



**HAL**  
open science

# Statistical modelling for the analysis of molecular diffusion and intracellular trafficking in fluorescence microscopy

Antoine Salomon

► **To cite this version:**

Antoine Salomon. Statistical modelling for the analysis of molecular diffusion and intracellular trafficking in fluorescence microscopy. Signal and Image Processing. Université de Rennes, 2023. English. NNT: 2023URENS125 . tel-04633804

**HAL Id: tel-04633804**

**<https://theses.hal.science/tel-04633804>**

Submitted on 3 Jul 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# THÈSE DE DOCTORAT DE

L'UNIVERSITÉ DE RENNES

ÉCOLE DOCTORALE N° 601

*Mathématiques, Télécommunications, Informatique, Signal, Systèmes,  
Électronique*

Spécialité : *AST – Signal, Image, Vision*

Par

**Antoine J.-F. SALOMON**

**Modélisations statistiques pour l'analyse de la diffusion des mo-  
lécules et du trafic intracellulaire en microscopie de fluorescence**

Thèse présentée et soutenue à Rennes, le 20 décembre 2023

Unité de recherche : Centre Inria de l'Université de Rennes (équipe SAIRPICO)

## Rapporteurs avant soutenance :

Bassam HAJJ                      CRCN CNRS, UMR168-CNRS, Institut Curie  
Jean-Baptiste SIBARITA      Ingénieur de Recherche CNRS, IINS, Université de Bordeaux

## Composition du Jury :

Président :	Olivier DAMERON	Fonction et établissement d'exercice ( <i>à préciser après la soutenance</i> )
Examineurs :	Bassam HAJJ	CRCN CNRS, UMR168-CNRS, Institut Curie
	Jean-Baptiste SIBARITA	Ingénieur de Recherche CNRS, IINS, Université de Bordeaux
	Leila MURESAN	Ingénieure de Recherche, Université de Cambridge
	Ludger JOHANNES	DR INSERM, U1143-INSERM/UMR3666-CNRS, Institut Curie
		Professeur, Université de Rennes
Dir. de thèse :	Charles KERVRANN	DR INRIA, Centre INRIA de l'université de Rennes



# TABLE OF CONTENTS

---

<b>Résumé en français</b>	<b>23</b>
Contexte et défis . . . . .	23
Contribution principale . . . . .	25
Organisation de la thèse . . . . .	26
Publications et communications . . . . .	29
<b>Summary</b>	<b>31</b>
Context and challenges . . . . .	31
Main contribution . . . . .	32
Organization of the thesis . . . . .	34
Publications and communications . . . . .	36
<b>1 Theory, observation and analytics of microscopic particle motion</b>	<b>39</b>
1.1 Stochastic processes and diffusions . . . . .	39
1.1.1 Stochastic process . . . . .	39
1.1.2 Brownian motion . . . . .	40
1.1.3 Diffusion process . . . . .	41
1.1.4 Stochastic Differential Equation (SDE) . . . . .	42
1.1.5 Fractional Brownian motion . . . . .	43
1.2 Particle motion modeling . . . . .	44
1.2.1 Langevin's approach . . . . .	44
1.2.2 Subdiffusion . . . . .	47
1.2.3 Superdiffusion . . . . .	51
1.3 Fluorescence microscopy techniques . . . . .	52
1.3.1 Conventional microscopy techniques . . . . .	53
1.3.2 The diffraction limit problem and its solutions . . . . .	62
1.3.3 Super-resolution microscopy techniques . . . . .	64
1.4 Bioparticle tracking techniques . . . . .	74
1.4.1 Multiple Hypothesis Tracking (MHT) . . . . .	75

TABLE OF CONTENTS

---

1.4.2	The U-TRACK algorithm . . . . .	76
1.5	Particle motion classification . . . . .	77
1.5.1	Particle trajectory classification . . . . .	78
1.5.2	Detection of motion class switching along particle trajectories . . . . .	80
1.5.3	Merging of motion classification and switching detection . . . . .	82
1.6	Simulating fluorescence microscopy particle dynamics with FLUOSIM simulator . . . . .	82
1.6.1	General principle . . . . .	83
1.6.2	Mathematical framework . . . . .	83
1.6.3	Performances . . . . .	85
1.7	Conclusion . . . . .	87
<b>2</b>	<b>Dense mapping of intracellular diffusion and drift from molecule tracks</b>	<b>89</b>
2.1	Related work . . . . .	90
2.1.1	Eulerian method for diffusion estimation . . . . .	90
2.1.2	Bayesian method for diffusion estimation . . . . .	92
2.1.3	Conclusion . . . . .	93
2.2	Dense Mapping: a hybrid Lagrangian-Eulerian method for diffusion and drift estimation . . . . .	94
2.2.1	Notations and definitions . . . . .	94
2.2.2	Dense diffusion and drift estimation . . . . .	96
2.2.3	Pseudo-code of the method . . . . .	98
2.3	Experimental evaluation of the algorithm . . . . .	101
2.3.1	Evaluation on simulated data sets . . . . .	102
2.3.2	Evaluation on real data sets . . . . .	113
2.4	Conclusion . . . . .	126
<b>3</b>	<b>Micro-domains and Confinements Domains Detection</b>	<b>133</b>
3.1	Introduction . . . . .	133
3.2	Notations and mathematical framework . . . . .	135
3.3	Most related work . . . . .	137
3.4	Modeling and simulation of trapped particles . . . . .	140
3.4.1	Modeling of particle dynamics . . . . .	140
3.4.2	Simulations of trajectories and microdomains . . . . .	142
3.5	Clustering of trajectories for micro-domain detection . . . . .	145

3.5.1	Classification of trajectories . . . . .	146
3.5.2	Representative trajectory points and spatial distribution . . . . .	148
3.5.3	A clustering algorithm: DBSCAN . . . . .	149
3.5.4	Detection of trapping areas . . . . .	152
3.6	Experimental results . . . . .	152
3.6.1	Evaluation on FLUOSIM simulated sequences . . . . .	152
3.6.2	Experiments on real TIRF images: dynamics of Rab11 involved in exocytosis . . . . .	156
3.7	Conclusion and discussion . . . . .	158
<b>Conclusion</b>		<b>161</b>
	Contributions of the thesis . . . . .	161
	Future work and extensions . . . . .	162
<b>A Appendix</b>		<b>165</b>
A.1	Derivation of the global binding and unbinding rates . . . . .	165
A.2	Additional results: langerin in RPE-1 cells . . . . .	166
<b>Bibliography</b>		<b>169</b>
<b>Abbreviations</b>		<b>181</b>



# LIST OF FIGURES

---

1	Différents types de mouvements de particules sont représentés. <b>(a)</b> La particule se déplace par diffusion libre. Ce type de mouvement résulte de collisions aléatoires avec les molécules du fluide environnant. <b>(b)</b> La particule se déplace par sous-diffusion. Ce phénomène est dû au fait que la particule est piégée dans un microdomaine. <b>(c)</b> La particule se déplace par superdiffusion. Les principales forces impliquées dans le mouvement de la particule sont déterministes. . . . .	25
2	Notre algorithme est capable de produire des cartes de la diffusion $\sigma$ et de la dérive $\mu$ au sein d'une cellule aussi bien en 2D et en 3D : <b>(a)</b> le paramètre de diffusion de protéines Gal-3 au sein d'une cellule SUM159 est représenté sous la forme d'un nuage de points en 3D ; <b>(b)</b> le paramètre de dérive de facteurs de transcription au sein d'une cellule est représenté en 2D sous la forme d'un nuage de vecteurs, eux-mêmes représentés sous forme de points dont la teinte et la saturation représentent respectivement la direction et l'amplitude. . . . .	27
1	Different types of particle motion are shown. <b>(a)</b> The particle undergoes free diffusion. This kind of motion results from random collisions with the surrounding fluid molecules. <b>(b)</b> The particle undergoes subdiffusion. This results from the particle being trapped in a microdomain. <b>(c)</b> The particle undergoes superdiffusion. The main forces involved in the particle's motion are deterministic. . . . .	33
2	Our algorithm is able to produce maps of diffusion $\sigma$ and drift $\mu$ within a cell in both 2D and 3D: <b>(a)</b> the diffusion parameter of Gal-3 proteins within a SUM159 cell is represented in the form of a 3D point cloud; <b>(b)</b> the drift parameter of transcription factors within a cell is represented in 2D as a vector cloud, which is itself represented as points whose hue and saturation represent direction and amplitude, respectively. . . . .	34

LIST OF FIGURES

---

1.1 Principle and light path of Wide-Field Fluorescence Microscopy. The excitation light is focused in the center of the pupil plane. (illustration extracted and edited from [Schermelleh et al. 2019], Fig. 1(a)) . . . . . 54

1.2 Principle and light path of confocal microscopy. The light beam fills the pupil plane, and is focused and scanned across the object plane. (illustration extracted and edited from [Schermelleh et al. 2019], Fig. 1(c)) . . . . . 55

1.3 Illustration of Snell’s law. (illustration available on Wikimedia Commons under the filename 'Snells law2.svg ', created and released in public domain by contributor Sawims) . . . . . 56

1.4 The intensity of the evanescent wave generated at the interface between two glass and aqueous phases decays exponentially with distance from the glass phase. (illustration extracted and edited from [Axelrod and Davidson n.d.], available on [www.olympus-lifescience.com](http://www.olympus-lifescience.com)) . . . . . 56

1.5 The evanescent intensities of polarized light relative to the incident intensity, as functions of the angle of incidence. Light passes through an interface composed of fused quartz and water (or an aqueous buffer solution) of refractive indexes  $n(1) = 1.46$  and  $n(2) = 1.33$  respectively, with a critical angle of  $\theta_c = 65.7^\circ$ . P and s-polarized evanescent intensities can be several times greater than the incident intensity for angles within 15 degrees of the critical angle. (illustration extracted and edited from [Axelrod and Davidson n.d.], available on [www.olympus-lifescience.com](http://www.olympus-lifescience.com)) . . . . . 57

1.6 Principle and light path of TIRF Microscopy. By focusing the excitation beam on the edge of the pupil plane, the light strikes the coverslip interface at a supercritical angle, generating an evanescent wave. (illustration extracted and edited from [Schermelleh et al. 2019], Fig. 1(b)) . . . . . 58

1.7 Principle and implementations of Light Sheet Fluorescence Microscopy. **(a)** A Selective/Single Plane Illumination Microscope (SPIM). A cylindrical lens focuses the laser beam along the z axis, creating a static light sheet that illuminates the field of view. **(b)** A Digital Scanned Laser Microscope (DSLM). A mirror/lens assembly focuses the laser beam into a single spot that is scanned across the volume to create a dynamic light sheet. (illustration extracted from [Stelzer et al. 2021]) . . . . . 60

1.8 3D image generation with LSFM. **(a)** By moving the specimen or light sheet relative to each other along the z axis, it is possible to acquire a stack of 2D images plane-by-plane, that together form a 3D image. **(b)** When the specimen is too large or dense to be penetrated by the laser beam, multidirectional imaging can be performed using a rotary stage that rotates the specimen around the y axis, allowing 3D image stacks to be acquired in multiple directions. **(c)** By moving the light sheet along the x axis and combining multiple acquisitions into a single image, a constant axial resolution can be achieved over a large field of view. (illustration extracted from [Stelzer et al. 2021]) . . . . . 61

1.9 The Airy pattern generated by a fluorophore through a perfect optical system, with  $\lambda = 509$  nm (GFP emission wavelength) and  $NA = 1.5$ .  $I$  designates the signal intensity, and  $I_0$  is its maximum at the center of the pattern. **(a)** Lateral view of the Airy pattern, as seen by the observer. Logarithmic scale is used to make the concentric ring more visible. In practice, the central Airy disk concentrates the  $\sim 84\%$  of the visible signal. **(b)** Relative intensity values of the radial cross-section of the Airy pattern represented in **(a)**. (illustration plotted according to the Point-Spread Function (PSF) equation in [Zhang, Zerubia, and Olivo-Marin 2007]) . . . 62

1.10 Simulation of the PSF generated by a microscope with  $\lambda = 557$  nm and  $NA = 1.3$ , as observed in the axial plane. To improve the visualization of the diffraction rings, the image was contrasted with a factor  $\gamma = 0.5$ . (illustration extracted and edited from [Cole, Jinadasa, and Brown 2011]) . 63

1.11 Principle and light path of Structured Illumination Microscopy. **(a)** Interference-based SIM. **(b)** Point-scanning SIM. (illustration extracted and edited from [Schermelleh et al. 2019], Fig. 1(d) and (e)) . . . . . 65

1.12 Comparison of 3D images of red-fluorescent microspheres of diameter  $0.12\ \mu\text{m}$ , illustrating the resolution and contrast improvement between WFFM and Structured Illumination Microscopy. **(a)** Lateral view of the microspheres obtained through WFFM. **(b)** Lateral view of the microspheres obtained through SIM. **(c)** Axial view of the microspheres obtained through WFFM. **(d)** Axial view of the microspheres obtained through SIM. **(e)** Histograms of lateral and axial Full Width at Half-Maximum (FWHM) of green-fluorescent beads observed by 3D SIM. (illustration extracted from [M. G. Gustafsson et al. 2008]) . . . . . 66

1.13 Principle and light path of Lattice Light Sheet Microscopy. (illustration extracted and edited from [Schermelleh et al. 2019], Fig. 1(h)) . . . . . 67

1.14 Comparison of light sheet generation methods. **(a)** The conventional LSFM approach sweeps a Gaussian beam across a plane to create the light sheet. **(b)** The use of a Bessel beam generates a sheet with a narrow core, albeit surrounded by a large number of bands. **(c and d)** Optical lattices create periodic interference patterns across the plane, with low peak intensity and low phototoxicity. The square lattice in **(c)** confines the excitation in the central plane, while the hexagonal lattice in **(d)** optimizes axial resolution. The columns from left to right depict: the intensity pattern at the rear pupil plane of the excitation objective (in arbitrary units); the intensity pattern at the xz plane of the excitation objective focus; the dithered or swept focal intensity pattern creating the light sheet; and the axial cross-section of the microscope’s overall PSF, respectively. (illustration extracted and edited from [Chen et al. 2014]) . . . . . 68

1.15 Principle and light path of Stimulated Emission Depletion microscopy. (illustration extracted and edited from [Schermelleh et al. 2019], Fig. 1(f)) . . . . . 70

1.16 Actin filaments in cells stained with 647-phalloidin. Stimulated emission depletion (STED) microscopy provides significant resolution improvements over those possible with confocal microscopy. (illustration created and released in CC-BY-SA-4.0 by Howard Vindin, available on Wikimedia Commons under the filename 'STED Confocal Comparison 50nm HWFM.png') 71

1.17 Principle of Single Molecule Localization Microscopy. (illustration extracted and edited from [Schermelleh et al. 2019], Fig. 1(g)) . . . . . 72

1.18	E. coli $\Delta$ tar cell with mEos-labeled Tar. (A), (B), (C) and (D) depict Differential interference contrast (DIC), Diffraction-limited epi-fluorescence (epi), TIR-PALM, and epi-PALM images of a single cell, respectively. (E) Superposition of (C) and (D). (F), (G) and (H) depict zooms on the left, middle and right boxed ROIs of (E), respectively. The scale bar in (A–E) indicates $1\ \mu\text{m}$ and the scale bar in (F–H) indicates $50\ \text{nm}$ . (illustration extracted from [Greenfield et al. 2009]) . . . . .	73
1.19	The fluosim interface and its parameters. . . . .	84
1.20	Comparison between SPT experiments and their corresponding FLUOSIM simulations. <b>(a)</b> Image sequence showing a COS-7 cell expressing GFP-Nrx1 $\beta$ sparsely labeled with Atto647N-conjugated GFP nanobody. <b>(b)</b> Zoom on the ROI in (a) (white rectangle) showing the trajectories tracked after image acquisition. <b>(c)</b> FLUOSIM simulation of particles in a geometry similar to the experiment illustrated in (a). <b>(d)</b> Zoom on the ROI in (c) representing trajectories simulated with FLUOSIM. (extracted and edited from [Lagardère et al. 2020]) . . . . .	86
2.1	The 2D simulated fluorescence microscopy video is generated with FLUOSIM. Real microscopy data are used to shape the cell into a crossbow. The sequence has a length of 400 frames, a time step of $0.005\ \text{s}$ , a spatial resolution of $0.16\ \mu\text{m}$ and includes 2500 moving particles. <b>(a)</b> Ground truth diffusion map is shown. The cell contains 20 trapping regions shown in blue where the diffusion coefficient is set to $0.04\ \mu\text{m}^2/\text{s}$ . The rest of the cell has a diffusion coefficient set to $0.16\ \mu\text{m}^2/\text{s}$ . <b>(b)</b> Raw video output is obtained from FLUOSIM. <b>(c)</b> Augmented video output is derived from (b). A fake grey background was added to the cell, in addition to Gaussian noise. A linear decrease of brightness was added to simulate photobleaching. . . . .	101
2.2	The ground truth tracks are directly extracted from the FLUOSIM-simulated data set and analyzed. <b>(a)</b> The set of all simulated particle trajectories is shown. <b>(b)</b> Particle trajectories are classified by motion type. <b>Free diffusion</b> is represented in <b>blue</b> , <b>subdiffusion</b> is represented in <b>green</b> and <b>superdiffusion</b> is represented in <b>red</b> . <b>(c)</b> The histogram of the input trajectory lengths is shown. . . . .	104

2.3 The results obtained with our method on simulated ground truth tracks are shown. **(a)** The point cloud output represents the spatial distribution of the diffusion coefficient inside the cell, in  $\mu\text{m}^2/\text{s}$ . **(b)** The point cloud is derived into a map by interpolation, with the same scale units. **(c)** The point cloud is derived into a diffusion histogram. The values displayed on top of the modes are also in  $\mu\text{m}^2/\text{s}$  . . . . . 105

2.4 The results obtained with method [Hozé, Nair, et al. 2012] on simulated ground truth tracks are shown. **(a-c)** Diffusion maps are generated using a coarse resolution parameter of  $1.2\mu\text{m}$ , with from left to right: the raw map, the Gaussian-filtered ( $\sigma = 0.8$ ) map, the Gaussian-filtered and interpolated diffusion map. **(d-f)** Diffusion maps are generated using a medium resolution parameter of  $0.7\mu\text{m}$ , and shown in the same order. **(g-i)** Diffusion maps are generated using a fine resolution parameter of  $0.2\mu\text{m}$ , and shown in the same order. The diffusion scale units are in  $\mu\text{m}^2/\text{s}$  . . . . . 106

2.5 The results obtained with method [Beheiry, Dahan, and Masson 2015] on simulated ground truth tracks are shown. **(a-c)** Diffusion maps are generated with the Voronoi tessellation mode, using respectively from left to right: the minimum number of zones (4700), the default number of zones (9401), and the maximum number of zones (14102). **(d)** A diffusion map is generated with the Quad-Tree mesh mode. The diffusion scale units are in  $\mu\text{m}^2/\text{s}$ . . . . . 107

2.6 The particles are tracked with ICY from the simulated video and analyzed. **(a)** The set of all simulated particle trajectories is shown. **(b)** Particle trajectories are classified by motion type. **Free diffusion** is represented in blue, **subdiffusion** is represented in green and **superdiffusion** is represented in red. **(c)** The histogram of the input trajectory lengths is shown. . . . . 109

2.7 The results obtained with our method on the data tracked with ICY from the FLUOSIM-simulated video, are shown. **(a)** The point cloud output represents the spatial distribution of the diffusion coefficient inside the cell, in  $\mu\text{m}^2/\text{s}$ . **(b)** The point cloud is derived into a map by interpolation, with the same scale units. **(c)** The point cloud is derived into a diffusion histogram. The values displayed on top of the modes are also in  $\mu\text{m}^2/\text{s}$  . . . . . 110

2.8 The results obtained with method [Hozé, Nair, et al. 2012] on the data tracked with ICY from the FLUOSIM-simulated video, are shown. **(a-c)** Diffusion maps are generated using a coarse resolution parameter of  $1.2\ \mu\text{m}$ , with from left to right: the raw map, the Gaussian-filtered ( $\sigma = 0.8$ ) map, the Gaussian-filtered and interpolated diffusion map. **(d-f)** Diffusion maps are generated using a medium resolution parameter of  $0.7\ \mu\text{m}$ , and shown in the same order. **(g-i)** Diffusion maps are generated using a fine resolution parameter of  $0.2\ \mu\text{m}$ , and shown in the same order. The diffusion scale units are in  $\mu\text{m}^2/\text{s}$  . . . . . 111

2.9 The results obtained with method [Beheiry, Dahan, and Masson 2015] on the data tracked with ICY from the FLUOSIM-simulated video, are shown. **(a-c)** Diffusion maps are generated with the Voronoi tessellation mode, using respectively from left to right: the minimum number of zones (611), the default number of zones (1223), and the maximum number of zones (1834). **(d)** A diffusion map is generated with the Quad-Tree mesh mode. The diffusion scale units are in  $\mu\text{m}^2/\text{s}$ . . . . . 112

2.10 The tracking data obtained from the transcription factors data set is analyzed. **(a)** The set of all simulated particle trajectories is shown. **(b)** Particle trajectories are classified by motion type. **Free diffusion** is represented in **blue**, **subdiffusion** is represented in **green** and **superdiffusion** is represented in **red**. **(c)** The histogram of the input trajectory lengths is shown. . . . . 115

2.11 The results obtained with our method on real data (transcription factors), are shown. **(a)** The point cloud output represents the spatial distribution of the diffusion coefficient inside the cell, in  $\mu\text{m}^2/\text{s}$ . **(b)** The point cloud is derived into a map by interpolation, with the same scale units. **(c)** The point cloud is derived into a diffusion histogram. The values displayed on top of the modes are also in  $\mu\text{m}^2/\text{s}$ . **(d)** This second point cloud represents the spatial distribution of the drift inside the cell. Each point represents a drift vector. The wheel indicates both the direction and the amplitude of the vectors. . . . . 116

2.12 The results obtained with method [Hozé, Nair, et al. 2012] on real data (transcription factors), are shown. **(a-c)** Diffusion maps are generated using a coarse resolution parameter of  $1.2\ \mu\text{m}$ , with from left to right: the raw map, the Gaussian-filtered ( $\sigma = 0.8$ ) map, the Gaussian-filtered and interpolated diffusion map. **(d-f)** Diffusion maps are generated using a medium resolution parameter of  $0.7\ \mu\text{m}$ , and shown in the same order. **(g-i)** Diffusion maps are generated using a fine resolution parameter of  $0.2\ \mu\text{m}$ , and shown in the same order. The diffusion scale units are in  $\mu\text{m}^2/\text{s}$  . . . . . 117

