal Diffusion at The Individual Level". In: *J. R. Soc. Interface* (2009).

[HB04] J. S. Han and J. D. Boeke. " A highly active synthetic mammalian retrotransposon." In: *Nature* 429 (2004), pp. 314–318.

[He+08] Y. He et al. "Random time-scale invariant diffusion and transport coefficients". In: *Phys. Rev. Lett.* 101 (2008), p. 058101.

[HF13] F. Höfling and T. Franosch. "Anomalous transport in the crowded world of biological cells". In: 046602 (2013). ISSN: 0034-4885.

[HLW08] B.I. Henry, T.A.M. Langlands, and S.L. Wearne. "Fractional cable models for spiny neuronal dendrities". In: *Phys. Rev. Lett* 100 (2008), pp. 128103/1–3.

[HS17] D. Holcman and Z. Schuss. "100 years after Smoluchowski: stochastic processes in cell biology". In: *Journal of Physics A: Mathematical and Theoretical* 50.9 (2017), p. 093002. ISSN: 1751-8113.

[IM13] A. Iomin and V. Méndez. "Reaction-subdiffusion front propagation in a comblike model of spiny dendrites". In: *Phys. Rev. E* 88 (1 2013), p. 012706.

[Ion+17] C. Ionescu et al. "The role of fractional calculus in modelling biological hhenomena: A review." In: *Communications in Nonlinear Science and Numerical Simulation* (2017).

[Jav+13] M. Javanainen et al. "Anomalous and normal diffusion of proteins and lipids in crowded lipid membranes". In: *Faraday Discuss.* 161 (0 2013), pp. 397–417.

[Jav+14] A. Javer et al. "Persistent super-diffusive motion of Escherichia coli chromosomal loci." In: *Nature communications* 5.May (2014), p. 3854. ISSN: 2041-1723.

[JDS97] B. Jacobs, L. Driscoll, and M. Schall. "Life-span dendritic and spine changes in areas 10 and 18 of human cortex: A quantitative golgi study". In: *The Journal of Comparative Neurology* 386 (4 Oct. 1997).

[Jeo+11] J.H. Jeon et al. "In Vivo Anomalous Diffusion and Weak Ergodicity Breaking of Lipid Granules". In: *Phys. Rev. Lett.* 106.4 (2011), p. 048103.

[Jeo+16] J.H. Jeon et al. "Protein Crowding in Lipid Bilayers Gives Rise to Non-Gaussian Anomalous Lateral Diffusion of Phospholipids and Protein". In: *Phys. Rev. X* 6 (2016), p. 021006.

[JM10] J.H. Jeon and R. Metzler. "Fractional Brownian motion and motion governed by the fractional Langevin equation in confined geometries". In: *Phys. Rev. E* 81 (2 2010), p. 021103.

[Jur+05] J. Jurka et al. "Repbase Update, a database of eukaryotic repetitive elements". In: *Cytogentic and Genome Research* 110 (2005), pp. 462–467.

[Jur00] J. Jurka. "Repbase Update: A database and an electronic journal of repetitive elements". In: *Trends in Genetics* 16.9 (2000), pp. 418–420.

[JW94] A. Janicki and A. Weron. "Simulation and chaotic behaviour of α-Stable function". In: *Marcel Dekker,New York* (1994).

[Kam81] N. G. van Kampen. *Stochastic Processes in Physics and Chemistry*. Amsterdam: North-Holland, 1981.

[Kar19] S. Karlin. "A first course in Stochastic processes". In: New York: Academic Press, 19. Chap. No VII Examples of continuous Time Markov Chains, pp. 189–209.

[KBG11] E. Kepten, I. Bronshtein, and Y. Garini. "Ergodicity convergence test suggests telomere motion obeys fractional dynamics". In: *Phys. Rev. E* 83 (4 2011), p. 041919.

[Kev00] J. Kevorkian. *Partial Differential Equations: Analytical Solution Techniques*. Texts in Applied Mathematics 35. Springer New York, 2000. ISBN: 978-1-4419-3139-9,978-1-4757-3266-5.

[KL97] M. G. Kidwell and D. Lisch. "Transposable elements as sources of variation in animals and plants". In: *Proc. Natl. Acad. Sci. USA*. 94(15), 1997, pp. 7704–7711.