2.13 The results obtained with method [Beheiry, Dahan, and Masson 2015] on real data (transcription factors), are shown. **(a-c)** Diffusion maps are generated with the Voronoi tessellation mode, using respectively from left to right: the minimum number of zones (86), the default number of zones (172), and the maximum number of zones (258). **(d)** A diffusion map is generated with the Quad-Tree mesh mode. The diffusion scale units are in  $\mu\text{m}^2/\text{s}$ . . . . . 118

2.14 The tracking data obtained from the mEOS2-calnexin plasmids in HEK-293 cells data set is analyzed. **(a)** The set of all simulated particle trajectories is shown. **(b)** Particle trajectories are classified by motion type. **Free diffusion** is represented in **blue**, **subdiffusion** is represented in **green** and **superdiffusion** is represented in **red**. **(c)** The histogram of the input trajectory lengths is shown. . . . . 119

2.15 The results obtained with our method on real data (mEOS2-calnexin plasmids in HEK-293 cells), are shown. **(a)** The point cloud output represents the spatial distribution of the diffusion coefficient inside the cell, in  $\mu\text{m}^2/\text{s}$ . **(b)** The point cloud is derived into a map by interpolation, with the same scale units. **(c)** The point cloud is derived into a diffusion histogram. The values displayed on top of the modes are also in  $\mu\text{m}^2/\text{s}$  . . . . . 120

2.16 The results obtained with method [Hozé, Nair, et al. 2012] on real data (mEOS2-calnexin plasmids in HEK-293 cells), are shown. Diffusion maps are generated using a resolution parameter of  $0.4\ \mu\text{m}$ , with from left to right: **(a)** the raw map, **(b)** the Gaussian-filtered ( $\sigma = 0.8$ ) map, **(c)** the Gaussian-filtered and interpolated diffusion map. The diffusion scale units are in  $\mu\text{m}^2/\text{s}$  . . . . . 121

---

2.17	The tracking data obtained from the TMR-labelled Halo-tagged CRT in HEK-293 cells data set is analyzed. <b>(a)</b> The set of all simulated particle trajectories is shown. <b>(b)</b> Particle trajectories are classified by motion type. <b>Free diffusion</b> is represented in <b>blue</b> , <b>subdiffusion</b> is represented in <b>green</b> and <b>superdiffusion</b> is represented in <b>red</b> . <b>(c)</b> The histogram of the input trajectory lengths is shown. . . . .	122
2.18	The results obtained with our method on real data (TMR-labelled Halo-tagged CRT in HEK-293 cells), are shown. <b>(a)</b> The point cloud output represents the spatial distribution of the diffusion coefficient inside the cell, in $\mu\text{m}^2/\text{s}$ . <b>(b)</b> The point cloud is derived into a map by interpolation, with the same scale units. <b>(c)</b> The point cloud is derived into a diffusion histogram. The values displayed on top of the modes are also in $\mu\text{m}^2/\text{s}$ . . .	123
2.19	The results obtained with method [Hozé, Nair, et al. 2012] on real data (TMR-labelled Halo-tagged CRT in HEK-293 cells), are shown. Diffusion maps are generated using a resolution parameter of $0.2\mu\text{m}$ , with from left to right: <b>(a)</b> the raw map, <b>(b)</b> the Gaussian-filtered ( $\sigma = 0.8$ ) map, <b>(c)</b> the Gaussian-filtered and interpolated diffusion map. The diffusion scale units are in $\mu\text{m}^2/\text{s}$ . . . . .	124
2.20	The 3D tracking data obtained from the Galectin-3 in SUM159 cell data set is analyzed. <b>(a)</b> The set of all simulated particle trajectories is shown. <b>(b)</b> Particle trajectories are classified by motion type. <b>Free diffusion</b> is represented in <b>blue</b> , <b>subdiffusion</b> is represented in <b>green</b> and <b>superdiffusion</b> is represented in <b>red</b> . . . . .	125
2.21	The length histogram derived from the trajectories displayed in Figure 2.20(a) is shown. . . . .	126
2.22	The results obtained with our method on real 3D data (Galectin-3 in SUM159 cell), are shown. <b>(a)</b> The 3D point cloud output represents the spatial distribution of the diffusion coefficient inside the cell, in $\mu\text{m}^2/\text{s}$ . <b>(b)</b> This 3D vector field represents the spatial distribution of the drift inside the cell. <b>Green vectors</b> correspond to <b>subdiffusion-related</b> drift whereas <b>red vectors</b> correspond to <b>superdiffusion-related</b> drift. . . . .	127
2.23	The diffusion histogram derived from the point cloud displayed in Figure 2.22(a) is shown. The values displayed on top of the modes are in $\mu\text{m}^2/\text{s}$ . . .	128

2.24 The results obtained with our method on real 3D data (CD-Mannose-6 Phosphate (CD-M6PR) receptor in HeLa cell), are shown. **(a)** The trajectories are represented inside a 3D view of deconvoluted fluorescence microscopy data. **(b)** The 3D point cloud output represents the spatial distribution of the diffusion coefficient inside the cell, in  $\mu\text{m}^2/\text{s}$ . **(c)** Particle trajectories are classified by motion type. **Free diffusion** is represented in blue, **subdiffusion** is represented in green and **superdiffusion** is represented in red. **(d)** The histogram of the input trajectory lengths is shown. . . . . 129

2.25 The results obtained with our method on real data (HIV-1 Gag in CD4 T cell), are shown as an evolving diffusion videomap of 9 frames, with a time interval overlapping of  $2/3$  between each frame. **(a)** to **(i)** The point cloud output represents the evolution of the spatial distribution of the diffusion coefficient inside the cell, in  $\mu\text{m}^2/\text{s}$ , from  $t = 0\text{ s}$  to  $t = 80\text{ s}$ . . . . . 130

2.26 The results obtained with our method on real data (HIV-1 Gag in CD4 T cell), are shown as an evolving drift videomap of 9 frames, with a time interval overlapping of  $2/3$  between each frame. **(a)** to **(i)** This second point cloud represents the evolution of the spatial distribution of the drift inside the cell, from  $t = 0\text{ s}$  to  $t = 80\text{ s}$ . Each point represents a drift vector, whose direction and amplitude are respectively represented by hue and saturation. **(q)** The hue-saturation colorwheel is used to interpret the color representation of drift vectors. . . . . 131

3.1 Illustrations of configurations with trapping areas. **(a)** Sketch with three trapping areas ( $\mathcal{S} = \mathcal{S}_1 \cup \mathcal{S}_2 \cup \mathcal{S}_3$ ); **(b)** Example of a simulated image at a given time  $t$ , depicting a crossbow-shaped cell and green particles in small trapping (blue) areas and in the cytosol  $\bar{\mathcal{S}}$  (orange). **(c)** ROI of **(b)**. . . . . 135

3.2 Drift field computed with the method of [Hozé, Nair, et al. 2012]. The FLUOSIM simulation comprises 30 frames of  $200 \times 200\text{ px}$  as the domain  $\mathcal{D}$  is a square of  $5\text{ }\mu\text{m}$  and the spatial resolution is  $\Delta x = 0.025\text{ }\mu\text{m}$  (see table 3.1). We compute  $\hat{\mu}(x)$  on each point of this  $200 \times 200$  lattice from trajectories **(a)** ( $r = 5\text{ px}$ ). We can clearly see the two microdomains  $\mathcal{S}_1$  and  $\mathcal{S}_2$  **(b)**: the vector field converges to the centers of these domains delineated by two red circles **(b)**. We can see that the vector field is orthogonal to the boundaries of the square  $\mathcal{D}$ . It is due the normal boundary condition on  $\partial\mathcal{D}$ . Elsewhere, the  $\hat{\mu}(x)$  is almost null, reflecting Brownian motion. . . . . 138

---

3.3	Positions of the particles simulated with FLUOSIM at time $t = 10$ s and $t = 100$ s (free particles are represented in blue, trapped particles are represented in red). The red circles delineate the regions $\mathcal{S}_1$ (large disk) and $\mathcal{S}_2$ (small disk). . . . .	143
3.4	Evolution of the proportions of trapped particles over time. (a) proportions of trapped particles in $\mathcal{D}$ ; (b) proportions of trapped particles in the trapping region $\mathcal{S}$ . The true proportion of trapped particles computed from the data, the proportions computed with model 3.5, and the asymptote of the model 3.5 are represented in red, blue, and green, respectively. We computed the red curve over $m = 2500$ trajectories simulated with FLUOSIM (see table 3.1). . . . .	144
3.5	Overview of our computational approach for detecting trapping areas. (a) input trajectories; (b) classification of trajectories into three diffusion groups (superdiffusion, subdiffusion, free diffusion); (c) representative points on trajectories; (d) display of subdiffusive representative points; (e) estimation of clusters associated to trapping areas; (f) detection of trapping areas.	145
3.6	Spatial distribution of the trapped particles in the FLUOSIM simulation; (a) the trajectories are represented by their mean point; (b) the trajectories are represented by $\mathbf{x}_{\lfloor n/2 \rfloor}$ . A trajectory is considered to be trapped if it is trapped during the whole period of observation. The red circles represent the boundaries of the trapping regions $\mathcal{S}_1$ (large circle) and $\mathcal{S}_2$ (small circle).	148
3.7	Clusters detected by DBSCAN ( $\kappa = 4$ ) on the set of particles detected as subdiffusive with the test procedure [Briane, Kervrann, and Vimond 2018] (single test with a significance level of 5%). We choose $\epsilon = 5.86$ derived with the Beta distribution assumption. The noisy points are in red, and points belonging to the same cluster are labelled with the same color (blue or green). The red circles represent the boundaries of the trapping regions $\mathcal{S}_1$ (large circle) and $\mathcal{S}_2$ (small circle). The black square delineates the borders of the whole domain $\mathcal{D}$ . . . . .	152
3.8	Simulated crossbow-shaped cell with seventeen trapping regions of different sizes and shapes. The white line at the bottom accounts for $16 \mu\text{m}$ . . . . .	153

3.9 Results of our procedure on the first crossbow-shaped cell simulation; (a) representative points of the trajectories detected as subdiffusive by the test procedure [Briane, Kervrann, and Vimond 2018] with a significance level of 5%; (b) trapping regions (yellow background, red boundaries) estimated with the convex hull from the clusters detected by DBSCAN. The true trapping regions are delineated with blue curves. . . . . 155

3.10 Results of our procedure on the second crossbow-shaped cell simulation; (a) representative points of the trajectories detected as subdiffusive by the test procedure [Briane, Kervrann, and Vimond 2018] with a significance level of 5%; (b) trapping regions (yellow background, red boundaries) estimated with the convex hull from the clusters detected by DBSCAN. The true trapping regions are delineated with blue curves. . . . . 157

3.11 Rab11-mCherry proteins within a M10 cell treated with Latrunculin and observed at several time points. The images are extracted from a 2D TIRFM sequence composed of 600 images (exposure time  $\Delta t = 100$  ms,  $dx = dy = 160$  nm). . . . . 158

3.12 Analysis and clustering of trajectories; (a) trajectory classification into three diffusion groups [Briane, Kervrann, and Vimond 2018]; (b) selection/clustering of subdiffusive trajectories: the point clusters are labeled with colors while individual subdiffusive trajectory are represented with black dots; (c) estimated diffusion coefficient map generated with our method presented in Chapter 2. . . . . 159

A.1 The tracking data obtained from the langerin in crossbow-shaped cells data set is analyzed. (a) The set of all simulated particle trajectories is shown. (b) Particle trajectories are classified by motion type. Free diffusion is represented in blue, subdiffusion is represented in green and superdiffusion is represented in red. (c) The histogram of the input trajectory lengths is shown. . . . . 167

A.2 The results obtained with our method on real data (langerin in crossbow-shaped cells), are shown. **(a)** The point cloud output represents the spatial distribution of the diffusion coefficient inside the cell, in  $\mu\text{m}^2/\text{s}$ . **(b)** The point cloud is derived into a map by interpolation, with the same scale units. **(c)** The point cloud is derived into a diffusion histogram. The values displayed on top of the modes are also in  $\mu\text{m}^2/\text{s}$  **(d)** This second point cloud represents the spatial distribution of the drift inside the cell. Each point represents a drift vector. The wheel indicates both the direction and the amplitude of the vectors. . . . . 168



# LIST OF TABLES

---

3.1	Parameters of the first FLUOSIM simulation. . . . .	143
3.2	Values of the parameter $\epsilon$ of DBSCAN obtained with different methods. The parameter $\epsilon$ is estimated assuming $\kappa = 4$ as recommended in [Ester et al. 1996]. The parameter $\epsilon$ is expressed in pixel units. The image size is $200 \times 200$ px. . . . .	151
3.3	FLUOSIM parameters to simulate the first crossbow-shaped cell. . . . .	153
3.4	Evaluation of the procedure on the crossbow-shaped cell simulation. . . . .	154
3.5	Proportion of subdiffusive trajectories inside the detected regions from the crossbow-shaped cell simulation. . . . .	154
3.6	FLUOSIM parameters to simulate the second crossbow-shaped cell. . . . .	155
3.7	Evaluation of the procedure on the second simulated crossbow-shaped cell. . . . .	156
3.8	Proportion of subdiffusive trajectories inside the detected regions from the crossbow-shaped cell simulation (2) to be compared ( $k^+/(k^+ + k^-) = 0.9091$ ). . . . .	156



# RÉSUMÉ EN FRANÇAIS

---

## Contexte et défis

Au cours des deux dernières décennies, de nombreuses technologies révolutionnaires ont vu le jour et ont permis la visualisation de tissus, de cellules, de protéines, de virus et de structures macromoléculaires à tous les niveaux de résolution spatiale (de 10 nm à 150 nm). La découverte des sondes de marquage par fluorescence (protéine fluorescente verte [GFP], prix Nobel de chimie 2008) et les récentes avancées en matière d'optique et de capteurs numériques (par exemple, PALM, STED et SIM) ont été des éléments clés ayant permis de dépasser la limite théorique de diffraction optique de 200 nm établie au 19<sup>ème</sup> siècle. En raison de ces percées technologiques et de leurs impacts dans les sciences de la vie, la microscopie optique a été saluée par des prix prestigieux, tels que le prix Nobel décernés aux inventeurs des concepts de microscopie à super-résolution (Eric Betzig, Stefan W. Hell et William E. Moerner en 2014). L'imagerie par microscopie de fluorescence est depuis devenue le fer de lance de la biologie moderne. Elle permet de générer des vidéos 3D+Temps de plusieurs dizaines de gigaoctets de données, décrivant des comportements cellulaires à long terme, à l'échelle nanométrique et avec une faible phototoxicité. Cette capacité à pouvoir suivre des événements cellulaires à l'échelle nanométrique s'avère d'une immense pertinence dans le domaine de la santé, notamment pour l'étude de la progression des cancers et infections virales.

De manière générale, les systèmes de microscopie de fluorescence enregistrent les signaux fluorescents émis par des molécules marquées par des protéines génétiquement modifiées à l'intérieur des cellules. Dans une configuration classique, les photons sont collectés et enregistrés au niveau d'un pixel donné (ou voxel en imagerie 3D). L'intensité de fluorescence mesurée est une valeur scalaire, généralement proportionnelle à la densité de molécules marquées représentant quelques dizaines de nanomètres dans un pixel/voxel. Les données d'imagerie de fluorescence étant des signaux en 2D-3D+Temps pouvant potentiellement représenter plusieurs espèces moléculaires marquées par fluorescence (images multicanaux), l'interprétation de ces signaux s'avère bien souvent complexe en traitement et analyse d'images. Plusieurs obstacles scientifiques doivent notamment être surmon-

tés. L'un des plus significatifs concerne la réduction du bruit et du flou observés dans les images. La difficulté augmente dès lors qu'il s'agit de traiter de grandes séries temporelles d'images. Une attention particulière doit alors être accordée à la faisabilité et à l'extensibilité des algorithmes développés. Un autre défi qui fait l'objet de cette thèse est d'estimer des cartes spatiales en 2D-3D et à haute résolution des dynamiques moléculaires à partir des données générées par les instruments de microscopie.

Dans le contexte du transport intracellulaire, on comprend encore très mal l'orchestration de tous les composants cellulaires impliqués dans les voies de transport d'endocytose [Lemaigre et al. 2023] [Johannes and Valades 2021] [Römer et al. 2007] [Renard et al. 2015] [Momboisse et al. 2022] et d'exocytose [Lachuer et al. 2023] [Pécot et al. 2018] [Basset et al. 2017] [Gidon et al. 2012], deux mécanismes respectivement impliqués dans les fonctions et dysfonctionnements dans la cellule tels que la morphogenèse ou dégénérescence des tissus, et l'invasion tumorale. Cette compréhension incomplète des mécanismes de transport n'a longtemps pas pu être améliorée par manque de technologies adaptées, c'est à dire une imagerie quantitative avec la bonne résolution spatio-temporelle, une fréquence d'images appropriée, sans photoblanchiment (diminution de la capacité des marqueurs fluorescents à émettre des photons lors d'excitations répétées), et sans phototoxicité sur l'échantillon vivant. Une première solution a été apportée par l'imagerie par microscopie de fluorescence 2D+temps, en particulier la microscopie TIRF [Axelrod 2003]. Celle-ci a permis des percées significatives, liées aux corrélations potentielles entre les molécules en interaction et la dynamique à la surface des cellules (membrane plasmique). Plus récemment, les progrès constants en biologie moléculaire et en optique ont permis d'imager en 3D+Temps des cellules vivantes. Cela conduit à l'acquisition de très grandes séries d'images : une seule expérience sur des cellules vivantes en microscopie à feuille de lumière en treillis (LLSM) [Chen et al. 2014] peut par exemple produire jusqu'à un téraoctet de données en une heure, avec une résolution spatiale de 100 à 200 nanomètres en 3D. Il est donc devenu nécessaire de développer de nouveaux paradigmes pour la reconstruction d'images, le suivi/estimation de mouvement des molécules en 3D, et l'estimation de paramètres biophysiques pour faire face aux gigantesques séries de volumes acquises avec des systèmes de microscopie très pointus. Afin de relever ces défis, les mathématiques appliquées, le traitement et l'analyse d'images et l'informatique doivent être associées à la biophysique et la biologie, et il est nécessaire d'obtenir une synergie durable entre ces différents domaines scientifiques.

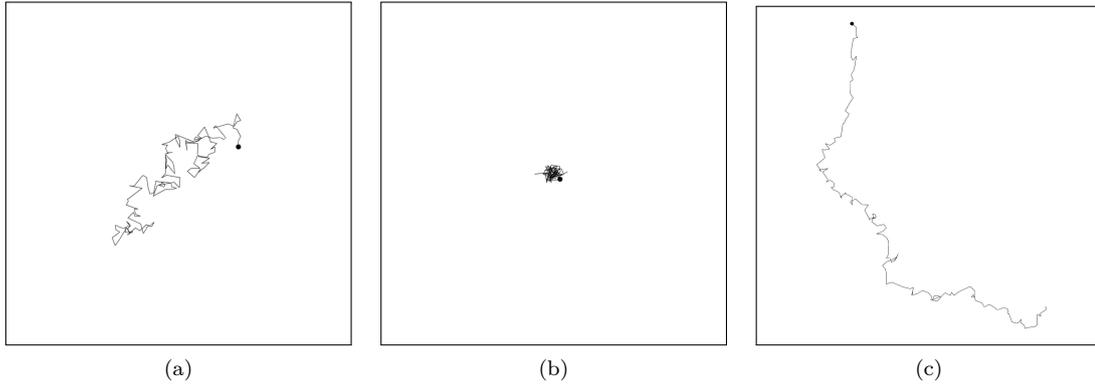


Figure 1 – Différents types de mouvements de particules sont représentés. **(a)** La particule se déplace par diffusion libre. Ce type de mouvement résulte de collisions aléatoires avec les molécules du fluide environnant. **(b)** La particule se déplace par sous-diffusion. Ce phénomène est dû au fait que la particule est piégée dans un microdomaine. **(c)** La particule se déplace par superdiffusion. Les principales forces impliquées dans le mouvement de la particule sont déterministes.

## Contribution principale

Dans cette thèse, nous étudions des méthodes dédiées à l'analyse de dynamiques des molécules à partir de caractéristiques basées sur le mouvement observées dans des images, en particulier les petites trajectoires issues d'algorithmes de suivi de particules. Ces trajectoires sont classées selon trois catégories de mouvement : mouvement brownien, diffusion confinée (sous-diffusion), diffusion dirigée (superdiffusion) (Figure 1). Des tests statistiques sont mis en place pour remplacer les méthodes conventionnelles (comme par exemple le déplacement quadratique moyen (MSD)) qui ne peuvent pas discriminer de manière fiable ces trois types de mouvement. La détection des changements de dynamiques le long de trajectoires longues est réalisée en transposant l'approche précédente sur des fenêtres temporelles glissantes le long des dites trajectoires. La principale contribution de cette thèse est une approche permettant d'estimer la mobilité moléculaire et le transport actif à partir des trajectoires calculées. Pour cela, nous supposons que le mouvement des molécules est régi par l'équation différentielle stochastique suivante :

$$d\mathbf{x} = \boldsymbol{\mu}(\mathbf{x})dt + \boldsymbol{\sigma}(\mathbf{x})d\mathbf{b},$$

où  $\mathbf{x} \in \mathbb{R}^d$ ,  $d \in \{2, 3\}$  désigne la position au temps  $t$  d'une molécule,  $\boldsymbol{\mu} \in \mathbb{R}^d$  est la dérive liée à la vitesse de transport actif appliquée à la particule au point  $\mathbf{x}$ ,  $\boldsymbol{\sigma} \in \mathbb{R}^+$

est le coefficient de diffusion variable dans l'espace, et  $\mathbf{b} \in \mathbb{R}^d$  est un mouvement brownien. Notre méthode est capable de générer des cartes à haute résolution spatiale des paramètres de diffusion  $\sigma(\mathbf{x})$  et de transport actif  $\boldsymbol{\mu}(\mathbf{x})$  à partir de séquences d'images 2D ou 3D (Figure 2). Contrairement aux paradigmes précédents consistant à reconstruire des images en super-résolution (par exemple SPT-PALM/STORM) à partir de plusieurs milliers d'images, nous fournissons directement des cartes à haute résolution spatiale de la mobilité moléculaire, calculées à partir d'un nombre réduit d'images conventionnelles de microscopie de fluorescence. Chaque quantité biophysique est calculée à des positions correspondant exactement aux coordonnées des sous-pixels données par les trajectoires moléculaires, fournissant ainsi des informations plus détaillées.

## Organisation de la thèse

Dans le premier chapitre de ce manuscrit, nous donnons une vue d'ensemble générale de la théorie statistique, des techniques de microscopie et des algorithmes utilisés dans le cadre de nos recherches. Nous commençons par exposer le cadre théorique en rapport avec le concept de processus stochastique, et en particulier le mouvement caractéristique des particules se déplaçant dans un fluide, dit mouvement « brownien ». Après avoir introduit quelques modèles et types de mouvements de particules, nous poursuivons en présentant les techniques de microscopie de fluorescence à la fois classiques et innovantes, en 2D et en 3D, limitées par la diffraction et super-résolues. Nos travaux exploitant les trajectoires individuelles de molécules, nous présentons quelques techniques et algorithmes de suivi. Enfin, nous exposons la méthode de classification du mouvement des particules de Vincent Briane [Briane, Kervrann, and Vimond 2018] [Briane, Vimond, Valades-Cruz, et al. 2020], que nous avons insérées dans notre schéma d'estimation.

Dans le Chapitre 2, nous proposons une nouvelle méthode de cartographie cellulaire visant à estimer de manière robuste les dynamiques intracellulaires (en particulier la diffusion et la dérive). Il est en effet primordial pour les biologistes de pouvoir aisément visualiser les dynamiques des protéines. Par ailleurs, il existe une forte demande d'outils algorithmiques capables d'automatiquement analyser des données de microscopie dans ce but. Après un passage en revue de quelques techniques dédiées à l'estimation de la diffusion et de la dérive, nous présentons en détail les éléments théoriques de notre méthode statistique. Ensuite, nous présentons des estimateurs à noyaux spatio-temporels propres à chacun des mouvements de particules (diffusion libre, sous-diffusion, superdiffusion). Nous

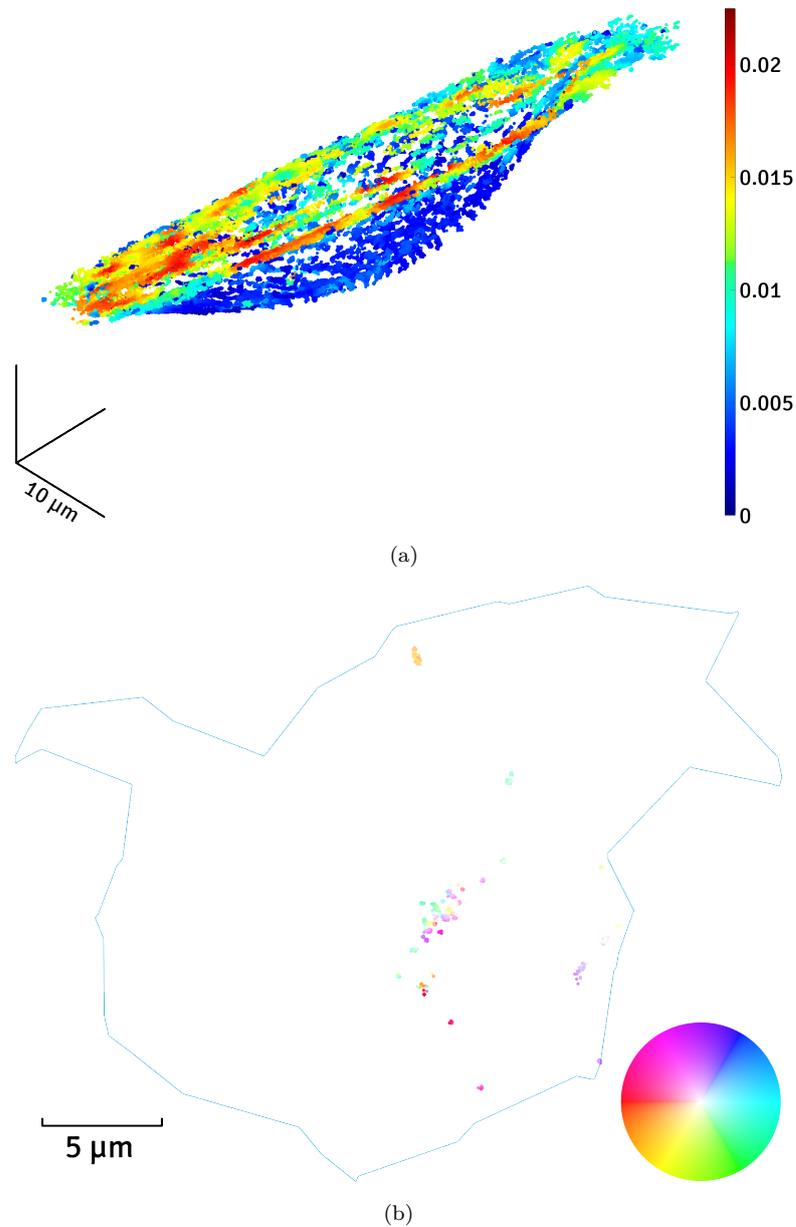


Figure 2 – Notre algorithme est capable de produire des cartes de la diffusion  $\sigma$  et de la dérive  $\mu$  au sein d'une cellule aussi bien en 2D et en 3D : **(a)** le paramètre de diffusion de protéines Gal-3 au sein d'une cellule SUM159 est représenté sous la forme d'un nuage de points en 3D ; **(b)** le paramètre de dérive de facteurs de transcription au sein d'une cellule est représenté en 2D sous la forme d'un nuage de vecteurs, eux-mêmes représentés sous forme de points dont la teinte et la saturation représentent respectivement la direction et l'amplitude.

évaluons la qualité et la robustesse de notre méthode hybride lagrangienne-eulérienne sur un ensemble étendu d’expériences, sur des jeux de données simulés et réels, en 2D et 3D. Nous comparons notre méthode à deux méthodes mentionnées dans l’état de l’art. Les cartes de diffusion et de dérive obtenues avec notre algorithme « DENSEMAPPING » sont plus précises que celles générées par ces dernières, et il est capable de générer des résultats exploitables même avec un nombre de trajectoires assez faible. Ceci procure un très net avantage par rapport aux méthodes existantes, principalement dédiées à la microscopie à super-résolution SPT-PALM. Au final, les biologistes peuvent étudier des dynamiques de protéines marquées par fluorescence avec des techniques de microscopie de fluorescence assez conventionnelles : TIRF, confocal, etc.

Dans le Chapitre 3, nous présentons une approche différente pour détecter les zones de piégeage (microdomaines ou aux zones de confinement) dans la cellule, correspondant aux régions où les molécules sont piégées. Après avoir défini le cadre mathématique et biophysique, nous présentons les travaux qui nous ont inspirés [Hozé, Nair, et al. 2012] [Masson, Dionne, et al. 2014] relatifs à la problématique de la détection de microdomaines. Le cœur de notre algorithme est basé sur une combinaison de méthodes de « clustering », utilisant également des procédures de classification de trajectoires capables de distinguer la sous-diffusion, la superdiffusion et le mouvement brownien. L’idée est d’identifier automatiquement les zones de piégeage où l’on observe une forte concentration de particules sous-diffusives. Nous évaluons notre preuve de concept sur des séquences artificielles générées grâce à l’algorithme FLUOSIM [Lagardère et al. 2020], un simulateur biophysique de mouvement de particules intracellulaires, et nous illustrons son potentiel sur des données réelles issues de la microscopie TIRF. Nous montrons notamment que même si des améliorations sont encore possibles en termes de robustesse, cette nouvelle procédure a un certain potentiel et s’avère très flexible par rapport aux méthodes existantes. Ce chapitre a été publié en collaboration avec V. Briane dans *Physical Biology* sous le titre « *A computational approach for detecting micro-domains and confinement domains in cells : a simulation study* » [Briane, Salomon, et al. 2020].

Dans le dernier chapitre, nous concluons, résumons nos contributions, et dressons quelques perspectives à cette thèse.

## Publications et communications

### Articles publiés en tant que premier auteur

- Salomon, Antoine, Cesar Augusto Valades-Cruz, Loïc Leconte et Charles Kervrann (mai 2020), « Dense mapping of intracellular diffusion and drift from single-particle tracking data analysis », dans : *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Barcelone, Espagne, pp. 966-970, DOI : 10.1109/ICASSP40776.2020.9054576. (présenté à l'oral)

### Articles publiés en tant que co-auteur

- Briane, Vincent, Myriam Vimond, Cesar Augusto Valades-Cruz, Antoine Salomon, Christian Wunder, Charles Kervrann (juin 2019), « A sequential algorithm to detect diffusion switching along intracellular particle trajectories », dans : *Bioinformatics* 36.1, pp. 317-329, DOI : 10.1093/bioinformatics/btz489.
- Briane, Vincent, Antoine Salomon, Myriam Vimond, Charles Kervrann (décembre 2019), « A computational approach for detecting micro-domains and confinement domains in cells: a simulation study », dans : *Physical Biology* 17.2, p. 025002, DOI : 10.1088/1478-3975/ab5e1d.

### Présentations orales

- Salomon, Antoine, Vincent Briane, Myriam Vimond, Cesar Augusto Valades-Cruz, Charles Kervrann (janvier 2019), « Analysing and quantifying intracellular particles movements in 3D LLSM data », *Quantitative Bio-Imaging - QBI* (Rennes, France), HAL : 02424609.
- Salomon, Antoine, Charles Kervrann (janvier 2020), « Precise mapping of intracellular diffusion and drift from SPT data analysis », *Quantitative Bio-Imaging - QBI* (Oxford, Royaume-Uni), HAL : 03087213.



# SUMMARY

---

## Context and challenges

During the past two decades many ground-breaking technologies emerged and allowed the visualization of tissues, cells, proteins, viruses, and macromolecular structures at all levels of spatial resolution (from 10 nm to 150 nm). The discovery of fluorescent labeling probes (Green Fluorescence Protein [GFP], Nobel Prize in chemistry 2008) and recent advances in optics and image sensors (e.g. PALM, STED and SIM) have been key developments to overcome the theoretical optical diffraction limit of 200 nm established in the 19th century. Because of these technological breakthroughs and their impacts in life sciences, light microscopy has been praised through prestigious awards, such as the Nobel Prize awarded to inventors of the concepts of super-resolution microscopy (Eric Betzig, Stefan W. Hell and William E. Moerner in 2014). Fluorescence microscopy imaging has since become the spearhead of modern biology. It is able to generate 3D+Time videos of dozens of gigabytes of data, depicting long-term nanoscale cell behaviors with low phototoxicity. This ability to follow nanoscale cellular events is proving to be of immense relevance in health, especially for the study of cancer progression and viral infections.

In general, all fluorescence microscopy systems record fluorescent signals emitted by molecules tagged with genetically engineered proteins within cells. In a conventional setup photons are collected and registered at a given pixel (or voxel in 3D imaging). The measured fluorescence intensity is a scalar value, generally proportional to the density of tagged-molecules representing a few dozens of nanometers within a pixel/voxel. As the fluorescence image data are 2D-3D+Time signals, potentially depicting several fluorescently tagged molecular species (multi-channel images), the interpretation of these signals is often challenging in image processing and analysis. Several scientific barriers in particular must be overcome. One of the most significant revolves around noise and blur reduction in images. The difficulty increases as soon as very large time series of images are dealt with. Particular attention must then be paid to the feasibility and scalability of the developed algorithms. Another challenge that is the subject of this thesis is to estimate 2D-3D spatial high-resolution maps of molecular motions from data generated

by microscopy instruments.

In the context of intracellular transport, the orchestration of all cellular components involved in the transport pathways of endocytosis [Lemaigre et al. 2023] [Johannes and Valades 2021] [Römer et al. 2007] [Renard et al. 2015] [Momboisse et al. 2022] and exocytosis [Lachuer et al. 2023] [Pécot et al. 2018] [Basset et al. 2017] [Gidon et al. 2012], two mechanisms respectively involved in cellular functions and defects such as tissue morphogenesis or degeneration and tumor cell invasion, is still very poorly understood. For a long time, this incomplete understanding of transport mechanisms could not be improved due to the lack of adapted technologies, that is in the right space-time resolution, at appropriate frame rate, without photo-bleaching (decreasing of ability for fluorescent tags to emit photons upon repeated excitation), and without photo-toxicity upon the live sample. A first answer to this challenge was proposed by 2D+time fluorescence microscopy imaging, in particular TIRF microscopy [Axelrod 2003]. It has allowed significant breakthroughs, related to the potential correlations between interacting molecules and dynamics at the cell surface (plasma membrane). More recently, the constant progress in molecular biology and optics has allowed for 3D+Time imaging of living cells. This leads to the acquisition of very huge image series: one single live cell experiment using cutting-edge Lattice Light Sheet Microscopy [Chen et al. 2014] can for instance produce up to one terabyte of data within an hour, at the spatial resolution of 100-200 nanometers in 3D. Therefore, it has become necessary to develop new paradigms for image reconstruction, 3D molecule tracking/motion estimation, and biophysical parameter estimation to face the huge series of volumes acquired with cutting-edge microscopy setups. In order to tackle these challenges, applied mathematics, image processing and analysis, and computer science have to be associated with biophysics and biology, and it is necessary to get a sustained synergy between these different scientific domains.

## **Main contribution**

In this thesis, given motion-based features observed in images, in particular tracklets (small trajectories) computed with single-particle tracking algorithms, we investigate methods dedicated to analyze the dynamics of molecules. Trajectories are classified into three motion categories: Brownian motion, confined diffusion (subdiffusion), directed diffusion (superdiffusion) (Figure 1). Statistical tests are set up to replace conventional methods (e.g. Mean Square Displacement (MSD)), which cannot reliably discriminate be-

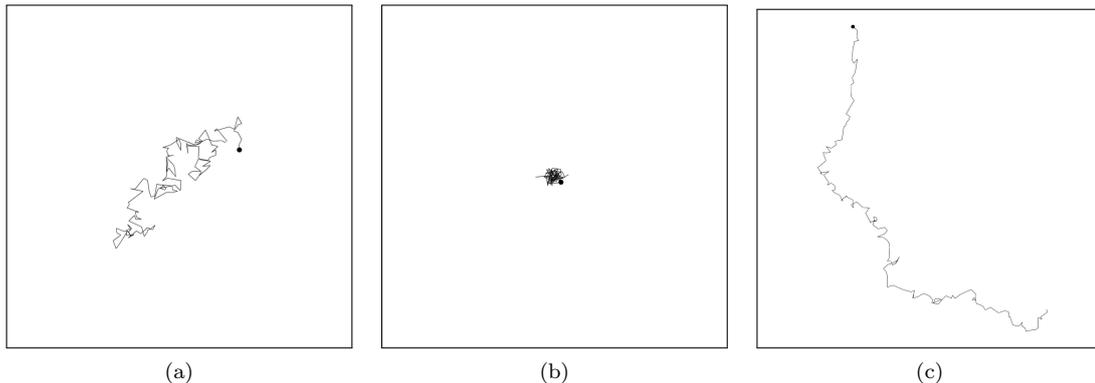


Figure 1 – Different types of particle motion are shown. **(a)** The particle undergoes free diffusion. This kind of motion results from random collisions with the surrounding fluid molecules. **(b)** The particle undergoes subdiffusion. This results from the particle being trapped in a microdomain. **(c)** The particle undergoes superdiffusion. The main forces involved in the particle’s motion are deterministic.

tween these three motion types. The detection of dynamics changes along long trajectories is performed by transposing the previous approach on sliding time windows along tracks. The main contribution of this thesis is an approach that can estimate molecular mobility and active transport from the computed trajectories. For this purpose, we assume that the motion of molecules is driven by the following stochastic differential equation:

$$d\mathbf{x} = \boldsymbol{\mu}(\mathbf{x})dt + \sigma(\mathbf{x})d\mathbf{b}, \quad (1)$$

where  $\mathbf{x} \in \mathbb{R}^d$ ,  $d \in \{2, 3\}$  denotes the position at time  $t$  of a molecule,  $\boldsymbol{\mu} \in \mathbb{R}^d$  is the drift related to active transport speed applied to the particle at point  $\mathbf{x}$ ,  $\sigma \in \mathbb{R}^+$  is the spatially-variable diffusion coefficient, and  $\mathbf{b} \in \mathbb{R}^d$  is a Brownian motion. Our method is able to produce spatially high-resolved maps of diffusion  $\sigma(\mathbf{x})$  and active transport  $\boldsymbol{\mu}(\mathbf{x})$  parameters from sequences of 2D or 3D images (Figure 2). Unlike previous paradigms which consist in reconstructing super-resolution images (e.g. SPT-PALM/STORM) from several thousand images, we directly provide spatial high-resolution maps of molecular mobility computed from a reduced number of conventional fluorescence microscopy images. Each biophysical quantity is calculated at positions corresponding exactly to the subpixel coordinates given by the molecule tracks, thus providing more detailed information.

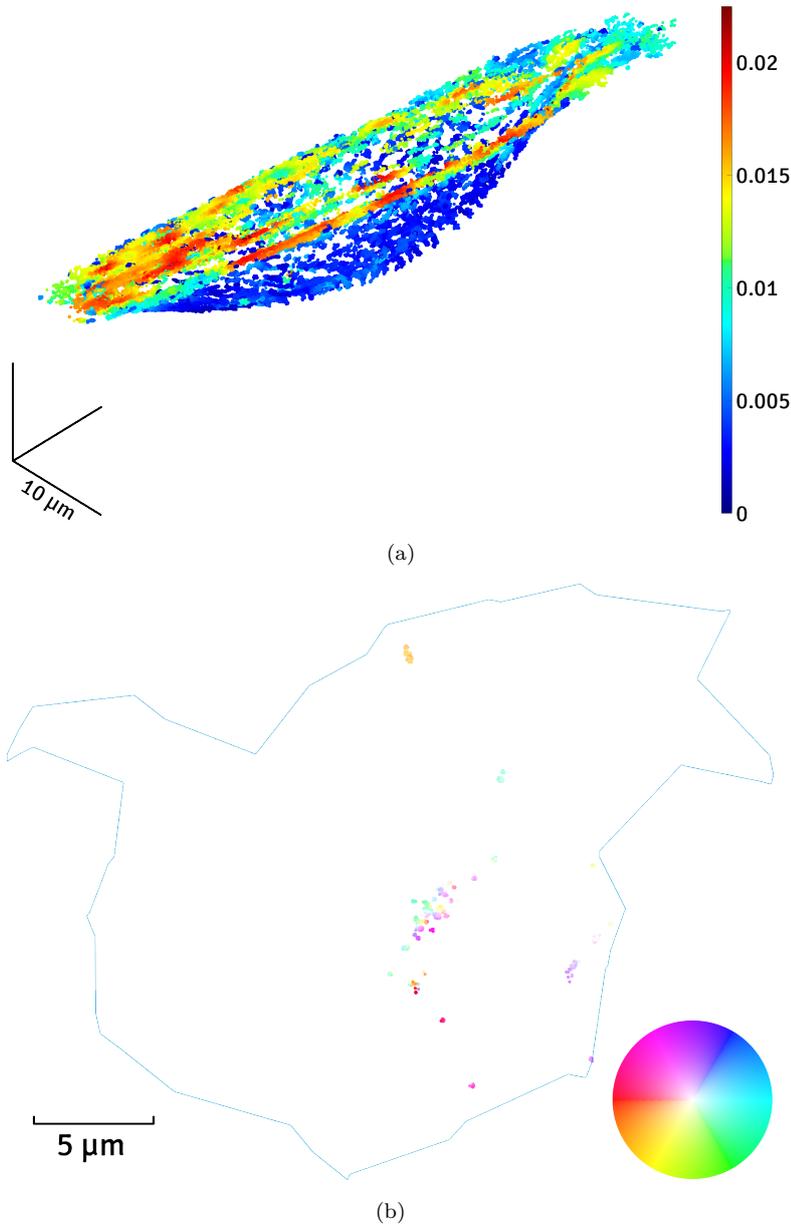


Figure 2 – Our algorithm is able to produce maps of diffusion  $\sigma$  and drift  $\mu$  within a cell in both 2D and 3D: **(a)** the diffusion parameter of Gal-3 proteins within a SUM159 cell is represented in the form of a 3D point cloud; **(b)** the drift parameter of transcription factors within a cell is represented in 2D as a vector cloud, which is itself represented as points whose hue and saturation represent direction and amplitude, respectively.

## Organization of the thesis

In the first chapter of this manuscript, we give a broad overview of the statistical theory, microscopy techniques and algorithms used in the context of our research. We start

by introducing the theoretical framework related to the concept of stochastic process, and in particular the characteristic so-called Brownian motion of particles moving in a fluid. After introducing a few models and types of particle motion, we continue by presenting both classic and groundbreaking, 2D and 3D, diffraction-limited and super-resolved fluorescence microscopy techniques. As our work exploits single molecule trajectories, we present a few tracking techniques and algorithms. Finally, we present Vincent Briane's particle motion classification method [Briane, Kervrann, and Vimond 2018] [Briane, Vimond, Valades-Cruz, et al. 2020], which we have inserted in our estimation scheme.

In Chapter 2, we propose a new cellular mapping method to robustly estimate dynamics (in particular diffusion and drift) in the entire cell. It is indeed of primary interest for biologists to be able to easily visualize protein dynamics. In addition, there is a high demand for algorithmic tools capable of automatically analyzing microscopy data to this end. After a review of a few techniques dedicated to the estimation of diffusion and drift, we present in details the theoretical elements of our statistical method. Next, we present space-time kernel estimators specific to each particle motion (free diffusion, subdiffusion, superdiffusion). We assess the quality and robustness of our hybrid Lagrangian-Eulerian method on an extensive set of experiments, on simulated and real data sets, in 2D and 3D. We compare our method to two methods mentioned in the state-of-the-art. The diffusion and drift maps obtained with our algorithm "DENSEMAPPING" are more accurate than those generated by the latter, and it is able to generate exploitable results even with a rather low number of trajectories. This provides a very clear advantage over existing methods, mainly dedicated to SPT-PALM super-resolution microscopy. In the end, biologists can study fluorescently tagged protein dynamics with fairly conventional fluorescence microscopy techniques: TIRF, confocal, etc.

In Chapter 3, we present a different approach to detect the trapping regions (microdomains or confinement areas) in the cell, corresponding to the regions where the molecules are trapped. After setting the mathematical and biophysical framework, we present the work [Hozé, Nair, et al. 2012] [Masson, Dionne, et al. 2014] related to the problem of microdomain detection that inspired us. The core of our algorithm is based on a combination of clustering methods, also using trajectory classification procedures able to distinguish subdiffusion, superdiffusion and Brownian motion. The idea is to automatically identify trapping areas where we observe a high concentration of subdiffusive particles. We evaluate our proof of concept on artificial sequences obtained with the algorithm FLUOSIM [Lagardère et al. 2020], a biophysics-based simulator for intracellular

particle motion, and we illustrate its potential on real TIRF microscopy data. We notably show that while there is still room for improvement in terms of robustness, this new procedure has some potential and is very flexible compared to existing methods. This chapter was published in collaboration with V. Briane in *Physical Biology* as "A computational approach for detecting micro-domains and confinement domains in cells: a simulation study" [Briane, Salomon, et al. 2020].