[Kra+16] D. Krapf et al. "Strange kinetics of bulk-mediated diffusion on lipid bilayers". In: *Phys. Chem. Chem. Phys.* 18 (18 2016), pp. 12633–12641.

[KRS08] R. Klages, G. Radons, and I. M. Sokolov, eds. *Anomalous Transport: Foundations and Applications*. Weinheim: WILEY–VCH Verlag GmbH & Co. KGaA, 2008.

[KSB06] H. Khan, A. Smit, and S. Boissinot. "Molecular evolution and tempo of amplification of human LINE-1 retrotransposons since the origin of primates". In: *Genome Research* 16.1 (2006), pp. 78–87.

[Kub66] R Kubo. "The fluctuation-dissipation theorem". In: *Reports on Progress in Physics* 29.1 (1966), pp. 255–284.

[Kun+10] G. Kunarso et al. "Transposable elements have rewired the core regulatory network of human embryonic stem cells". In: *Nat Genet* 42.7 (2010), pp. 631–634.

[LC05] A. Le Rouzic and P. Capy. "The first steps of transposable elements invasion: Parasitic strategy vs. genetic drift". In: *Genetics* 169.2 (2005), pp. 1033–1043.

[LC06] A. Le Rouzic and P. Capy. "Population genetics models of competition between transposable element subfamilies". In: *Genetics* 174.2 (2006), pp. 785–793.

[Lin+15] S. Linquist et al. "Applying ecological models to communities of genetic elements: The case of neutral theory". In: *Molecular Ecology* 24.13 (2015), pp. 3232–3242.

[Lut01] E. Lutz. "Fractional Langevin equation". In: *Physical Review E* 64 (5 2001).

[MA13] V. Méndez and A.Iomin. "Comb-like models for transport along spiny dendrites". In: *Chaos, Solitons and Fractals* 53.Supplement C (2013), pp. 46 –51. ISSN: 0960-0779.

[Mag+09] M. Magdziarz et al. "Fractional Brownian motion versus the Continuous-Time Random Walk: A simple test for subdiffusive dynamics". In: *Phys. Rev. Lett.* 103 (2009), p. 180602.

[Mag06] R.L. Magin. *Fractional Calculus in Bioengineering*. Begell House Publishers, 2006. Chap. 8.

[Mai10] F. Mainardi. *Fractional Calculus and Waves in Linear Viscoelasticity*. 1st. Imperial College Press, London, 2010.

[Mai96] F. Mainardi. "Fractional relaxation-oscillation and fractional diffusion-wave phenomena". In: *Chaos, Solitons and Fractals* 7 (1996), pp. 1461–1477.

[Mai97] F. Mainardi. "Fractional Calculus: Some Basic problems in Continuum and Statistical Mechanics". In: *Fractals and Fractional Calculus in Continuum Mechanics*. Ed. by A. Carpinteri and F. Mainardi. CISM Courses and Lecture Notes, Vol. 378. Springer-Verlag, Wien, 1997, pp. 291–348. ISBN: 3-211-82913-X.

[Man+15] C. Manzo et al. "Weak Ergodicity Breaking of Receptor Motion in Living Cells Stemming from Random Diffusivity". In: *Phys. Rev. X* 5 (2015), p. 011021.

[Mar84] O. I. Marichev. "Handbook of Integral Transforms of Higher Transcendental Functions. Theory and Algorithmic Tables." In: *ZAMM - Journal of Applied Mathematics and Mechanics / Zeitschrift für Angewandte Mathematik und Mechanik* 64 (6 1984).

[Mas+14] P. Massignan et al. "Nonergodic subdiffusion from Brownian motion in an inhomogeneous medium". In: *Phys. Rev. Lett.* 112.15 (2014), p. 150603.

[Met+14] R. Metzler et al. "Anomalous diffusion models and their properties: non-stationarity, non-ergodicity, and ageing at the centenary of single particle tracking". In: *Phys. Chem. Chem. Phys.* 16 (2014), p. 24128.

[MG+16] D. Molina-García et al. "Fractional kinetics emerging from ergodicity breaking in random media". In: *Phys. Rev. E* 94 (2016), p. 052147.

[MGP15] C. Manzo and M. F. Garcia-Parajo. "A review of progress in single particle tracking: from methods to biophysical insights". In: *Rep. Progr. Phys.* 78 (2015), p. 124601.

[MH01] M. L. Mears and C. A. Hutchinson. "The evolution of modern lineages of mouse L1 elements". In: *J. Mol. Evol* 52 (2001), pp. 51–62.

[MJC16] R. Metzler, J.H. Jeon, and A.G. Cherstvy. "Non-Brownian diffusion in lipid membranes: Experiments and simulations". In: *Biochimica et Biophysica Acta (BBA) - Biomembranes* 1858.10 (2016). Biosimulations of lipid membranes coupled to experiments, pp. 2451 –2467. ISSN: 0005-2736.

[MK00] R. Metzler and J. Klafter. "The random walk's guide to anomalous diffusion: a fractional dynamics approach". In: *Phys. Rep.* 339.1 (2000), pp. 1–77.

[MLGP10] M. Munoz-López and J. L. García-Pérez. "DNA transposons: nature and applications in genomics." In: *Current genomics* 11.2 (2010), pp. 115–28.

[MLP01] F. Mainardi, Y. Luchko, and G. Pagnini. "The fundamental solution of the space-time fractional diffusion equation". In: *Fract. Calc. Appl. Anal.* 4.2 (2001), pp. 153–192.

[MM09] A. Mura and F. Mainardi. "A class of self-similar stochastic processes with stationary increments to model anomalous diffusion in physics". In: *Integr. Transf. Spec. F.* 20.3–4 (2009), pp. 185–198.

[MMP10] F. Mainardi, A. Mura, and G. Pagnini. "The M-Wright function in time-fractional diffusion processes: A tutorial survey". In: *Int. J. Differ. Equations* 2010 (2010), p. 104505.

[Mon64] E.W. Montroll. "Random walks on lattices". In: *Proc. Symp. Appl. Math., Am. Math. Soc.* 16 (1964), pp. 193–220.

[MP08a] F. Mainardi and P. Pironi. "The Fractional Langevin Equation: Brownian Motion Revisited". In: *Extracta Mathematicae* 10 (1 2008), pp. 140–154.

[MP08b] A. Mura and G. Pagnini. "Characterizations and simulations of a class of stochastic processes to model anomalous diffusion". In: *J. Phys. A: Math. Theor.* 41 (2008), p. 285003.

[MPC15] A. A. Mir, C. Philippe, and G. Cristofari. "euL1db: The European database of L1HS retrotransposon insertions in humans". In: *Nucleic Acids Research* 43 (2015), pp. D43–D47.

[MPS05] F. Mainardi, G. Pagnini, and R. K. Saxena. "Fox H functions in fractional diffusion". In: *J. Comput. Appl. Math.* 178 (2005), pp. 321–331.

[MTM08] A. Mura, M. S. Taqqu, and F. Mainardi. "Non-Markovian diffusion equations and processes: Analysis and simulations". In: *Physica A* 387 (2008), pp. 5033–5064.

[Mur11] A. Mura. *Non-Markovian Stochastic Processes and Their Applications: From Anomalous Diffusion to Time Series Analysis*. Ph.D. Thesis, Physics Department, University of Bologna, 2008. Lambert Academic Publishing, 2011.

[NSS02] E. A. Nimchinsky, B. L. Sabatini, and K. Svoboda. "SRUCTURE AND FUNCTION OF DENDRITIC SPINES". In: *Annual Review of Physiology* 64 (1 Mar. 2002).

[Pag12] G. Pagnini. "Erdélyi–Kober fractional diffusion". In: *Fract. Calc. Appl. Anal.* 15.1 (2012), pp. 117–127.

[Pag13] G. Pagnini. "The M-Wright function as a generalization of the Gaussian density for fractional diffusion processes". In: *Fract. Calc. Appl. Anal.* 16.2 (2013), pp. 436–453.

[Pag14] G. Pagnini. "Short note on the emergence of fractional kinetics". In: *Physica A* 409 (2014), pp. 29–34.

[PCG09] P. Paradisi, R. Cesari, and P. Grigolini. "Superstatistics and renewal critical events". In: *Cent. Eur. J. Phys.* 7 (2009), pp. 421–431.

[PMM12] G. Pagnini, A. Mura, and F. Mainardi. "Generalized fractional master equation for self-similar stochastic processes modelling anomalous diffusion". In: *Int. J. Stoch. Anal.* 2012 (2012), p. 427383.