In the last chapter, we conclude, summarize our contributions, and draw some perspectives to this thesis.

## Publications and communications

### Articles published as first author

- Salomon, Antoine, Cesar Augusto Valades-Cruz, Loïc Leconte et Charles Kervrann (May 2020), « Dense mapping of intracellular diffusion and drift from single-particle tracking data analysis », in: *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Barcelona, Spain, pp. 966-970, DOI: 10.1109/ICASSP40776.2020.9054576. (*presented orally*)

### Articles published as co-author

- Briane, Vincent, Myriam Vimond, Cesar Augusto Valades-Cruz, Antoine Salomon, Christian Wunder, Charles Kervrann (June 2019), « A sequential algorithm to detect diffusion switching along intracellular particle trajectories », in: *Bioinformatics* 36.1, pp. 317-329, DOI: 10.1093/bioinformatics/btz489.
- Briane, Vincent, Antoine Salomon, Myriam Vimond, Charles Kervrann (December 2019), « A computational approach for detecting micro-domains and confinement domains in cells: a simulation study », in: *Physical Biology* 17.2, p. 025002, DOI: 10.1088/1478-3975/ab5e1d.

### Oral presentations

- Salomon, Antoine, Vincent Briane, Myriam Vimond, Cesar Augusto Valades-Cruz, Charles Kervrann (January 2019), « Analysing and quantifying intracellular particles movements in 3D LLSM data », *Quantitative Bio-Imaging - QBI* (Rennes,

France), HAL: 02424609.

- Salomon, Antoine, Charles Kervrann (January 2020), « Precise mapping of intracellular diffusion and drift from SPT data analysis », *Quantitative Bio-Imaging - QBI* (Oxford, Royaume-Uni), HAL: 03087213.



# THEORY, OBSERVATION AND ANALYTICS OF MICROSCOPIC PARTICLE MOTION

---

In this chapter, we present an overview of our research framework in both statistics, microscopy and computer science. We start by presenting the statistical theory related to the concepts of stochastic processes and diffusion. In particular, we present a few models and types of particle motion from the theoretical point of view. Second, our work being based on fluorescence microscopy data, we introduce a few microscopy techniques, both classic, groundbreaking, 2D, 3D, diffraction-limited and super-resolved. In addition, since we use single particle trajectories extracted from image sequences, we present several techniques and algorithms involved in tracking. Finally, we present particle motion classification methods, and in particular Vincent Briane's work that has been embedded in our estimation scheme.

## 1.1 Stochastic processes and diffusions

We present the probabilistic tools used to define diffusion processes, which are paramount for modelling intracellular dynamics, with a focus on  $d$ -dimensional processes,  $d \in \{2, 3\}$ . We define the notion of stochastic processes and focus on Brownian motion, which is the reference diffusion process. Furthermore, we introduce the the concept of Stochastic Differential Equation (SDE), and we present extensions of Brownian motion like fractional Brownian motion.

### 1.1.1 Stochastic process

A stochastic process is a function indexing random variables by a mathematical set. In the context of our research, random variables typically depict particle positions, and thus  $d$ -dimensional.

Let  $(\Omega, \mathcal{F}, P)$  be a probability space where  $\Omega$  is the sample space,  $\mathcal{F}$  a field and  $P$  a probability measure. A  $d$ -dimensional stochastic process is mathematically defined as:

$$\begin{aligned} I \times \Omega &\rightarrow \mathbb{R}^d \\ (t, \omega) &\mapsto \mathbf{x}(t, \omega), \end{aligned}$$

where  $I$  is a time interval. In what follows, we denote this application  $(\mathbf{x}_t)_{t \in I}$ , or shortly  $(\mathbf{x}_t)$ . A path of the stochastic process  $(\mathbf{x}_t)_{t \in I}$ , also called trajectory, is the application

$$\begin{aligned} I &\rightarrow \mathbb{R}^d \\ t &\mapsto \mathbf{x}(t, \omega). \end{aligned}$$

A stochastic process may be seen as an application from  $\Omega$  to the set of functions from  $I = [0, T]$  to  $\mathbb{R}^d$ . As previously mentioned, we consider only the stochastic processes whose trajectories are continuous, that is for almost  $\omega \in \Omega$ ,  $t \rightarrow \mathbf{x}_t(\omega)$  is continuous.

### 1.1.2 Brownian motion

During summer 1827, Scottish botanist Robert Brown was observing grains of pollen suspended in water, when he noticed the erratic motion of smaller particles ejected from the pollen grains. This observation is at the origin of both name and theory related to Brownian motion. The equations of Brownian motion were established in 1905 by Albert Einstein while its probabilistic framework was developed by Norbert Wiener in 1923.

In this section, we first focus on the one-dimensional Brownian motion, viewed as a Gaussian process, before extending the definition to the  $d$ -dimensional Brownian motion.

**Definition:** The one-dimensional Brownian motion  $(b_t)$  is a stochastic process with the following properties:

- $(b_t)$  is a process with independent increments: for all  $t > s$ ,  $b_t - b_s$  is independent of the field generated by the historic of the process  $(b_u)_{u \in [0, s]}$  until the time  $s$ .
- For all  $t > s$ ,  $b_t - b_s$  has normal distribution with mean 0 and variance  $t - s$ .
- The paths of  $(b_t)$  are almost surely continuous.

**Gaussian process:** A Gaussian process is a process for which all the finite-dimensional distributions are multivariate normal.

**Theorem 1** *A Brownian motion started at zeros is a Gaussian process with zero mean and covariance function  $\min(t, s)$ . Conversely, a Gaussian process with zero mean and covariance  $\min(t, s)$  is a Brownian motion.*

**Multivariate Brownian motion:** By extension, a  $d$ -dimensional Brownian motion ( $d \in 2, 3$ ) is defined as a random vector  $\mathbf{b}_t = (b_t^1, \dots, b_t^d)$ , where all coordinates  $b_t^i$  are independent one-dimensional Brownian motions.

### 1.1.3 Diffusion process

**Markov property:** The Markov property is central and amounts to considering that the future behavior of a process depends only on its current state and is independent of its past. In a discrete  $d$ -dimensional setup, the Markov property is defined as follows:

$$P(\mathbf{x}_j = k | \mathbf{x}_{j-1}, \dots, \mathbf{x}_0) = P(\mathbf{x}_j = k | \mathbf{x}_{j-1}). \quad (1.1)$$

As we consider time-continuous stochastic processes, the historic of the process  $\mathbf{x}_{j-1}, \dots, \mathbf{x}_0$  in the discrete case is replaced by a field  $\mathcal{F}_t$  at time  $t$ :

$$P(\mathbf{x}_{t+s} \in A | \mathcal{F}_t) = P(\mathbf{x}_{t+s} \in A | \mathbf{x}_t), \quad (1.2)$$

where  $A \subset \mathbb{R}^d$ . The difference between (1.1) and (1.2) lies in the different nature of the events  $\{\mathbf{x}_j = k\}$  (i.e. countable) and  $\{\mathbf{x}_{t+s} \in A\}$  (i.e. continuous space  $\mathbb{R}^d$ ), yielding the following theorem:

**Theorem 2** *The Brownian motion ( $\mathbf{b}_t$ ) has the Markov property.*

**Diffusions:** A diffusion process ( $\mathbf{x}_t$ ) is a continuous time process which possesses the Markov property and for which the sample paths are continuous. Any diffusion process satisfies the following conditions [Karlin and Taylor 1981]:

$$\lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} P(\|\mathbf{x}_{t+\Delta t} - \mathbf{x}_t\| > \epsilon | \mathbf{x}_t = \mathbf{x}) = 0, \quad \forall \epsilon > 0, \quad \forall \mathbf{x} \in \mathbb{R}^d, \quad (1.3)$$

where  $\|\cdot\|$  denotes the Euclidean norm. This basically means that large displacements of magnitude exceeding  $\epsilon > 0$  are very unlikely over sufficiently small intervals. Second, the following limits exist and constitute two parameters called drift and diffusion coefficient,

respectively:

$$\begin{aligned}\boldsymbol{\mu}(\mathbf{x}, t) &:= \lim_{\Delta t \rightarrow 0} \frac{\mathbb{E}[\mathbf{x}_{t+\Delta t} - \mathbf{x}_t | \mathbf{x}_t = \mathbf{x}]}{\Delta t}, \quad \forall \mathbf{x} \in \mathbb{R}^d, \\ \sigma(\mathbf{x}, t) &:= \lim_{\Delta t \rightarrow 0} \frac{\mathbb{E}[\|\mathbf{x}_{t+\Delta t} - \mathbf{x}_t\|^2 | \mathbf{x}_t = \mathbf{x}]}{\Delta t}, \quad \forall \mathbf{x} \in \mathbb{R}^d.\end{aligned}$$

The drift parameter  $\boldsymbol{\mu}(\mathbf{x}, t) : \mathbb{R}^d \times \mathbb{R}^+ \rightarrow \mathbb{R}^d$  is the deterministic component of a diffusion process. In the biological field, it can for instance characterize the active transport of vesicles propelled by molecular motors. The diffusion parameter  $\sigma(\mathbf{x}, t) : \mathbb{R}^d \times \mathbb{R}^+ \rightarrow S_+^d$  is linked to the stochastic component of a diffusion process, where  $S_+^d$  is the set of positive semi-definite matrix of size  $d$ . It can for instance characterize the number of collisions that a particle in a fluid undergoes.

It follows that Brownian motion is a diffusion process, for which the drift is the null function, and the diffusion coefficient is a constant.

#### 1.1.4 Stochastic Differential Equation (SDE)

A common framework to analyze diffusion processes is related to stochastic differential equations, of which they can be considered as solutions. At starting point, diffusion models were established to represent the motion of particles in a fluid as submitted to a deterministic force for one part, and a random force caused by random collisions with others particles for a second part.

Let us denote  $\mathbf{x}_t \in \mathbb{R}^d$  the position of the particle at time  $t$  and  $(\mathbf{b}_t)$  a  $d$ -dimensional Brownian motion. We assume that  $\mathbf{x}_t = \mathbf{x}$ . The displacement of the particle between  $t$  and  $t + \Delta t$  may be represented by

$$\mathbf{x}_{t+\Delta t} - \mathbf{x} \approx \boldsymbol{\mu}(\mathbf{x}, t)\Delta t + \sigma(\mathbf{x}, t)(\mathbf{b}_{t+\Delta t} - \mathbf{b}_t). \quad (1.4)$$

where  $\Delta t$  is the time interval,  $\boldsymbol{\mu}(\mathbf{x}, t)$  is the drift and  $\sigma(\mathbf{x}, t)$  is the diffusion coefficient. The term  $\sigma(\mathbf{x}, t)(\mathbf{b}_{t+\Delta t} - \mathbf{b}_t)$  is the random component of the motion caused by the random collisions with the surrounding fluid. The diffusion coefficient  $\sigma(\mathbf{x}, t)$  can represent the number of collisions, which generally increase with the fluid temperature.

**Definition:** In (1.4), the displacement of the particle  $\mathbf{x}_{t+\Delta t} - \mathbf{x}$  is approximated by a Gaussian random variable of mean  $\boldsymbol{\mu}(\mathbf{x}, t)\Delta t$  depending on the drift and of variance

$\sigma(\mathbf{x}, t)\sqrt{\Delta t}$  depending on the diffusion coefficient. If  $\Delta t$  is small and represented by infinitesimal time step  $dt$ , and motion is small, that is  $\mathbf{b}_{t+\Delta t} - \mathbf{b}_t \approx d\mathbf{b}_t$  and  $\mathbf{x}_{t+\Delta t} - \mathbf{x}_t = d\mathbf{x}_t$ , we get

$$d\mathbf{x}_t = \boldsymbol{\mu}(\mathbf{x}_t, t)dt + \sigma(\mathbf{x}_t, t)d\mathbf{b}_t, \quad (1.5)$$

with  $(\mathbf{b}_t)$  a  $d$ -dimensional Brownian motion,  $\boldsymbol{\mu}(\mathbf{x}, t) : \mathbb{R}^+ \times \mathbb{R}^d \rightarrow \mathbb{R}^d$  and  $\sigma(\mathbf{x}, t) : \mathbb{R}^+ \times \mathbb{R}^d \rightarrow \mathcal{M}^d$  ( $\mathcal{M}^d$  denoting the set of square matrix of size  $d$ ).

This equation is called a stochastic differential equation (SDE), where  $(\mathbf{x}_t)$  is the unknown process.

**Solution of SDE:** There are two types of solutions respectively called strong and weak solutions. Let  $\mathcal{F}_t$  be the field induced by the initial condition  $\mathbf{x}_0$  and the Brownian motion  $(\mathbf{b}_t)$  which drives the stochastic differential (1.5). Equation (1.5) has a strong solution  $(\mathbf{x}_t)$  on the probability space  $(\Omega, \mathcal{F}, P)$  with respect to  $(\mathbf{b}_t)$  and initial condition  $\mathbf{x}_0$  if the stochastic process  $\mathbf{x}_t$  satisfies (1.5), has continuous paths and that  $\mathbf{x}_t$  is  $\mathcal{F}_t$ -measurable for all  $t$ . In other words,  $\mathbf{x}_t$  is the output of a system parametrized by  $\boldsymbol{\mu}(\mathbf{x}, t)$  and  $\sigma(\mathbf{x}, t)$  whose input is the Brownian motion  $(\mathbf{b}_t)$ . It reflects the principle of causality of the system.

A weak solution of a SDE consists in building at the same time a couple of processes  $(\mathbf{x}_t, \mathbf{b}_t)$  where  $(\mathbf{x}_t)$  is a solution of the SDE driven by the Brownian  $(\mathbf{b}_t)$ .

The solution of the stochastic differential equation is written as [Klebaner 2012]

$$\mathbf{x}_t = \mathbf{x}_0 + \int_0^t \boldsymbol{\mu}(\mathbf{x}_s, s)ds + \int_0^t \sigma(\mathbf{x}_s, s)d\mathbf{b}_s, \quad (1.6)$$

### 1.1.5 Fractional Brownian motion

Fractional Brownian motion (fBm) was introduced to model scale-invariant phenomena processes showing long-range dependence. First, [Kolmogorov 1941] developed a turbulence theory based on two hypotheses of scale invariance. Then, [Hurst 1951] observed hydrological events invariant to changes in scale. Finally, [Mandelbrot and Ness 1968] defined fractional Brownian motion: "fBm of exponent  $\mathfrak{h}$  is a moving average of  $db(t)$ , in which past increments of  $b(t)$  are weighted by the kernel  $(t - s)^{2\mathfrak{h}-1}$ ". The parameter  $\mathfrak{h}$  is known as the Hurst index or Hurst parameter. More recently, [Coutin and Z. Qian 2002] give the following definition of a  $d$ -dimensional fractional Brownian motion:

**Definition:** A fractional Brownian motion in dimension  $d > 1$  is the random vector  $\mathbf{b}_t^{\mathfrak{h}} = (b_t^{\mathfrak{h},1}, \dots, b_t^{\mathfrak{h},d})$  where all coordinates  $b_t^{\mathfrak{h},i}$  are independent one-dimensional fractional Brownian motions of Hurst parameter  $0 < \mathfrak{h} < 1$ .

A  $d$ -dimensional fBm reduces to a  $d$ -dimensional Brownian motion in the case  $\mathfrak{h} = 1/2$ . Accordingly, we can extend the stochastic differential equation (1.5) to define a SDE driven by a fBm of Hurst index  $0 < \mathfrak{h} < 1$ :

$$d\mathbf{x}_t = \boldsymbol{\mu}(\mathbf{x}_t, t)dt + \sigma(\mathbf{x}_t, t)d\mathbf{b}_t^{\mathfrak{h}}. \quad (1.7)$$

The SDE driven by Brownian motion (1.5) is of the form of the SDE (1.7) with  $\mathfrak{h} = 1/2$ .

## 1.2 Particle motion modeling

In this thesis, we focus on Brownian motion, subdiffusion and superdiffusion, which are the three main types of diffusion studied in biophysics to model intracellular particle dynamics, as well as the following SDE derived from Langevin:

$$d\mathbf{x} = \boldsymbol{\mu}(\mathbf{x})dt + \sigma(\mathbf{x})d\mathbf{b}, \quad (1.8)$$

where  $\mathbf{x} \in \mathbb{R}^d$ ,  $d \in \{2, 3\}$  designates the particle position,  $\boldsymbol{\mu} \in \mathbb{R}^d$  is the drift,  $\sigma \in \mathbb{R}^+$  is the diffusion coefficient, and  $\mathbf{b} \in \mathbb{R}^d$  is standard Gaussian white noise.

### 1.2.1 Langevin's approach

[Langevin 1908] characterizes the particle motion through the  $d$ -dimensional (Langevin) equation

$$m\ddot{\mathbf{x}}(t) = -\gamma\dot{\mathbf{x}}(t) + \mathbf{f}_s(t), \quad (1.9)$$

where  $\dot{\mathbf{x}} : \mathbb{R}^+ \rightarrow \mathbb{R}^d$  designates the velocity of the particle,  $m$  is its mass,  $\gamma \in \mathbb{R}_+^*$  is the friction coefficient and  $\mathbf{f}_s : \mathbb{R}^+ \rightarrow \mathbb{R}^d$  is a stochastic force resulting from the collisions with the surrounding particles. If we assume that the particles are spherical, we have  $\gamma = 6\pi ka$ , where  $k$  is the viscosity coefficient of the surrounding liquid and  $a$  is the radius of the particle.

### Ornstein-Uhlenbeck process

Uhlenbeck and Ornstein make two additional assumptions concerning  $\mathbf{f}_s$  [Uhlenbeck and Ornstein 1930]. First, all colliding particles including the particle of interest are similar, have the same initial speed  $\dot{\mathbf{x}}_0$ , and

$$\mathbb{E}[\mathbf{f}_s(t)] = \mathbf{0}_d,$$

where  $\mathbf{0}_d$  is the null vector of  $\mathbb{R}^d$ . That is the mean of  $\mathbf{f}_s(t)$  is null over a large number of independent collisions. Second, the authors state that the autocorrelation function of  $\mathbf{f}_s$  is a function of  $t - s$  with a very sharp peak of width equal to the duration of a single collision at  $t - s = 0$ . [Van Kampen 2007] remarks that as long as the collision time is shorter than all other relevant times, we can more conveniently use the Dirac function as follows:

$$\mathbb{E}[\mathbf{f}_s(t)\mathbf{f}_s(s)^T] = C\delta(t - s)\mathbf{I}_d, \quad (1.10)$$

where  $C \in \mathbb{R}_+^*$  is a constant,  $\delta$  is the Dirac function and  $\mathbf{I}_d$  is the identity matrix of size  $d$ . The idea is that each collision is practically instantaneous and that successive collisions are uncorrelated.

If we further assume that  $\mathbf{f}_s$  is Gaussian, we entirely define this process as a Gaussian process which is determined by its first two moments.  $\mathbf{f}_s(t)$  is then called a Gaussian white noise, and can be defined as the derivative of a Brownian motion in an informal way (Brownian motion technically being nowhere differentiable) [Karlin and Taylor 1981]:

$$\mathbf{f}_s(t) = \sigma \dot{\mathbf{b}}(t). \quad (1.11)$$

Finally, we can rewrite (1.9) as the  $d$ -dimensional SDE

$$m d\mathbf{v}(t) = -\gamma \mathbf{v}(t)dt + \sigma d\mathbf{b}(t), \quad (1.12)$$

where  $\mathbf{v} := \dot{\mathbf{x}}$ .

The solution of SDE (1.12) is known as the Ornstein-Uhlenbeck process, which is a

Gaussian process such that

$$\begin{aligned}\mathbb{E}[\mathbf{v}(t)] &= \mathbf{0}_d, \\ \mathbb{E}[\mathbf{v}(t)\mathbf{v}(s)^T] &= \frac{\sigma^2}{2\gamma m} e^{-(\gamma/m)|t-s|} \mathbf{I}_d.\end{aligned}\tag{1.13}$$

When  $t \rightarrow \infty$ , the mean square velocity verifies [Waterston, Beaufort, and Strutt 1892]

$$\lim_{t \rightarrow \infty} \mathbb{E}[\|\mathbf{v}(t)\|^2] = d \frac{k_B T}{m},\tag{1.14}$$

where  $d \in \{2, 3\}$  is the dimensionality of the process,  $k_B$  is the Boltzmann constant and  $T$  is the temperature. Each component of the velocity vector has the same variance:

$$\lim_{t \rightarrow \infty} \mathbb{E}[v_i(t)^2] = \frac{k_B T}{m}, \quad i = 1, \dots, d.\tag{1.15}$$

From (1.13) and (1.15), we deduce

$$\sigma = \sqrt{2\gamma k_B T}.$$

Finally, the Brownian motion of [Langevin 1908] is defined as

$$\mathbf{x}(t) = \int_0^t \mathbf{v}(s) ds,\tag{1.16}$$

where  $\mathbf{v}$  is the Ornstein-Uhlenbeck process solution of the SDE (1.12). Due to the Gaussian nature of  $\mathbf{v}$ ,  $\mathbf{x}$  is also a Gaussian process.

### Mean Square Displacement

Einstein and Langevin both established that particle motion asymptotically exhibits a linear mean square displacement (MSD) [Einstein 1905] [Langevin 1908].

In the  $d$ -dimensional case, Einstein shows on one hand that the mean square displacement of the Brownian motion is

$$\mathbb{E}[\|\mathbf{x}_t - \mathbf{x}_0\|^2] = d2\sigma t.$$

On the other hand, he demonstrates that

$$\sigma = \frac{RT}{n\gamma},$$

where  $R$  is the gas constant,  $T$  designates the temperature,  $n$  is the amount of substance, and  $\gamma$  is the friction coefficient. We can thus write

$$\mathbb{E} [\|\mathbf{x}_t - \mathbf{x}_0\|^2] = d \frac{2k_B T}{\gamma} t,$$

where  $k_B = R/n$  is the Boltzmann constant.

Assuming  $\mathbf{x}_0 = 0$  for the sake of simplicity, Langevin showed that

$$\begin{aligned} \mathbb{E} [\|\mathbf{x}_t - \mathbf{x}_0\|^2] &= \sum_{i=1}^d \mathbb{E} \left[ \int_0^t \int_0^t v_i(s) v_i(u) ds du \right], \\ &= d \int_0^t \int_0^t \mathbb{E} [v_1(s) v_1(u)] ds du, \\ &= d \frac{2k_B T}{\gamma} \left( t - \frac{m}{\gamma} (1 - e^{-(\gamma/m)t}) \right), \\ &= d \frac{2k_B T}{\gamma} t + o(t), \end{aligned}$$

where  $o(t) \rightarrow 0$  as  $t \rightarrow \infty$ , leading us to the same conclusion as Einstein.

In conclusion, Brownian motion is nowhere differentiable according to Einstein, and then has a rough but continuous path. On the other hand, the particle motion defined by Langevin is differentiable due to its definition as an Ornstein-Uhlenbeck process integration (Equation (1.16)). Its path is then smooth.

### 1.2.2 Subdiffusion

In this section, we present subdiffusion processes and their mathematical transcriptions. They notably include confined and anomalous diffusion, which are translations of several biological scenarios.

## Anomalous subdiffusion

According to [Meroz and Sokolov 2015], an anomalous diffusion process  $(\mathbf{x}_t)$  is characterized by a MSD growing non-linearly in time, such that

$$\mathbb{E} \left[ \|\mathbf{x}_t - \mathbf{x}_0\|^2 \right] \propto t^\alpha, \quad (1.17)$$

with  $\alpha \neq 1$ . More specifically,  $(\mathbf{x}_t)$  is called anomalously subdiffusive if  $0 < \alpha < 1$ . We present three models of anomalous subdiffusion, the first two being solutions of a fBm-driven SDE (1.7).

**Fractional Brownian motion** This first model is simply fBm with a Hurst index  $0 < \mathfrak{h} < 1/2$ . Its MSD is given by (1.17), with  $\alpha = 2\mathfrak{h} < 1$ . In [Jeon et al. 2011], it is used to characterize the longer-time motion of a lipid granule in the crowded environment of a yeast cell, and in presence of plastic elements that generate correlations in the granule trajectory.

**Generalized Langevin Equation (GLE)** Kou modeled the motion of particles being slowed down by contrary current caused by the viscoelastic properties of the cytoplasm with stochastic differential equations driven by fBm with Hurst index  $1/2 < \mathfrak{h} < 1$  [Kou 2008]. Moreover, Zwanzig and Chandler proposed the following generalized Langevin equation (GLE) [Zwanzig 2001] [Chandler 1987]:

$$m\ddot{\mathbf{x}}(t) = -\gamma \int_{-\infty}^t \dot{\mathbf{x}}(u)K(t-u)du + \mathbf{g}(t), \quad (1.18)$$

where the velocity  $\dot{\mathbf{x}}$  is convoluted with a kernel  $K$ , and  $\mathbf{g}(t)$  is a noise with memory replacing the memory-less white noise  $\mathbf{f}_s(t)$ . The autocorrelation function of  $\mathbf{g}$  is given by

$$\mathbb{E} \left[ \mathbf{g}(t)\mathbf{g}(t)^T \right] = \sigma K(t-s)\mathbf{I}_d. \quad (1.19)$$

If we choose  $K$  as the Dirac function  $\delta$ , GLE (1.18) is equivalent to the Langevin equation (1.9) and the condition on the second moment (1.19) is equivalent to the condition (1.10). [Kou 2008] chooses to define  $\mathbf{g}(t)$  as fractional Gaussian noise (1.6) with Hurst index  $1/2 < h < 1$  for exhibiting long range dependence. From condition (1.19),

the authors deduce the kernel  $K$  (noted now  $K_{\mathfrak{h}}$ )

$$K_{\mathfrak{h}}(t) = 2\mathfrak{h}(2\mathfrak{h} - 1)|t|^{2\mathfrak{h}-2}.$$

Then, the related stochastic differential equation is

$$m d\mathbf{v}(t) = -\gamma \left( \int_{-\infty}^t \mathbf{v}(u) K(t-u) du \right) dt + \sigma d\mathbf{b}^{\mathfrak{h}}(t),$$

where  $\mathbf{v} := \dot{\mathbf{x}}$  and  $(\mathbf{b}_t^{\mathfrak{h}})$  is a fBm with  $1/2 < \mathfrak{h} < 1$ . Finally, [Kou 2008] shows that the integrated process  $\mathbf{x}_t = \int \mathbf{v}(u) du$  verifies

$$\mathbb{E} [\|\mathbf{x}_t - \mathbf{x}_0\|^2] \propto t^{2-2\mathfrak{h}}$$

as  $t \rightarrow \infty$ . It fulfills the MSD condition (1.17) asymptotically with  $\alpha = 2 - 2\mathfrak{h} < 1$  for  $1/2 < \mathfrak{h} < 1$ .

**Continuous Time Random Walk (CTRW)** It is frequent particles switching between two states, respectfully binding events where the particle is attached to a spot, and movement between two binding spots. [Scher and Montroll 1975] introduced the concept of continuous time random walk (CTRW) to represent such dynamics. Formally, a particle performs random jumps whose step length is generated by a probability density with finite second moments. The waiting times between jumps are assumed to be distributed according to a probability distribution  $\psi(t)$ . If  $\psi(t)$  has a finite first moment that is  $\int t\psi(t)dt < \infty$ , then the mean square displacement of the CTRW is time-linear. For instance, we can use the exponential distribution

$$\psi(t) = (1/\tau)e^{-t/\tau}, \quad t \in \mathbb{R}^+,$$

where  $\tau > 0$  is called the characteristic time. In this case, the random walk has the Markov property (due to the memoryless property of the exponential distribution). If on the contrary we have  $\int t\psi(t)dt = \infty$ , then the mean square displacement of the CTRW is given by (1.17). A typical choice is a power law distribution

$$\psi(t) = 1/(1 + t/\tau)^{1+\alpha}, \quad t \in \mathbb{R}^+,$$

where  $\tau > 0$  is the characteristic time and we have  $0 < \alpha < 1$ .

## Confined diffusion

We now present another type of subdiffusion, namely confined diffusion. According to [Saxton 1994][Monnier, Guo, et al. 2012], a confined diffusion process ( $\mathbf{x}_t$ ) is characterized by a MSD of the following form:

$$\mathbb{E} [\|\mathbf{x}_t - \mathbf{x}_0\|^2] = \frac{r_c^2}{a} \left( 1 - \beta e^{-\zeta \sigma^2 / (2r_c^2)} \right), \quad (1.20)$$

where  $r_c$  is the characteristic size of the confinement region,  $a$  is a scale parameter,  $\beta$  and  $\zeta$  depend on the shape of the region, and  $\sigma \in \mathbb{R}^+$  designates the constant diffusion coefficient. In what follows, we focus on one model of confined diffusion: diffusion in a potential well.

**Diffusion in a potential well** Generally speaking, it is possible for a particle to be subjected to an external, deterministic force in addition to the stochastic force caused by the collisions with surrounding fluid particles. In such a case, the Langevin equation (1.9) is extended with an extra term and becomes

$$m\ddot{\mathbf{x}}(t) = -\gamma\dot{\mathbf{x}}(t) + \mathbf{f}_d(t) + \mathbf{f}_s(t), \quad (1.21)$$

where  $\mathbf{f}_d : \mathbb{R}^d \times \mathbb{R}_+ \rightarrow \mathbb{R}^d$  designates the external deterministic force applied to the particle, and  $\mathbf{f}_s(t) = \sigma\dot{\mathbf{b}}(t)$  as stated in (1.11).

More specifically, [Kramers 1940] introduced a subdiffusion model where a particle is attracted to a spot by an external force derived from a potential well, originally to describe chemical reactions. In this case, (1.21) can be rewritten as

$$m\mathbf{d}\mathbf{v}(t) = -\gamma\mathbf{d}\mathbf{x}(t) - \nabla U(\mathbf{x}_t)dt + \sigma\mathbf{d}\mathbf{b}(t),$$

where  $\nabla$  denotes the gradient operator, and  $U$  designates the potential well. Furthermore, if we assume that the so-called overdamped condition [Van Kampen 2007] is satisfied, the acceleration term  $m\mathbf{d}\mathbf{v}(t)$  is negligible and the viscosity is very large, implying  $\gamma \rightarrow \infty$ . Consequently, the model reduces to

$$\gamma\mathbf{d}\mathbf{x}(t) = -\nabla U(\mathbf{x}_t)dt + \sigma\mathbf{d}\mathbf{b}(t). \quad (1.22)$$

Also, we assume that the potential  $U$  is uni-modal, suggesting that the particle is trapped

in a single domain. In this case,  $U$  can be approximated by a polynomial of order 2 as follows:

$$U(x_1, \dots, x_d) = (1/2) \sum_{i=1}^d k_i (x_i - \theta_i)^2,$$

where  $k_i \in \mathbb{R}^+$ ,  $\theta_i \in \mathbb{R}$  and  $d$  is the dimensionality of the process. The parameter  $k_i$  characterizes the strength of attraction of the potential well, in relation to the well depth, while  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_d)$  is the equilibrium position towards which the particle is attracted. The SDE (1.22) can then be rewritten as

$$dx_i(t) = -\lambda_i(x_i(t) - \theta_i)dt + \sigma db_i(t), \quad i = 1, \dots, d, \quad (1.23)$$

where  $\lambda_i = k_i/\gamma > 0$ . The solution of the SDE (1.23) is the Ornstein-Uhlenbeck process, as in the case of Equation (1.12), but with a different parametrization, notably with the extra  $\theta_i$  parameters. Since it is a Gaussian process with normal stationary distribution, the MSD of the process is given by

$$\mathbb{E} [\|\mathbf{x}_t - \mathbf{x}_0\|^2] = \sigma^2(1 - e^{-\lambda t}) \sum_{i=1}^d (1/\lambda_i), \quad (1.24)$$

when  $\mathbf{x}_0$  is drawn with the stationary distribution.

If  $k_i = k$  for  $i = 1, \dots, d$ , we have  $\lambda_i = \lambda$  for  $i = 1, \dots, d$  and (1.23) can be rewritten as

$$d\mathbf{x}(t) = -\lambda(\mathbf{x}(t) - \boldsymbol{\theta})dt + \sigma d\mathbf{b}(t), \quad i = 1, \dots, d.$$

Then, the MSD becomes

$$\mathbb{E} [\|\mathbf{x}_t - \mathbf{x}_0\|^2] = \frac{d\sigma^2(1 - e^{-\lambda t})}{\lambda}. \quad (1.25)$$

The MSD in (1.20) is notably equal to (1.25) with  $r_c^2 = \sigma^2/(2\lambda)$ ,  $a = 2/d$  and  $\beta = \zeta = 1$ .

### 1.2.3 Superdiffusion

In biophysics, the representation of superdiffusion has not received much attention. The most common model is Brownian with drift, for which the particle is subjected to a constant force in addition to the stochastic force caused by the collisions with surrounding fluid particles. We saw earlier that in such a case, the Langevin equation can be extended

to Equation (1.21). In the overdamped condition, (1.21) becomes

$$d\mathbf{x}(t) = \boldsymbol{\mu}(t)dt + \sigma d\mathbf{b}(t),$$

where  $\boldsymbol{\mu} = (\mu_1, \dots, \mu_d) \in \mathbb{R}^d$  is the constant drift parameter modelling the velocity of the molecular motor.

The MSD of the directed Brownian motion is then given by

$$\mathbb{E} [\|\mathbf{x}_t - \mathbf{x}_0\|^2] = \|\boldsymbol{\mu}\|^2 t^2 + d\sigma^2 t,$$

where the linear term comes from the Brownian motion, while the quadratic term is due to the constant drift. In absence of Brownian component, the MSD becomes fully quadratic and the motion is described as ballistic, that is the particle has a straight motion.

**Anomalous superdiffusion** Anomalous superdiffusions are the superdiffusive analogue to anomalous subdiffusions. The MSD of an anomalous superdiffusion ( $\mathbf{x}_t$ ) is characterized by a MSD which is proportional to the monome  $t^\alpha$ :

$$\mathbb{E} [\|\mathbf{x}_t - \mathbf{x}_0\|^2] \propto t^\alpha, \tag{1.26}$$

but this time with  $1 < \alpha < 2$ .

**Fractional Brownian motion** Superdiffusion can also be modeled by the fractional Brownian motion, with Hurst parameter  $1/2 < \mathfrak{h} < 1$ . In fact, we know that the MSD of the fBm is given by Equation (1.26). We note however that the use of the fractional Brownian motion in biophysics is mainly related to subdiffusion.

### 1.3 Fluorescence microscopy techniques

The principle of optical microscopy is to illuminate small objects and magnify them through a system of lenses. Optical microscopy can use a diverse set of optical phenomena. In particular, a fluorescence microscope uses fluorescence to highlight objects. Fluorescence is a form of luminescence where the absorption of electromagnetic radiation by a substance causes it to emit light. The specimen under study is usually illuminated with radiation of one or several wavelengths called excitation wavelength, which are then

absorbed by compounds called fluorophores that emit light of one or several wavelengths, called emission wavelength. Emission wavelength are usually longer than excitation wavelengths. In our framework, fluorophores are attached to particles of interest in a biological specimen, allowing to study their motion.

In Subsection 1.3.1, we give a summary of existing conventional microscopy techniques. Then in 1.3.2, we introduce the diffraction limit problem and how it can be resolved through super-resolution microscopy. In Subsection 1.3.3, we present some super-resolution fluorescence microscopy techniques.

### 1.3.1 Conventional microscopy techniques

We define conventional fluorescence microscopy techniques as any technique that is limited by the optical phenomenon of diffraction, which will be presented in the next subsection.

#### Wide-Field Fluorescence Microscopy (WFFM)

Wide-Field Fluorescence Microscopy or WFFM is the most rudimentary fluorescence microscopy technique. In this configuration, illustrated in Figure 1.1, the specimen is illuminated as a whole with radiation of one or several wavelength, which activates all fluorophores present in the specimen.

The advantages of this technique are its simplicity, its low maintenance cost and the possibility of directly observing the specimen through the ocular with the naked eye. Its main problem is that optical microscopes have a very short depth of field, and the objects can only be visualized in a thin plane of focus. The indiscriminate activation of fluorophores, even when they are out-of-focus, causes a lack of contrast, sharpness and resolution. This drawback can partially be corrected with either computational deconvolution [Sibarita 2005], or optical techniques such as the use of thin apertures (e.g. with a f-number of  $f/1.5$ ). But the power of deconvolution is limited, and thin apertures reduce the brightness of produced images. This last problem can in turn be corrected by computerized means, but at the cost of a higher noise level.

#### Confocal microscopy

In confocal microscopy the excitation light beam is concentrated by lenses on a single point of the specimen, as illustrated in Figure 1.2. The emitted fluorescence light is then

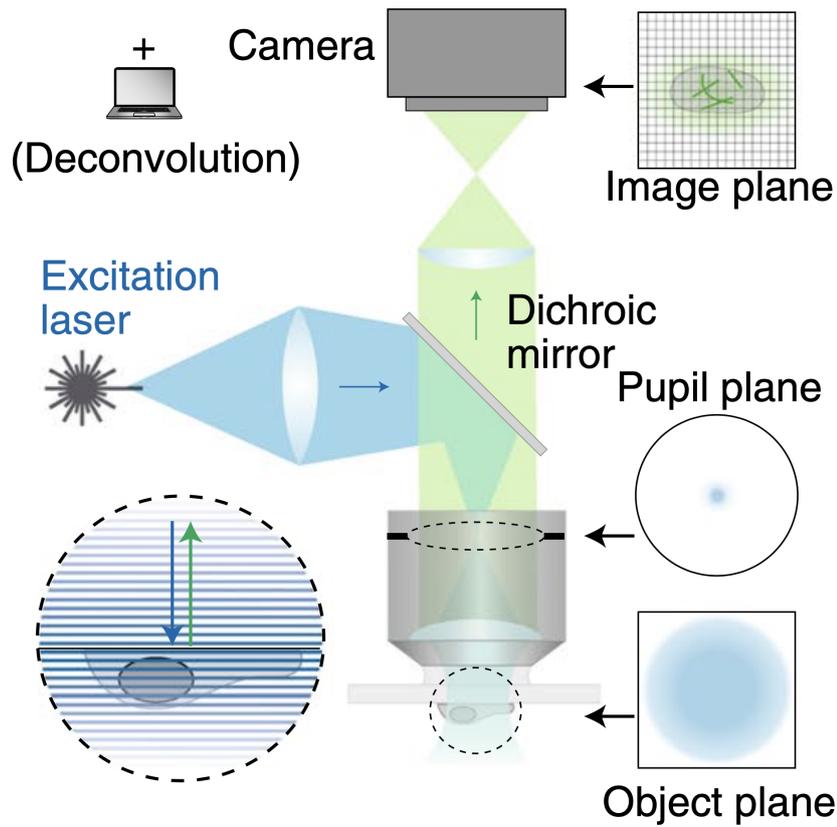


Figure 1.1 – Principle and light path of Wide-Field Fluorescence Microscopy. The excitation light is focused in the center of the pupil plane. (illustration extracted and edited from [Schermelleh et al. 2019], Fig. 1(a))

filtered by a pinhole placed below the detector to remove any signal emitted from out-of-focus zones of the specimen. Different techniques of horizontal scanning, such as point scanning or spinning disk, can be used to generate an image of the specimen as a whole. 3D confocal microscopy is possible by generating an image on several planes of focus and then reconstructing a 3D image by computerized means.

By removing the out-of focus signal, this method has the advantage of producing sharper images with better contrast than WFFM. However, this is done at the expense of fluorescence intensity since much of the emitted light is blocked by the pinhole, lowering the obtained image brightness. Another problem is slowness of the scanning process that does not allow for live imaging, not to mention the impossibility of observation through the naked eye. The temporal resolution of time sequences generated through this technique is consequently limited.

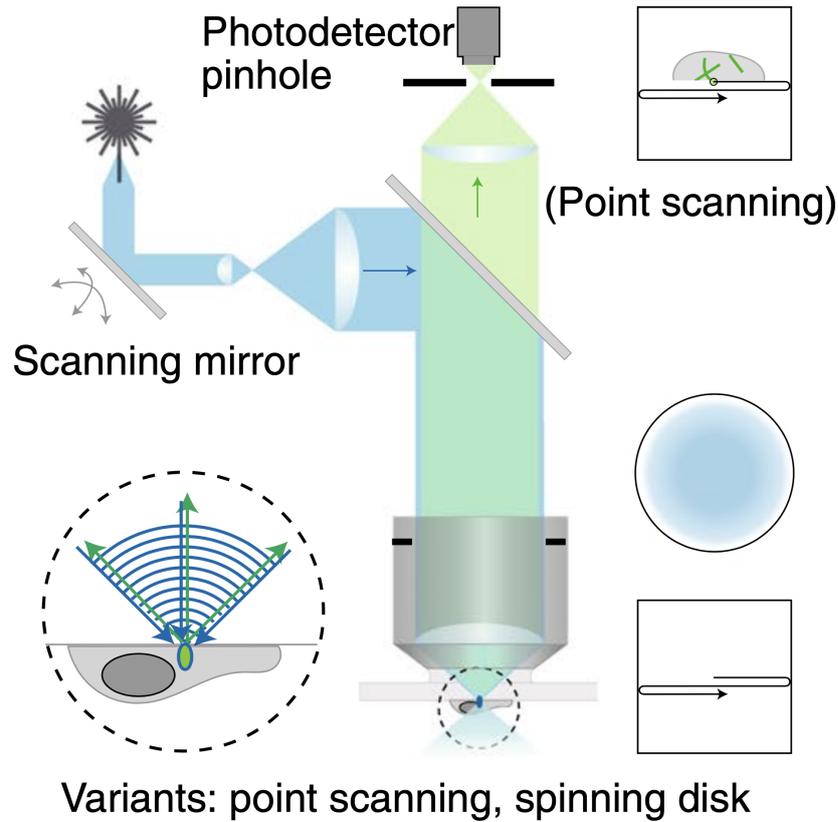


Figure 1.2 – Principle and light path of confocal microscopy. The light beam fills the pupil plane, and is focused and scanned across the object plane. (illustration extracted and edited from [Schermelleh et al. 2019], Fig. 1(c))

### Total Internal Reflection Fluorescence Microscopy (TIRF)

As its name suggests, TIRF microscopy uses the optical phenomenon of total internal refraction. Consider two isotropic media of respective refraction indices  $n_1$  and  $n_2$  separated by a flat interface, and a light beam passing through the first medium hitting said interface. As illustrated in Figure 1.3, Snell's law states that

$$n_1 \sin \theta_1 = n_2 \sin \theta_2 ,$$

where angle  $\theta_1$  and  $\theta_2$  designate the angles of incidence and refraction, respectively.