[PMM13] G. Pagnini, A. Mura, and F. Mainardi. "Two-particle anomalous diffusion: Probability density functions and self-similar stochastic processes". In: *Phil. Trans. R. Soc. A* 371 (2013), p. 20120154.

[Pod99] I. Podlubny. *Fractional Differential Equations*. 1st. Mathematics in Science and Engineering 198. Academic Press, San Diego, 1999.

[PP16] G. Pagnini and P. Paradisi. "A stochastic solution with Gaussian stationary increments of the symmetric space-time fractional diffusion equation". In: *Fract. Calc. Appl. Anal.* 19.2 (2016), pp. 408–440.

[PVF] E. Pascale, E. Valle, and A. V. Furano. In: *A(-ad. Stci. USA* ().

[Reg+13] B. M. Regner et al. "Anomalous Diffusion of Single Particles in Cytoplasm". In: *Biophysical Journal* 104.8 (2013), pp. 1652 –1660. ISSN: 0006-3495.

[Rey03] A. M. Reynolds. "Superstatistical mechanics of tracer-particle motions in turbulence". In: *Phys. Rev. Lett.* 91 (2003), p. 084503.

[Ris89] H. Risken. *The Fokker-Planck Equation. Methods of Solution and Applications*. second. Springer-Verlag, 1989.

[RRM11] R. Rebollo, M. T. Romanish, and D. L. Mager. "Transposable Elements: An Abundant and Natural Source of Regulatory Sequences for Host Genes". In: *Annual Review of Genetics* 46 (2011), p. 1.

[SaB13] A. Sookdeo, C. M. Hepp M. aMcClure, and S. Boissinot. "Revisiting the evolution of mouse LINE-1 in the genomic era." In: *Mobile DNA* 4.1 (2013).

[SAL92] H. Holden S. Albeverio J.E. Fenstad and T. Lindstrom. *W.R. Schneider, "Grey noise." In: 'Ideas and Methods in Mathematical Analysis. Stochastics, and Applications*. Vol. 1. Cambridge University Press, 1992.

[San+04] J. M. Sancho et al. "Diffusion on a solid surface: Anomalous is normal". In: *Physical Review Letters* 92.25 I (2004), pp. 250601–1. ISSN: 00319007.

[San+06] F. Santamaria et al. "Anomalous diffusion in Purkinje cell dendrites caused by spines". In: *Neuron* 52 (4 2006).

[San+11] F. Santamaria et al. "The diffusional properties of dendrites depend on the density of dendritic spines." In: *The European Journal of Neuroscience* (4 2011).

[SBD13] F. Serra, V. Becher, and H. Dopazo. "Neutral Theory Predicts the Relative Abundance and Diversity of Genetic Elements in a Broad Array of Eukaryotic Genomes". In: *PLoS ONE* 8 (2013), p. 6.

[Sel+06] M. C. Seleme et al. "Extensive individual variation in L1 retrotransposition capability contributes to human genetic diversity." In: *Proc. Natl. Acad Sci. U. S. A* 103 (2006), pp. 6611–6616.

[SGM04] E. Scalas, R. Gorenflo, and F. Mainardi. "Uncoupled continuous-time random walks: Solution and limiting behavior of the master equation". In: *Phys. Rev. E* 69 (2004), p. 011107.

[SKR05] C.J. Struchiner, M.G. Kidwell, and J.M.C. Ribeiro. "Population dynamics of transposable elements: copy number regulation and species invasion requirements". In: *Journal of Biological Systems* 13.4 (2005), pp. 455–475.

[Sli+17] O. Sliusarenko et al. "A novel model for anomalous transport in biological systems". In: *Poster at International conference on statistical physics (SIGMAPHI2017)* (2017).

[SM80] M. Suetsugu and P. Mehraein. "Spine distribution along the apical dendrites of the pyramidal neurons in Down's syndrome". In: *Acta Neuropathologica* 50 (3 1980).

[Smi+95] A. F. Smit et al. "Ancestral, mammalian-wide subfamilies of LINE-1 repetitive sequences." In: *Journal of molecular biology* 246.3 (1995), pp. 401–417.

[Spo16] V Sposini. *A numerical study of fractional diffusion through a Langevin approach in random media.* Master Thesis in Physics, University of Bologna, Prof. G. Castellani and G. Pagnini. 2016.