If  $n_1 > n_2$ , we can define the critical incidence angle as  $\theta_c = \sin^{-1}(\frac{n_2}{n_1})$ . Total internal refraction happens when  $\theta_1 > \theta_c$ , as the light beam is almost totally reflected by the interface. However, a small portion of the incident beam passes through the interface, propagating parallel to the plane of incidence and creating an electromagnetic field close

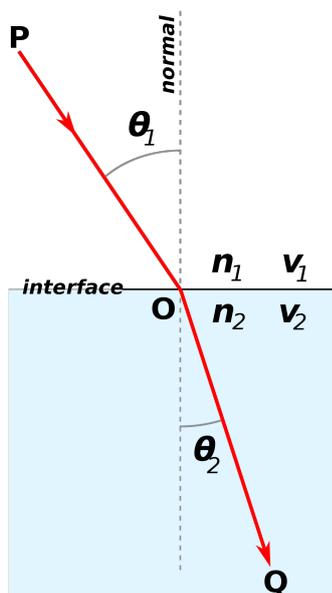


Figure 1.3 – Illustration of Snell's law. (illustration available on Wikimedia Commons under the filename 'Snells law2.svg', created and released in public domain by contributor Sawims)

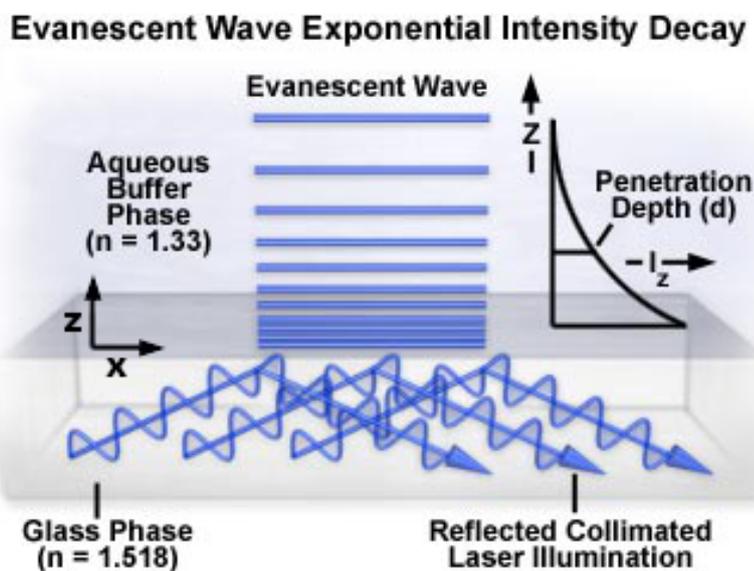


Figure 1.4 – The intensity of the evanescent wave generated at the interface between two glass and aqueous phases decays exponentially with distance from the glass phase. (illustration extracted and edited from [Axelrod and Davidson n.d.], available on [www.olympus-lifescience.com](http://www.olympus-lifescience.com))

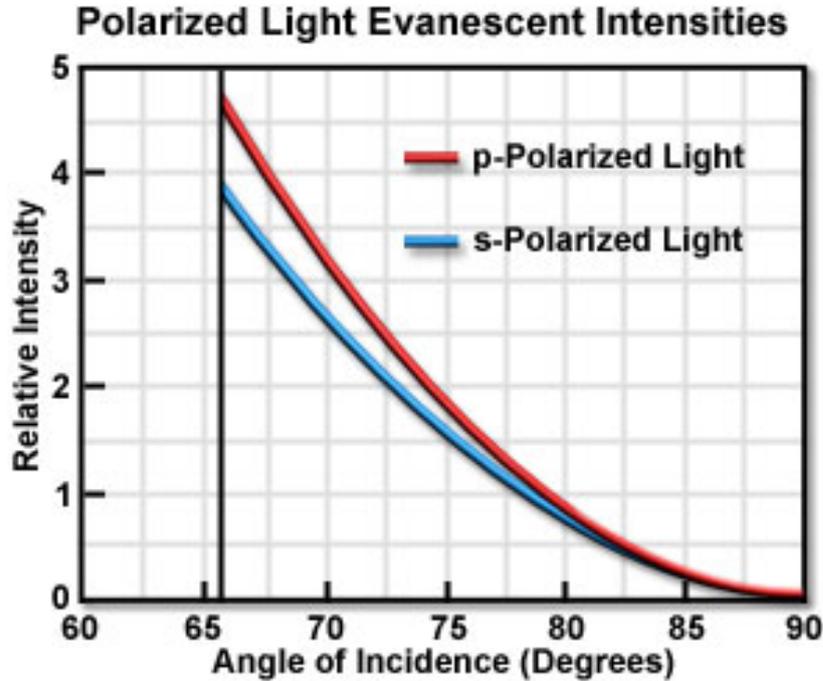


Figure 1.5 – The evanescent intensities of polarized light relative to the incident intensity, as functions of the angle of incidence. Light passes through an interface composed of fused quartz and water (or an aqueous buffer solution) of refractive indexes  $n(1) = 1.46$  and  $n(2) = 1.33$  respectively, with a critical angle of  $\theta_c = 65.7^\circ$ . P and s-polarized evanescent intensities can be several times greater than the incident intensity for angles within 15 degrees of the critical angle. (illustration extracted and edited from [Axelrod and Davidson n.d.], available on [www.olympus-lifescience.com](http://www.olympus-lifescience.com))

to the interface. This field is called the evanescent field, and its intensity  $I$  exponentially decays along the  $z$ -direction, normal to the interface:

$$I = I_0 e^{-z/d}, \quad (1.27)$$

where  $I_0$  is the intensity of the evanescent field at the interface and  $d$ , called penetration depth, is a function of the wavelengths of incident light  $\lambda_0$  and the refractive indices  $n_1$  and  $n_2$ :

$$d = \frac{\lambda_1}{4\pi} \left( n_1^2 \sin^2 \theta_1 - n_2^2 \right)^{-1/2}.$$

The intensity decay phenomenon of the evanescent field is illustrated in Figure 1.4.

It is possible to deduce the polarization properties of the evanescent field using Fresnel's equations (see details in [Martin-Fernandez, Tynan, and Webb 2013]). Notably, it

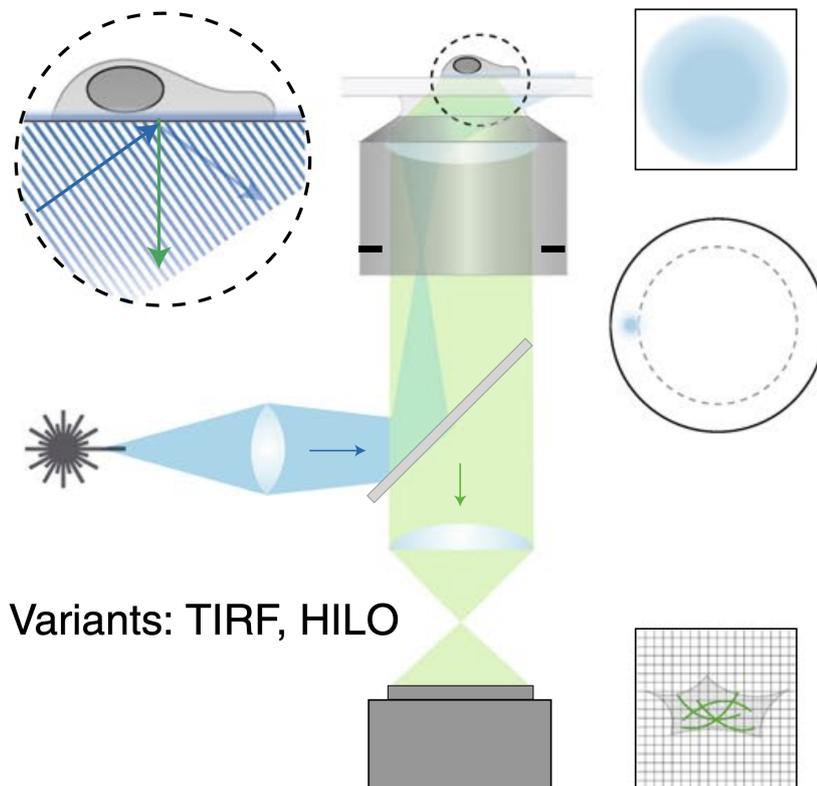


Figure 1.6 – Principle and light path of TIRF Microscopy. By focusing the excitation beam on the edge of the pupil plane, the light strikes the coverslip interface at a supercritical angle, generating an evanescent wave. (illustration extracted and edited from [Schermelleh et al. 2019], Fig. 1(b))

can be shown that with polarization,  $I_0$  can be several times greater than the intensity of the incident wave as illustrated in Figure 1.5.

Axelrod, Burghardt and Thompson were the first to take advantage of this phenomenon in fluorescence microscopy [Axelrod, Burghardt, and Thompson 1984]. In TIRF microscopy, the microscope's glass coverslip and the aqueous medium constituted by the specimen in water act as the two interfaced media. As illustrated in Figure 1.6, a laser beam is directed at the interface with an angle of incidence greater than the critical angle  $\theta_c$ . A small, flat portion of the specimen is thus illuminated by the evanescent wave generated near the coverslip, in a range of 100 – 200 nm which is smaller than the conventional resolutions. This technique is therefore much less phototoxic than WFFM and allows for longer time acquisitions if new fluorophores enter the evanescent field. Also, planar illumination delivers high contrast as out-of-focus activation of fluorophores is low. Unlike confocal microscopy, no scanning process is involved and live-imaging is possible,

enabling high temporal resolution. These advantages, combined with the low cost of the setup, make TIRF microscopy a technique of choice for the study of cell membranes.

TIRF microscopy is also adaptable in three dimensions by recording changes in intensity within the evanescent wave [Allersma et al. 2004], by using Differential Nanometry (DiNa) [Saffarian and Kirchhausen 2008] or with Simultaneous Two-wavelength Axial Ratiometry (STAR) [Stabley et al. 2015]. But one of the most popular method is 3D Multi-Angle TIRF microscopy (3D MA-TIRFM) [Loerke, Stühmer, and Oheim 2002] [Boulanger et al. 2014] [Soubies et al. 2019]. Using multiple 2D images acquired with variable incident angles  $\theta_1$  as an input, it is able to reconstruct 3D stacks generally by solving a pseudo-deconvolution problem by considering a dedicated transfer function depending on incident angle. It is a high-speed technique able to generate stacks with a resolution of 30 – 50 nm along the axial direction versus 200 nm in lateral direction, and up to 500 nm in depth. The aforementioned reconstruction problem becomes very challenging, especially because we aim to achieve a high axial resolution ( $\sim 50$  nm) of up to 1000 nm in depth. The approach implies the sequential acquisition of MA-TIRF stacks while decreasing frame rate. It is also worth noting that calibrating the penetration depth of the evanescent wave can increase the precision of the technique [Oheim et al. 2019].

### **Light-Sheet Fluorescence Microscopy (LSFM)**

Light Sheet Fluorescence Microscopy (LSFM), whose principle is illustrated in Figure 1.7, consists of illuminating the specimen orthogonally to the direction of observation with a thin sheet of light [Stelzer et al. 2021].

As illustrated in Figure 1.7, two methods co-exist to generate the light sheet, which in practice is most often approximated by a Gaussian beam. In selective/single plane illumination microscopy (SPIM), a cylindrical lens is used to focus the excitation laser beam into the  $z$  axis. Another method called Digital Scanned Laser Microscopy (DSLM) generates a dynamic light sheet by rapidly scanning the excitation laser in the lateral axis thanks to motorized mirrors. In both cases, only the fluorophores present within the focal volume of the microscope are illuminated.

The generation of 3D images, illustrated in Figure 1.8, is performed by translating the sample through the light sheet along the axial axis, or by moving the light sheet through the specimen. 2D images are therefore acquired at a regular interval and can be stacked to form a 3D image. In the case of large and opaque specimen in which the laser beam cannot penetrate, the scanning can be performed in multiple directions by

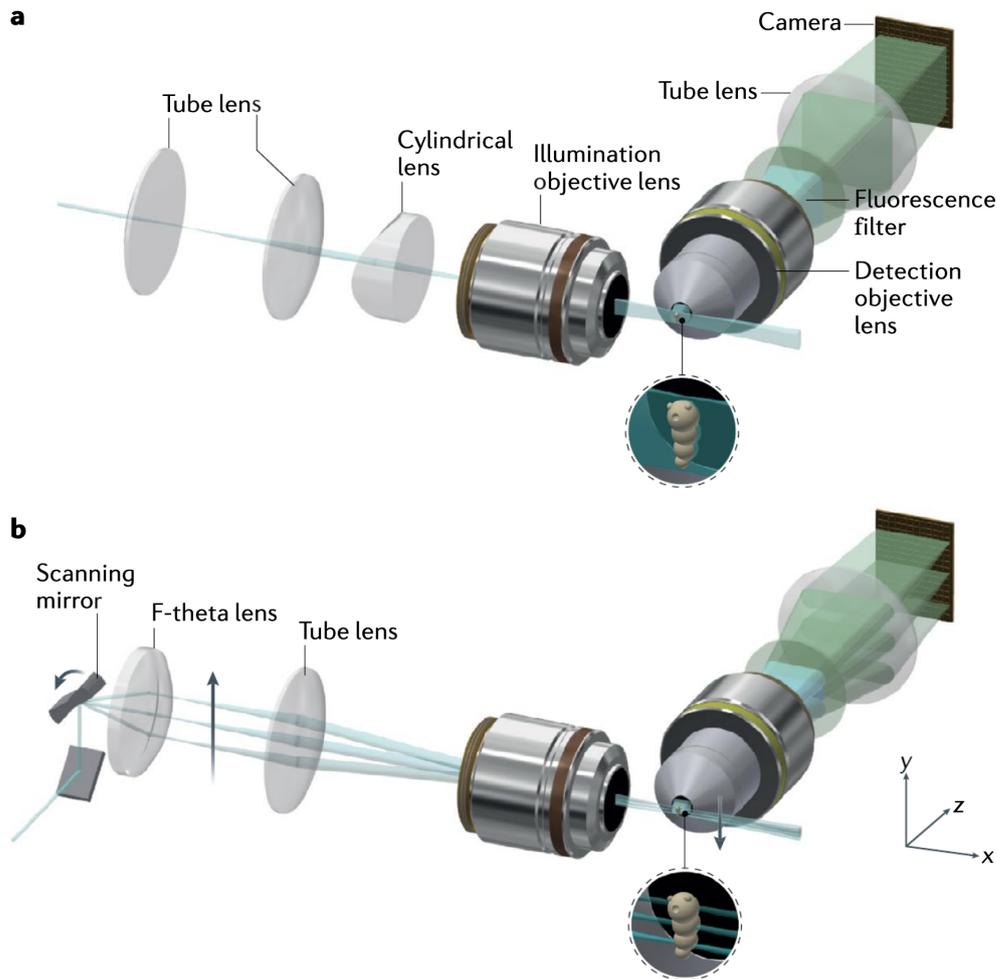


Figure 1.7 – Principle and implementations of Light Sheet Fluorescence Microscopy. **(a)** A Selective/Single Plane Illumination Microscope (SPIM). A cylindrical lens focuses the laser beam along the  $z$  axis, creating a static light sheet that illuminates the field of view. **(b)** A Digital Scanned Laser Microscope (DSLIM). A mirror/lens assembly focuses the laser beam into a single spot that is scanned across the volume to create a dynamic light sheet. (illustration extracted from [Stelzer et al. 2021])

physically rotating the specimen. Finally, constant axial resolution over the field of view can be achieved by moving the light sheet along the lateral axis and combining several acquisitions into a single image by computational means.

The true optical sectioning capability of light sheet microscopes allows for axial resolution and reduced photobleaching and phototoxicity since only a fraction of the specimen with focal range is illuminated. For comparison, the specimen is irradiated three to five times less than with confocal microscopy. This technique also offers good contrast as out-

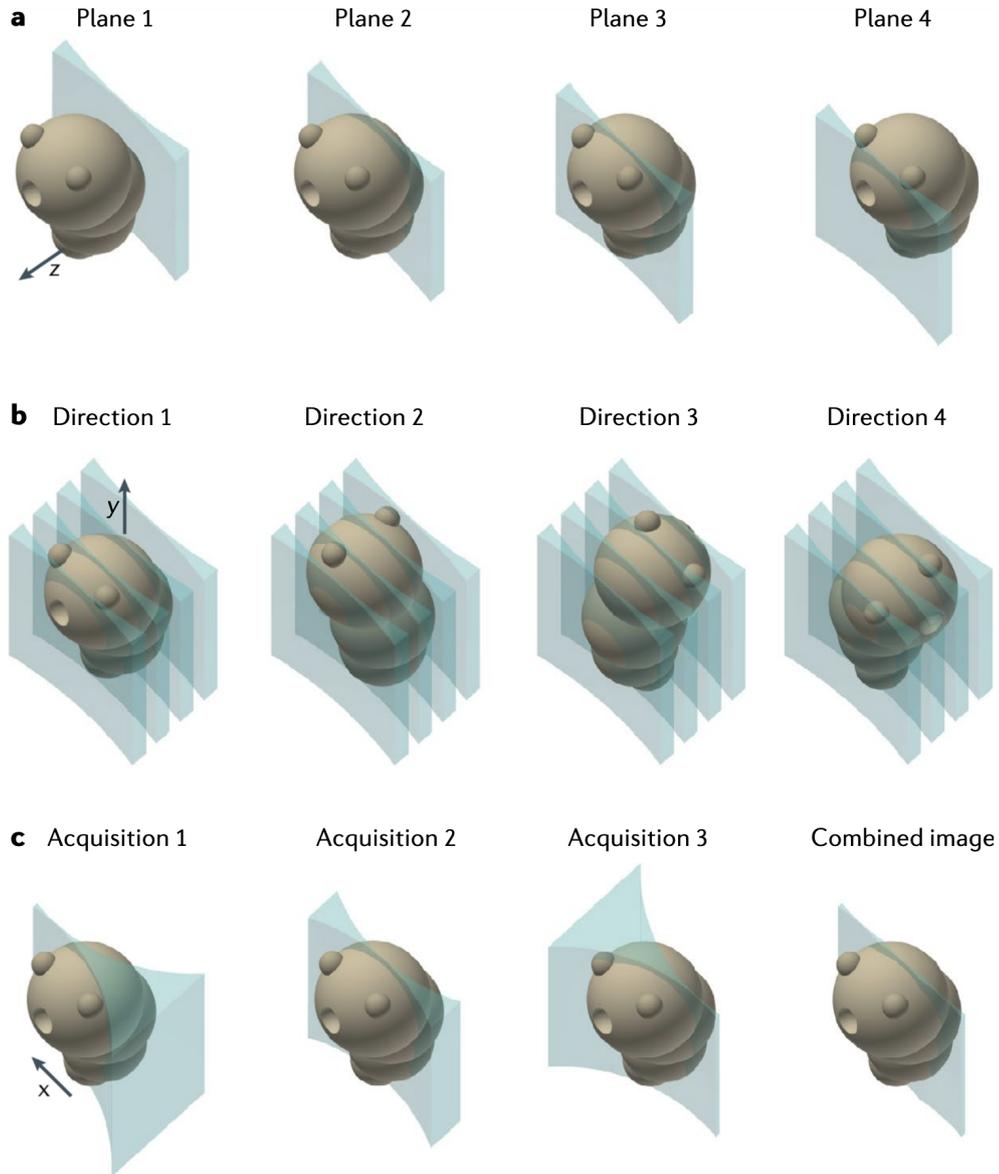


Figure 1.8 – 3D image generation with LSFM. **(a)** By moving the specimen or light sheet relative to each other along the  $z$  axis, it is possible to acquire a stack of 2D images plane-by-plane, that together form a 3D image. **(b)** When the specimen is too large or dense to be penetrated by the laser beam, multidirectional imaging can be performed using a rotary stage that rotates the specimen around the  $y$  axis, allowing 3D image stacks to be acquired in multiple directions. **(c)** By moving the light sheet along the  $x$  axis and combining multiple acquisitions into a single image, a constant axial resolution can be achieved over a large field of view. (illustration extracted from [Stelzer et al. 2021])

of-focus signal is almost nonexistent. Furthermore, the use of cameras enables millions of

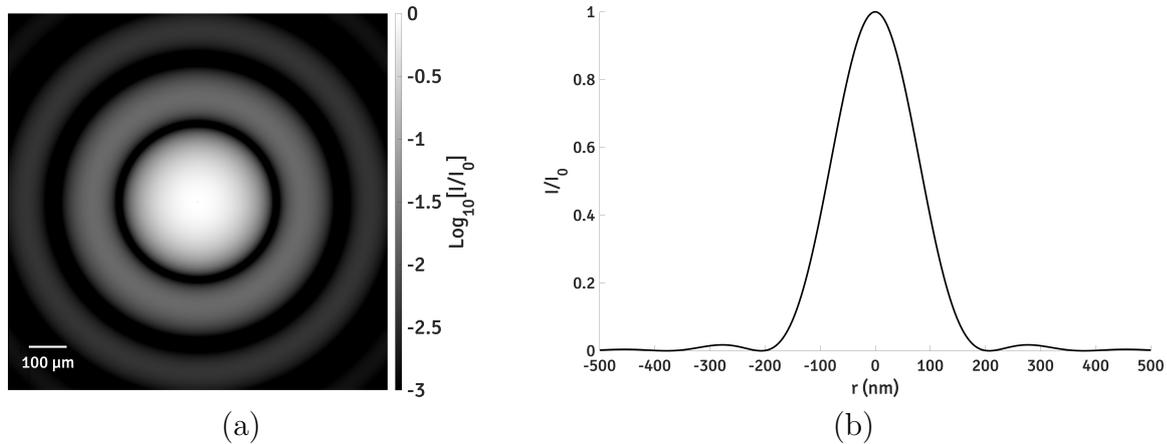


Figure 1.9 – The Airy pattern generated by a fluorophore through a perfect optical system, with  $\lambda = 509$  nm (GFP emission wavelength) and  $NA = 1.5$ .  $I$  designates the signal intensity, and  $I_0$  is its maximum at the center of the pattern. **(a)** Lateral view of the Airy pattern, as seen by the observer. Logarithmic scale is used to make the concentric ring more visible. In practice, the central Airy disk concentrates the  $\sim 84\%$  of the visible signal. **(b)** Relative intensity values of the radial cross-section of the Airy pattern represented in **(a)**. (illustration plotted according to the Point-Spread Function (PSF) equation in [Zhang, Zerubia, and Olivo-Marin 2007])

voxels to be acquired in parallel, meaning that tens to thousands of images can be recorded in a matter of seconds. These advantages make LSM a highly adaptable technique for 3D imaging in biomedical fields.

### 1.3.2 The diffraction limit problem and its solutions

In this section, we explain how the optical phenomenon of diffraction creates a resolution limit for conventional fluorescence microscopy techniques, and how this limit can be overcome.

The Point Spread Function (PSF) of an optical system (such as a microscope) refers to the image it gives of an infinitely small light source (point source) when focused on it. In particular, a perfect lens with circular aperture focused on a point source produces a pattern called Airy pattern [Airy 1835], composed of a central light spot designated as Airy disk, surrounded by concentric rings (see Figure 1.9). The Airy disk concentrates  $\sim 84\%$  of the total light intensity coming from the emitter. The PSF of an optical system is caused by diffraction which is the product of light wave interference in the image plane.



Figure 1.10 – Simulation of the PSF generated by a microscope with  $\lambda = 557 \text{ nm}$  and  $\text{NA} = 1.3$ , as observed in the axial plane. To improve the visualization of the diffraction rings, the image was contrasted with a factor  $\gamma = 0.5$ . (illustration extracted and edited from [Cole, Jinadasa, and Brown 2011])

In 3D, the PSF takes the form of a ellipsoid, stretched along the axial axis as illustrated in Figure 1.10. In space-invariant systems where the PSF stays the same everywhere, fluorescent images are the result of the convolution of the PSF with the intensity of the object. It is common to approximate the PSF by a Gaussian distribution [Zhang, Zerubia, and Olivo-Marin 2007].

The Rayleigh criterion tells us that two distinct light sources can be resolved if they are separated by at least the radius of the Airy disk, which value is given by  $r = 0.61\lambda/\text{NA}$ , where  $\lambda$  is the wavelength of the light emitted by the light source, and  $\text{NA}$  is the numerical aperture of the objective [Rayleigh 1879]. This implies that any conventional microscope using visible light has a resolution limit of  $\sim 200 - 300 \text{ nm}$  laterally (axially, the reso-

lution is limited to  $\sim 500 - 700$  nm), no matter its degree of perfection. In fluorescence microscopy, fluorescently tagged particles are between 1 to 10 nm in size, but appear as bright spots, and cannot be distinguished if separated by less than this resolution limit. Figure 1.9 illustrates a case where a particle tagged with GFP, which has an emission wavelength of  $\lambda = 509$  nm, is observed under a lens of numerical aperture of  $\text{NA} = 1.5$ , producing a spot of  $\sim 200$  nm radius.

The diffraction limit can be pushed back by different means. The use of large numerical apertures, such as  $\text{NA} = 1.5$ , attenuates the diffraction phenomenon by reducing the radius of the Airy disks. Also, computational tools such as deconvolution algorithms [Sibarita 2005] enable to counteract the diffraction if the PSF is known. In practice however, its efficiency is limited by the presence of noise in the images. To fully overcome the diffraction limit, it is necessary to use cutting-edge technology such as super-resolution, using PSF models to localize the positions of single particles ([Huang et al. 2008], [Shroff, White, and Betzig 2013]). We detail some super-resolution techniques in the next section.

### 1.3.3 Super-resolution microscopy techniques

Breaking the optical resolution beyond the diffraction limit while preserving the advantages of fluorescence microscopy and the specificity of molecular imaging has been a long-standing goal. In this section, we present a general overview of the most popular super-resolution microscopy (SRM) techniques that allow to achieve this goal. More detailed presentations of SRM techniques are available with [Liu, Lavis, and Betzig 2015] and [Schermelleh et al. 2019], with a complete overview both commercially available and emerging techniques and their characteristics, as well as an evaluation of their advantages and weaknesses, specified biological applications and future development paths.

SRM can be broadly separated in two main groups: Structured Illumination Microscopy (SR-SIM) and diffraction-unlimited SRM, which notably includes Stimulated Emission Depletion microscopy (STED) and Single Molecule Localization Microscopy (SMLM). These techniques are based on classic WFFM, confocal or TIRF setups, including more sophisticated technologies such as multifocus microscopy [Abrahamsson et al. 2013] [Hajj, Wisniewski, et al. 2014] [Hajj, El Beheiry, and Dahan 2016]. Additionally, LSM can be adapted in super resolution with the notable example of Lattice Light Sheet Microscopy (LLSM).

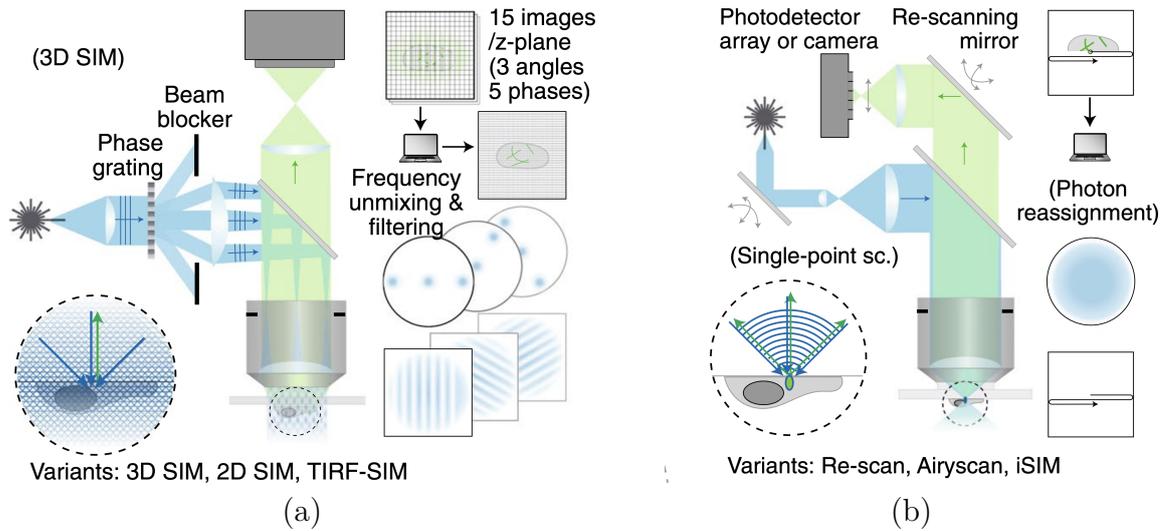


Figure 1.11 – Principle and light path of Structured Illumination Microscopy. **(a)** Interference-based SIM. **(b)** Point-scanning SIM. (illustration extracted and edited from [Schermelleh et al. 2019], Fig. 1(d) and (e))

### Structured Illumination Microscopy (SIM)

Structured Illumination Microscopy (SIM) [M. G. Gustafsson 2000] [M. G. Gustafsson et al. 2008] consists in interfering laser beams to illuminate the specimen with a series of structured light stripe patterns, the minimum distance between stripes nearing the resolution limit. These stripes interact with conventionally non-observable ‘high frequency’ sample features, producing larger-scale interference similar to Moiré effects that can be observed in addition to conventional wide-field image. These contain encoded information about high spatial frequency features of the specimen. A series of images is acquired with translated and rotated stripes to gather high frequency information in all lateral directions. This information is then decoded by computational means and reassembled in frequency space to reconstruct a contrast-enhanced image. Interference-based SIM can be based on both WFFM and TIRF microscopy, but a variant called point-scanning SIM uses confocal microscopy as a base setup, as illustrated in Figure 1.11.

This technique is able to double of the resulting resolution compared to diffraction-limited microscopy techniques, with 100 - 130 nm laterally and 300 - 400 nm axially in the case of 3D SIM, as illustrated in Figure 1.12. Combined with TIRF microscopy and ultra-high apertures such as  $NA = 1.7$ , the lateral resolution of linear SIM can be further

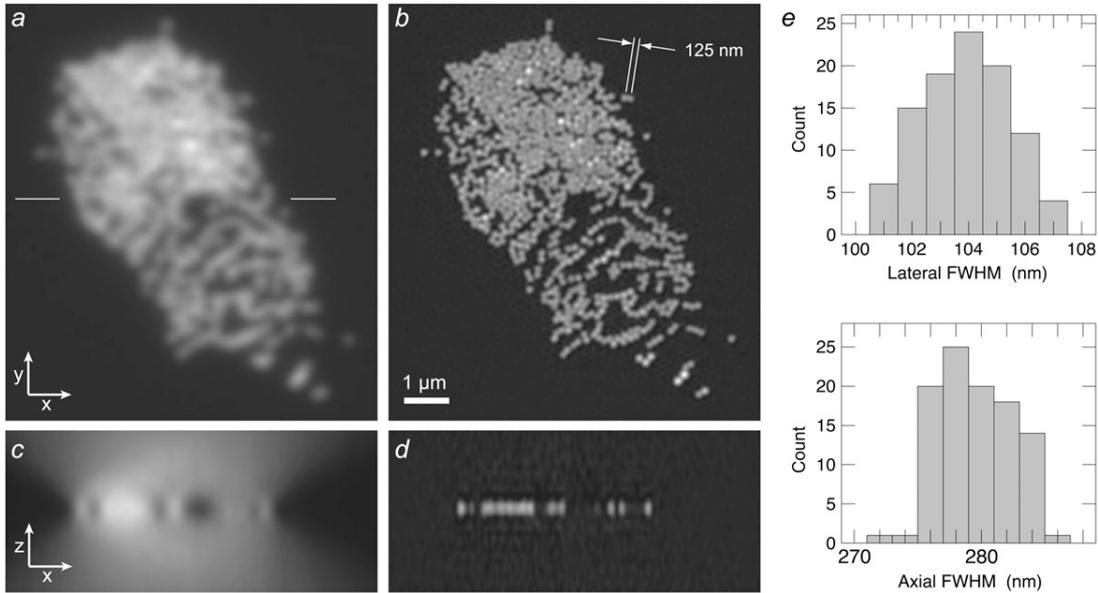
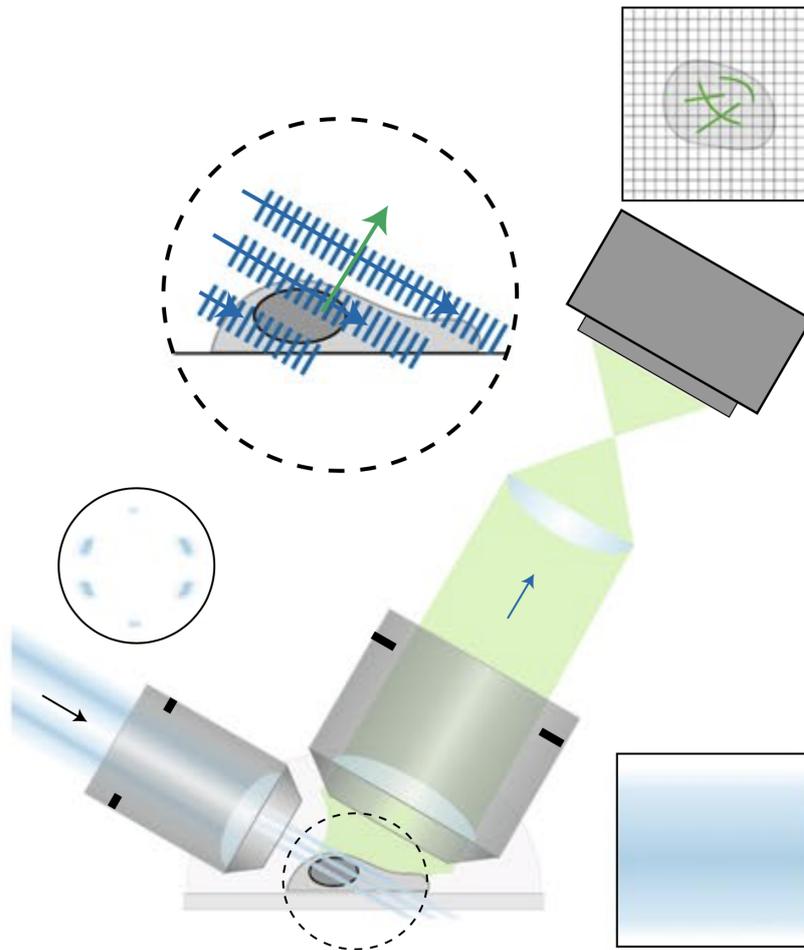


Figure 1.12 – Comparison of 3D images of red-fluorescent microspheres of diameter  $0.12\ \mu\text{m}$ , illustrating the resolution and contrast improvement between WFFM and Structured Illumination Microscopy. **(a)** Lateral view of the microspheres obtained through WFFM. **(b)** Lateral view of the microspheres obtained through SIM. **(c)** Axial view of the microspheres obtained through WFFM. **(d)** Axial view of the microspheres obtained through SIM. **(e)** Histograms of lateral and axial Full Width at Half-Maximum (FWHM) of green-fluorescent beads observed by 3D SIM. (illustration extracted from [M. G. Gustafsson et al. 2008])

improved to  $\sim 80\ \text{nm}$ . It is technically possible to reach even higher resolutions with specific methods such as reversible photoswitching non-linear SIM (NL-SIM) or parallelized RESOLFT [Rego et al. 2011] [Chmyrov et al. 2013].

SIM is particularly adapted to live-cell imaging and high throughput microscopy. It is a very photon-efficient approach, enabling multiple color imaging and 3D observation with conventional fluorophores. On the negative side, it requires to carefully align and calibrate the microscope setup and mathematical post-process the obtained data. This increases the risk of apparition of reconstruction artifacts, which require in-depth knowledge to detect and neutralize. Additionally, even if they technically break the conventional resolution limit, SR-SIM based methods remain fundamentally linked to the laws of diffraction. Rather than suppressing them, these methods simply push back the limits of conventional techniques.



## Variants: Bessel beam, LLS-SIM

Figure 1.13 – Principle and light path of Lattice Light Sheet Microscopy. (illustration extracted and edited from [Schermelleh et al. 2019], Fig. 1(h))

### Lattice Light-Sheet Microscopy (LLSM)

Lattice Light-Sheet Microscopy (LLSM, illustrated in Figure 1.13) [Chen et al. 2014] is an improved variant of LSFM that uses an ultra-thin light sheet to illuminate the specimen. As a matter of fact, conventional LSFM setups use Gaussian laser beams to generate the light sheet, which remain too thick for subcellular imaging (see Figure 1.14(a)). A first response to this limitation was the introduction of Bessel beams, generating thinner sheets of submicrometer thickness (see Figure 1.14(b)). LLSM goes one step further, generating sheets of light from 2D optical lattices (see Figure 1.14(c) and (d)). Optical lattices are

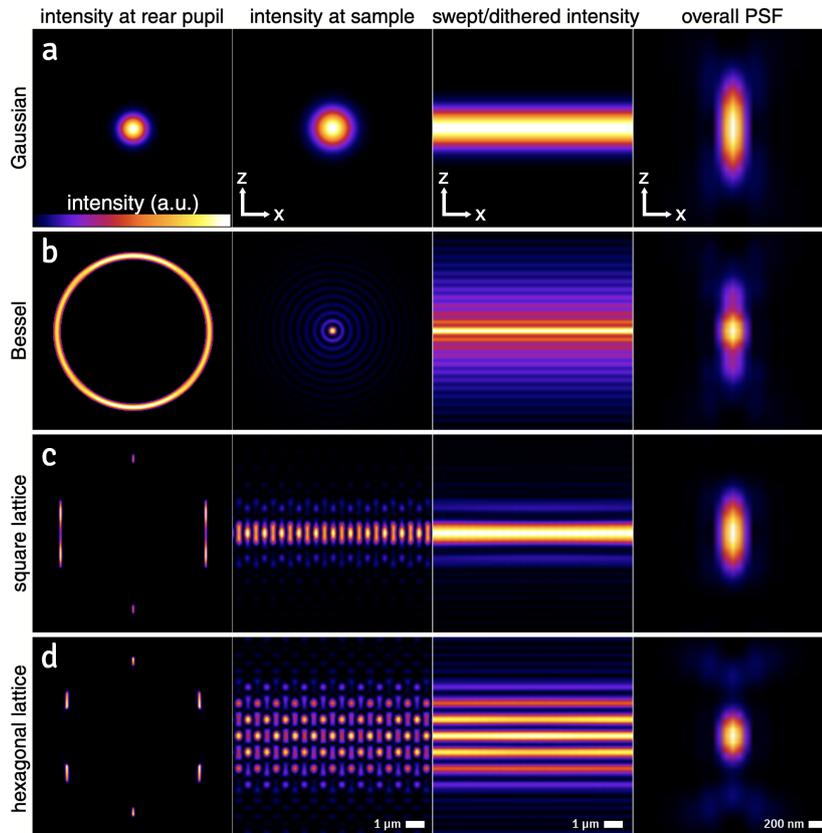


Figure 1.14 – Comparison of light sheet generation methods. **(a)** The conventional LSFM approach sweeps a Gaussian beam across a plane to create the light sheet. **(b)** The use of a Bessel beam generates a sheet with a narrow core, albeit surrounded by a large number of bands. **(c and d)** Optical lattices create periodic interference patterns across the plane, with low peak intensity and low phototoxicity. The square lattice in **(c)** confines the excitation in the central plane, while the hexagonal lattice in **(d)** optimizes axial resolution. The columns from left to right depict: the intensity pattern at the rear pupil plane of the excitation objective (in arbitrary units); the intensity pattern at the  $xz$  plane of the excitation objective focus; the dithered or swept focal intensity pattern creating the light sheet; and the axial cross-section of the microscope’s overall PSF, respectively. (illustration extracted and edited from [Chen et al. 2014])

2D or 3D periodic interference patterns that can be generated by coherently superposing plane waves. In particular, they have notably been used as a mean to trap atoms at very low temperatures. In LLSM, 2D optical lattices can either be dithered to generate a uniform light sheet in high speed mode, or stepped through the specimen in SIM mode to go beyond the diffraction limit.

The resulting ultra-thin light sheet enables high axial resolution with very low back-

ground signal compared to conventional LSFM. LLSM is capable of resolving at the diffraction limit and beyond, with a quasi-isotropic resolution of  $230 \times 230 \times 370$  nm in dithered mode and  $150 \times 230 \times 280$  nm in SIM mode. Compared to Bessel beam LSFM or spinning disk confocal microscopy, the low-peak excitation intensity of LLSM greatly reduces photo-bleaching and photo-toxicity, enabling acquisitions over longer periods of time. LLSM also achieves high acquisition speeds, recording from 40 to 200 2D planes per second, enabling hundreds or even thousands of 3D images separated by subsecond intervals to be imaged. However, high-speed imaging of 3D images results in large amounts of generated data, which requires cutting-edge image processing algorithms capable of operating at live-cell recording regimes. In addition, the assembly constraints of LLS microscopes result in a particular angle of the detection objective, making it necessary to "deskew" the recorded images before processing them.

### **Diffraction-unlimited super-resolution microscopy**

Contrary to SR-SIM, diffraction-unlimited SRM techniques (also referred to as nanoscopy) are theoretically able to suppress the diffraction limit, pushing resolution towards infinitesimally small scales. Their basic principle is to switch fluorescence emissions on and off. Deterministic techniques such as Stimulated Emission Depletion (STED) target the specimen with focused laser beams to trigger the on/off switching, while others such as Single Molecule Localization Microscopy (SMLM) stochastically induce switching via wide-field illumination, using algorithmic tools to detect events and reconstruct the image. However, in experimental practice and particularly in live-cell imaging, strong irradiation levels, fluorophores density, and long acquisition times limit the achievable resolution.

### **Stimulated Emission Depletion (STED)**

Stimulated Emission Depletion (STED) is probably the most common of diffraction-unlimited SRM techniques. Its principle was first patented by Okhonin [Okhonin 1991], before being developed again by Hell and Wichmann [Hell and Wichmann 1994], unaware of the existence of the first patent, and was first experimentally demonstrated by Klar and Hell [Klar and Hell 1999]. A standard STED setup consists of a confocal microscope where a depletion beam overlays the conventional excitation beam, as illustrated in Figure 1.15. This depletion beam features at least one local intensity minimum, usually positioned in the focal center (donut shape). Its role is inhibit or deplete fluorescent probes to restrict the excitation zone(s) to a sub-diffraction scale. STED offers the possibility of adjusting

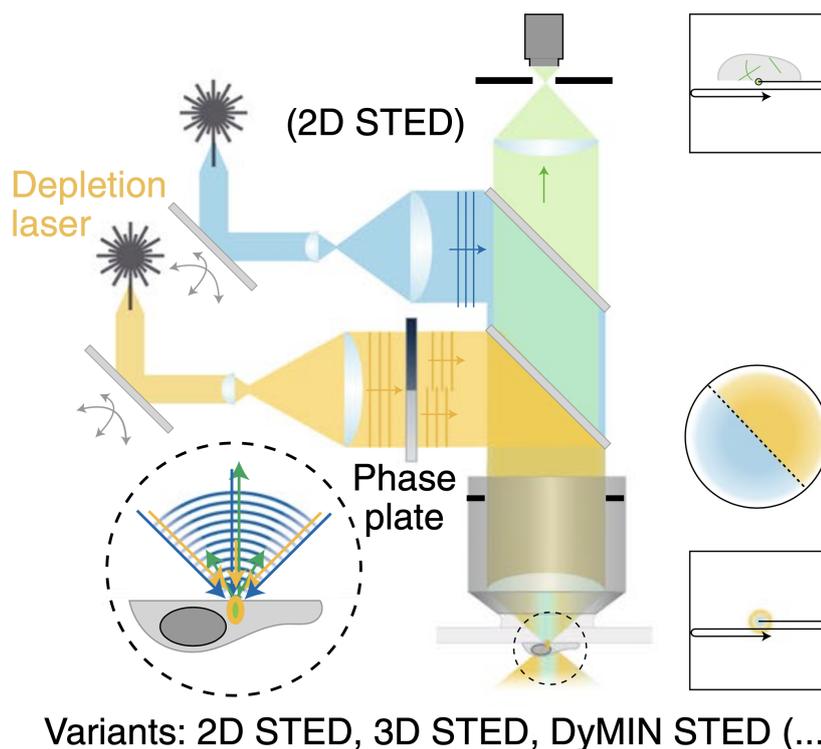


Figure 1.15 – Principle and light path of Stimulated Emission Depletion microscopy. (illustration extracted and edited from [Schermelleh et al. 2019], Fig. 1(f))

the resolution by tuning its laser power. This allows the user to weigh up the balance between resolution and photo-damaging effects, which can be particularly useful in live-cell imaging when used with custom fluorescent tags and optimized scanning protocols. Live cells can also be imaged using reversible, saturable optical linear fluorescence transitions (RESOLFT) microscopy, which uses reversibly photoswitchable fluorescent tags.

In terms of performance, commercial STED installations are capable of achieving lateral resolution of 60 - 120 nm in cells, while specialized laboratories and teams can go down to 30 - 80 nm. 3D STED setups are available, allowing a free choice between increasing lateral and axial resolution. It is also possible to perform two-color imaging (and even more channels through the conventional confocal mode) with a wide range of fluorophores, the best performance being obtained with specifically STED-optimized dyes.

STED is generally considered as easy to use as standard confocal microscopes, of which it is a complementary modality. While direct observation is impossible, computational post-processing is not an absolute requirement, although deconvolution is often useful. In

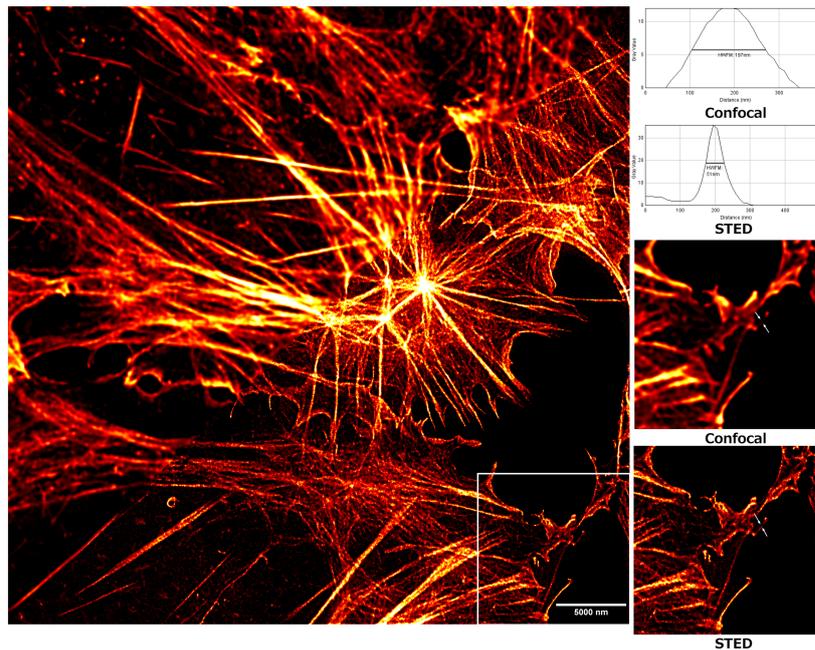


Figure 1.16 – Actin filaments in cells stained with 647-phalloidin. Stimulated emission depletion (STED) microscopy provides significant resolution improvements over those possible with confocal microscopy. (illustration created and released in CC-BY-SA-4.0 by Howard Vindin, available on Wikimedia Commons under the filename 'STED Confocal Comparison 50nm HWFM.png')

2D, STED is particularly suited to imaging filamentous or vesicular structures such as those illustrated in Figure 1.16, thanks to its exceptional lateral resolution. In volumetric imaging, 3D STED offers excellent performance when it comes to render thick and densely packed structures. On the other hand, STED has the same drawbacks as conventional confocal microscopy when it comes to point scanning, as acquisition times increase with sample size. This translates, for example, into a relatively slow imaging speed for entire cells. The reduction of effective fluorescence volume further accentuates this problem, as it shrinks the size of the scanning steps, in addition to decreasing the amount of signal detected.