[Sta] B. Stankovìc. "On the function of E.M. Wright". In: *Publ. de l'InstitutMathèmatique, Beograd, Nouvelle Sèr.* 10 (), pp. 113–124.

[SVM17] G. Castellani S. Vitali and F. Mainardi. "Time Fractional Cable Equation And Applications in Neurophysiology." In: *Chaos, Solitons and Fractals* 102 (2017), pp. 467–472.

[SW89] W. R. Schneider and W. Wyss. "Fractional diffusion and wave equations". In: *J. Math. Phys.* 30.1 (1989), pp. 134–144.

[Tam+15] M. V. Tamm et al. "Anomalous Diffusion in Fractal Globules". In: *Phys. Rev. Lett.* 114 (17 2015), p. 178102.

[Tay22] G. I. Taylor. "Diffusion by Continuous Movements". In: *Proceedings of the London Mathematical Society* s2-20.1 (1922), pp. 196–212.

[Tes+12] A. Testori et al. "M. De Bortoli, DCora', and M. Caselle." In: *The role of Transposable Elements in shaping the combinatorial interaction of Transcription Factors BMC genomics* 13.1 (2012).

[TMS14] W. Teka, T.M. Marinov, and F Santamaria. "Neuronal spike timing adaptation described with a fractional leaky integrate-and-fire model." In: *PLOS Computational Biology* 10 (2014).

[TN+04] I. M. Tolić-Nørrelykke et al. "Anomalous diffusion in living yeast cells". In: *Phys. Rev. Lett.* 93 (2004), p. 078102.

[Tuc88] H.C. Tuckwell, ed. *Introduction to Theoretical Neurobiology - Vol.1: Linear Cable Theory and Dendritic Structure*. Cambridge University Press, 1988.

[Tul16] F. di Tullio. *Fractional diffusion in random media on the basis of Gaussian stochastic processes.* Master Thesis in Mathematics, "Roma Tre" University, Prof. R. Spigler and Dr. G. Pagnini. 2016.

[TZ13] J. Tang and S. Zhou. "Hybrid niche-neutral models outperform an otherwise equivalent neutral model for fitting coral reef data". In: *Journal of Theoretical Biology* 317 (2013), pp. 212–218.

[VFB09] S. Venner, C. Feschotte, and C. Biémont. "Dynamics of transposable elements: towards a community ecology of the genome". In: *Trends in Genetics* 25.7 (2009), pp. 317–323.

[VM17] S. Vitali and F. Mainardi. "Fractional Cable Model for Signal Conduction in Spiny Neuronal Dendrites". In: *AIP Conference Proceedings* Accepted (2017).

[Vol+03] I. Volkov et al. "Neutral theory and relative species abundance in ecology." In: *Nature* 424.6952 (2003), pp. 1035–1037.

[Vol06] J. N. Volff. "Turning junk into gold: Domestication of transposable elements and the creation of new genes in eukaryotes". In: *BioEssays* 28.9 (2006), pp. 913–922.

[Wei+01] W. E. I. Wei et al. "Human L1 Retrotransposition:" in: *Society* 21.4 (2001), pp. 1429–1439.

[Wei+11] A. V. Weigel et al. "Ergodic and nonergodic processes coexist in the plasma membrane as observed by single-molecule tracking." In: *Proc. Natl. Acad. Sci. USA* 108.16 (2011), pp. 6438–43.

[Wei96] T.F. Weiss, ed. *Cellular Biophysics, Vol. 2: Electrical Properties*. Bradford Book, 1996.

[Wic+07] T. Wicker et al. "A unified classification system for eukaryotic transposable elements". In: *Nature Reviews Genetics* 8.12 (2007), pp. 973–982.

[WW00] G. Wilk and Z. Włodarczyk. "Interpretation of the nonextensivity parameter q in some applications of Tsallis statistics and Lévy distributions". In: *Phys. Rev. Lett.* 84 (2000), pp. 2770–2773.

[Yue+14] F. Yue et al. "A comparative encyclopedia of DNA elements in the mouse genome." In: *Nature* 515.7527 (2014), pp. 355–64.

[ZDK15] V. Zaburdaev, S. Denisov, and J. Klafter. "Lévy walks". In: *Rev. Mod. Phys.* 87 (2 2015), pp. 483–530.