### Single Molecule Localization Microscopy (SMLM)

Single Molecule Localization Microscopy (SMLM), illustrated in Figure 1.17, is a family of unlimited-SR techniques that achieve super-resolution under wide-field or TIRF illumination by stochastically activating (or switching on and off) fluorophores in space

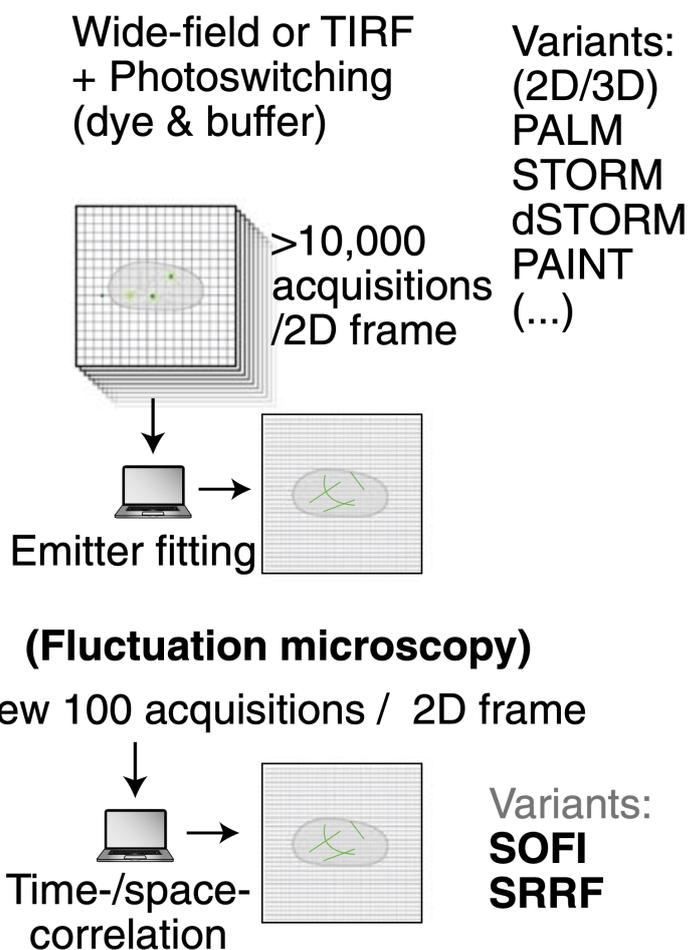


Figure 1.17 – Principle of Single Molecule Localization Microscopy. (illustration extracted and edited from [Schermelleh et al. 2019], Fig. 1(g))

and time. This enables the acquisition of long image sequences with very sparse fluorescence signal, which can be computer-processed to extract the precise localization of single molecules from their isolated PSF. 3D localization is possible through PSF engineering, multiplane or interferometric approaches, or even a combination of these techniques [Hajj, El Beheiry, Izeddin, et al. 2014] [Abrahamsson et al. 2013] [Hajj, Wisniewski, et al. 2014] [Hajj, El Beheiry, and Dahan 2016]. The obtained precision is nanometric and depends mainly on the amount of signal detected per activated fluorophore. The concatenation of positions detected over thousands of frames enables a pointillistic reconstruction of the observed specimen.

There exists a large number of SMLM techniques, differing mainly in the way switching is performed. The list includes, but is not limited to, Photo-Activated Localization Mi-

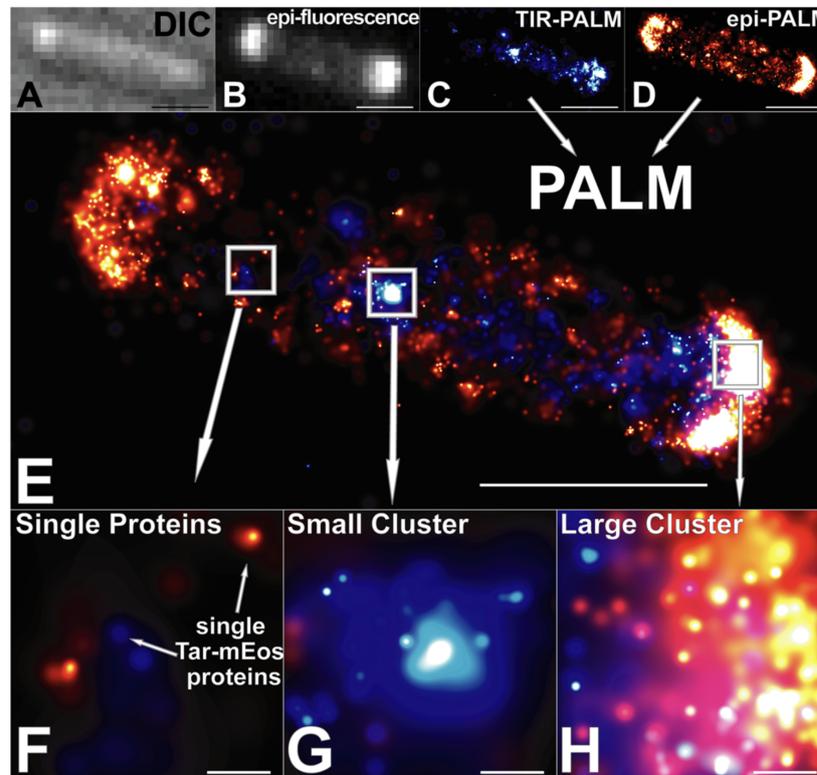


Figure 1.18 – *E. coli*  $\Delta tar$  cell with mEos-labeled Tar. (A), (B), (C) and (D) depict Differential interference contrast (DIC), Diffraction-limited epi-fluorescence (epi), TIR-PALM, and epi-PALM images of a single cell, respectively. (E) Superposition of (C) and (D). (F), (G) and (H) depict zooms on the left, middle and right boxed ROIs of (E), respectively. The scale bar in (A–E) indicates 1  $\mu\text{m}$  and the scale bar in (F–H) indicates 50 nm. (illustration extracted from [Greenfield et al. 2009])

croscopy (PALM, comparative performances illustrated in Figure 1.18) [Betzig et al. 2006], Stochastic Optical Reconstructed Microscopy (STORM) [Rust, Bates, and Zhuang 2006] and direct STORM (dSTORM) [Endesfelder and Heilemann 2015], Binding-Activated Localization Microscopy (BALM) [Schön et al. 2011], Point Accumulation for Imaging in Nanoscale Topography (PAINT) [Sharonov and Hochstrasser 2006] and a number of variants. Combining SMLM with TIRF or HILO (Highly Inclined and Laminated Optical sheet) techniques can significantly improve the efficiency of detection and the signal-to-background ratio.

SMLM microscopes are generally capable of achieving very high optical resolution reaching 20 nm laterally and 50 nm axially. The photon count is the main factor on which the localization precision of the detected PSF centroid depends. In the optical axis, the

focal depth of the image plane is a limiting factor that can be overcome by the addition of astigmatic/helical optical distortions or bi-plane detection techniques. These are able to improve optical resolution to below 100 nm in most cases, at the expense of lateral precision. On its part, structural resolution (such as the ability to distinguish biological features like filaments) depends on the density of the specimen labeling and switching properties.

In addition to their very high resolution, the main advantage of SMLM techniques is their low cost and their ability to be implemented on conventional wide-field microscopes, including Light Field Microscopy (LFM) setups [Sims et al. 2020]. The overall operating complexity is however shifted to the preparation of the biological specimen, as well as the data analysis and the reconstruction process. Regarding the last point, high label densities or inappropriate photoswitching rates can for instance induce false detection of emitters. Other technical requirements include specifically photoswitchable/activatable fluorescent labels (e.g. in PALM) or special blinking-inducing buffer conditions for conventional dyes (e.g. in (d)STORM). The speed of SMLM is generally severely limited by the long acquisition times associated with the thousands of camera frames required to reconstruct a single image plane. Additionally, these acquisition times must be short enough so that the motion of the observed structure does not generate artifacts on the image reconstruction. These limitations strongly inhibit the applicability of SMLM for live cell imaging, although methods such as fluctuation analysis, Super-resolution Optical Fluctuation Imaging (SOFI) [Dertinger et al. 2009] or Super-Resolution Radial Fluctuations (SRRF) [N. Gustafsson et al. 2016] offer to make a trade-off between optical and temporal resolution, enabling live-cell recordings.

## 1.4 Bioparticle tracking techniques

In the previous section, we presented various fluorescence microscopy techniques as a means of imaging microscopic objects in 2D and 3D. To extract useful information from the images, it is often necessary to apply processing steps. In biological microscopy, image processing and analysis have become essential tools for automatically and reliably quantifying sub-cellular and molecular biological processes in space and time. With data collection becoming increasingly complex, as do the techniques used to acquire it, the signal and image processing algorithms used to perform these operations are gaining in importance.

In particular, computational methods for analyzing the motion of bioparticles, which has been described individually and collectively by a number of models [Bressloff and Newby 2013] [Hozé and Holcman 2017], are of particular interest in cell biology. The motivation for developing such tools stems from the complexity of the dynamic processes observed, such as protein association, dissociation and recombination. Most existing methods are probabilistic tracking techniques for fluorescent bioparticles [Chenouard, Smal, et al. 2014]. Tracking bioparticles in both 2D and 3D image sequences with high precision is a difficult task, requiring optimization or simulation to handle the estimation of several thousand tracks [Sbalzarini and Koumoutsakos 2005] [Smal et al. 2008] [Jaqaman et al. 2008] [Chenouard, Bloch, and Olivo-Marin 2013] [Sibarita 2014] [Racine, Hertzog, et al. 2006] [Racine, Sachse, et al. 2007] [Roudot, Ding, et al. 2017]. Single-Particle Tracking (SPT) methods generally attempt to optimally associate the position of detected fluorescent spots with appropriate algorithms, while overcoming non-linear constraints [Chenouard, Bloch, and Olivo-Marin 2013]. Several methods have been designed to handle specific types of bioparticle motion [Chenouard, Bloch, and Olivo-Marin 2013] [Spilger et al. 2020] and were evaluated in 2014 [Chenouard, Smal, et al. 2014]: all trackers were highly competitive, although giving their best on their target scenario (see [Pécot et al. 2018]).

In this section, we focus on two tracking methods that were used in our work: Multiple Hypothesis Tracker [Reid 1979] [Blackman 2004] [Chenouard, Bloch, and Olivo-Marin 2013] and U-TRACK [Jaqaman et al. 2008] [Roudot, Ding, et al. 2017].

### 1.4.1 Multiple Hypothesis Tracking (MHT)

Multiple Hypothesis Tracking is one of the most popular methods for tracking multiple targets, and is used in a wide variety of fields such as radar detection, video surveillance or fluorescence microscopy. It is based on the idea of forming at each frame several linking hypotheses between the positions of detected targets (such as fluorescent spots), constructing probability trees of possible complete trajectories. This idea of multiple hypothesis propagation was originally developed in [Singer, Sea, and Housewright 1974], before being algorithmically implemented in [Reid 1979] for civil and military radar purposes. It has since been derived in a number of variants designed for various fields [Blackman 2004].

In the past decade, it has been adapted to fluorescence microscopy in [Chenouard, Bloch, and Olivo-Marin 2013], to make it capable of simultaneously tracking thousands of targets in biological image sequences. To this end, the authors propose a unified Bayesian

framework suitable for nanoscale target tracking in cellular environments, incorporating realistic models for target existence, particle motion and fluorescent image features. They then propose to adapt the Multiple Hypothesis Tracking algorithm to the proposed framework, in order to handle the large scale of the problem with a favorable trade-off between computational cost and model complexity. Track trees provide an exhaustive representation of possible solutions, while a linear programming algorithm with non-convex constraints ensures a reasonable computing time. Process consistency is ensured by integrating all tracking decisions into a single optimization problem.

The authors carried out two experiments on synthetic and real data to test and compare their method with various state-of-the-art particle-tracking techniques. Both data sets consisted of low signal-to-noise ratio fluorescence images depicting cluttered biological samples. The proposed method proved remarkably robust to clutter, and was the only one to deliver high-quality results under poor imaging conditions. As a final note, the algorithm was implemented in JAVA and is available as an open-source plug-in for the Icy platform [Chaumont et al. 2012].

### 1.4.2 The U-TRACK algorithm

U-TRACK is a popular particle tracking algorithm developed with MATLAB [Jaqaman et al. 2008]. It was designed to address important tracking challenges, such as high densities of particles in the image sequence, the presence of different types of motion, particles merging/splitting, or gaps in trajectories resulting from particle disappearance or detection failure.

The algorithm operates in two steps: first, using combinatorial optimization and in particular the linear assignment problem, it links detected particles positions between consecutive frames; second, using a similar approach, it links the obtained track segments into complete trajectories, closing gaps and detecting merging/splitting events between initial track segments. The method also uses recursive tracking in forward and reverse time directions. Both steps use spatial optimization for particle and track segment assignation, which increases tracking accuracy in the case of high particle density, while the second step additionally uses time optimization. U-TRACK waits until the second step to construct all possible paths from the initial track segments obtained in the first step, unlike MHT which builds in all frames all possible paths from detected particle positions. This makes it possible to solve the assignment problem with a powerful desktop computer, as it considerably reduces the space of potential assignments.

U-TRACK is a robust algorithm, capable of predicting abrupt transitions between particle motion types and recovering from them, as well as being suitable for a wide range of tracking tasks in live-cell imaging. Its two-step optimization brings it close to the theoretically best, but computationally unimplementable, Multiple-Hypothesis Tracking approach. In [Jaqaman et al. 2008], it was successfully tested on two biological case studies involving GTPase dynamin and CD36 receptors, which depended crucially on the algorithm’s ability to produce accurate tracks in dense particle fields and to detect merge and split events, respectively. As a final note, a 3D-capable version of U-TRACK was recently developed and presented in [Roudot, Legant, et al. 2022].

## 1.5 Particle motion classification

The motion of a particle in the cytoplasm can be roughly classified into three classes. In general, particle motion is a stochastic process, resulting from all shocks between the particles and the environment. In what follows, we describe conventional situations. First, free diffusion (or Brownian motion) corresponds to pure diffusion (no obstacle and no collision), with no drift. The Langevin equation (2.1) can be re-written in a simplified form as

$$d\mathbf{x} = \sigma(\mathbf{x})d\mathbf{b}.$$

Second, the particle can be bound to a static point. In this case (subdiffusion) stochastic motion is restricted to a small area around the binding point. The Ornstein-Uhlenbeck process has been proposed to model this motion, and can be seen as a particular type of Langevin stochastic process, with (2.1) re-written as

$$d\mathbf{x} = \theta(\mathbf{x}_0 - \mathbf{x})dt + \sigma(\mathbf{x})d\mathbf{b},$$

where  $\theta$  designates a constant corresponding to the strength of the binding and  $\mathbf{x}_0$  the binding point.

Third, superdiffusion corresponds to a mostly deterministic motion. The particle moves in a given direction and motion is not restricted to a small area. This case typically happens when the particle is being actively transported within the cell by molecular motors and (2.1) is unchanged.

Several characteristics, including the length and direction of the particle trajectories, the different modes (superdiffusion, subdiffusion, free diffusion) of diffusion governed by

parameters (diffusion coefficient, drift), and the transitions between two distinct diffusion modes, are used to characterize specific intracellular mechanisms and processes. An important task is to assign a diffusion mode to each trajectory and to estimate the underlying parameters. The MSD (Mean Square Displacement) method is the most popular in this context, but is not very robust in distinguishing Brownian motion from confined motion. The classification errors increase as the trajectory length decreases and the trajectories are noisy (e.g., localization errors). Bayesian and learning-based approaches allow to obtain higher classification rates but require a large numbers of simulations [Monnier, Barry, et al. 2015] [Granik et al. 2019].

In [Briane, Kervrann, and Vimond 2018], the authors proposed a fast statistical method requiring no simulation, which given a trajectory as an input, is able to assign it an individual motion class. The main advantage of the Briane’s method is that it can efficiently handle short trajectories from 10 points and beyond. Unlike the optimization algorithm to find the Maximum A Posteriori in the Bayesian setting, no prior is required. The approach is low time consuming and requires no memory capacity. The same particle being able to change the way it moves following its interaction with its cellular environment, the same authors proposed in [Briane, Vimond, Valades-Cruz, et al. 2020] a statistical method to detect switches between two types of motion along the trajectory. This method is able to divide a trajectory into classified subtrajectories. In what follows, we use the two aforementioned methods to classify particle tracks. We end up with the set  $\mathcal{X} \subset \mathbb{R}^d, d \in \{2, 3\}$  of all points from all trajectories, divided by trajectories, subtrajectories, and motion class.

### 1.5.1 Particle trajectory classification

Let  $\mathbb{X}_n = (\mathbf{x}_{t_0}, \dots, \mathbf{x}_{t_n})$  be a trajectory generated from an unknown  $d$ -dimensional ( $d = 2$  or  $d = 3$ ) diffusion process  $(\mathbf{x}_t)$ . We define three hypotheses:

**Hypothesis 0 (Test hypothesis)**  $(\mathbf{x}_t)$  is a *Brownian motion*.

**Hypothesis 1**  $(\mathbf{x}_t)$  is a *subdiffusive process*.

**Hypothesis 2**  $(\mathbf{x}_t)$  is a *superdiffusive process*.

The procedure of [Briane, Kervrann, and Vimond 2018] allows one to assess from which type of diffusion the observed trajectory is generated with two statistical tests:

1. Hypothesis 0 versus Hypothesis 1
2. Hypothesis 0 versus Hypothesis 2

The two procedures are then aggregated to build a three-decision procedure. Finally, a multiple test procedure is proposed to reduce the rate of false alarms. This procedure allows one to control the number of false detections, i.e. the number of trajectories detected as non-Brownian when in fact they are.

Let  $D_n$  be the maximal distance of the process from its starting point and  $T_n$  the standardized distance:

$$T_n = \frac{D_n}{\sqrt{(t_n - t_0)\hat{\sigma}_n^2}},$$

$$D_n = \max_{i=A,\dots,n} \|\mathbf{x}_{t_i} - \mathbf{x}_{t_0}\| ,$$

where  $\hat{\sigma}_n$  is a estimator of  $\sigma$  measured on one particular trajectory. Briane, Kervrann and Vimond propose two estimators for  $\sigma$ , of first and second order respectively [Briane, Kervrann, and Vimond 2018]:

$$\hat{\sigma}_{1,n}^2 = \frac{1}{dn\Delta t} \sum_{j=1}^n \|\mathbf{x}_{t_j} - \mathbf{x}_{t_{j-1}}\|^2 ,$$

$$\hat{\sigma}_{2,n}^2 = \frac{1}{2dn\Delta t} \sum_{j=1}^n \left\| (\mathbf{x}_{t_{j+1}} - \mathbf{x}_{t_j}) - (\mathbf{x}_{t_j} - \mathbf{x}_{t_{j-1}}) \right\|^2 .$$

A low  $T_n$  means that the process remains close to its initial position for the period  $[t_0, t_n]$ : this is probably subdiffusion. On the contrary, if  $T_n$  is large, it means that the process is likely to move away from its starting point as a superdiffusive process. This measure  $T_n$  introduces an order into the diffusion processes solution of the SDE (1.9). It can then be used to classify them into the different diffusion classes: free diffusion, superdiffusion and subdiffusion. We want a test whose null hypothesis is that the trajectory comes from a Brownian motion, the gold standard process in biophysics. As a consequence,  $T_n$  must be a pivotal statistic under the Hypothesis 0.

Let  $q_n(\alpha)$  be the quantile of  $T_n$  of order  $\alpha \in (0, 1)$  when  $(\mathbf{x}_t)$  is a Brownian motion and  $R_{1,\alpha}$  and  $R_{2,\alpha}$  two critical regions defined as

$$R_{1,\alpha} = T_n < q_n(\alpha) ,$$

$$R_{2,\alpha} = T_n > q_n(1 - \alpha) .$$

If  $\mathbb{X}_n \in R_{1,\alpha}$ , we choose Hypothesis 1. If  $\mathbb{X}_n \in R_{2,\alpha}$ , we choose Hypothesis 2. Otherwise, we do not reject Hypothesis 0.

## 1.5.2 Detection of motion class switching along particle trajectories

In addition to classifying trajectories by motion type as described in the previous subsection, Briane, Vimond and Valades-Cruz introduce a method for detecting motion changes within trajectories, thus enabling us to divide them into sub-trajectories of different motion types [Briane, Vimond, Valades-Cruz, et al. 2020]. The procedure is as follows:

1. Detect potential change points ,
2. Group the potential points into clusters ,
3. Estimate a simple change point in each cluster .

We detail each step in the next subsections.

### Detecting the potential change points

Let  $\mathbb{X}_n = (\mathbf{x}_{t_0}, \dots, \mathbf{x}_{t_n})$  and  $1 \leq k \leq n/2$ . For all index  $i$  such that  $t_k \leq t_i \leq t_{n-k}$ , we consider two subtrajectories of size  $k$  starting at  $\mathbf{x}_{t_i}$ ,

- the forward trajectory  $\mathbb{X}_i^+ = \{\mathbf{x}_{t_i}, \mathbf{x}_{t_{i+1}}, \dots, \mathbf{x}_{t_{i+k}}\}$ ,
- the backward trajectory  $\mathbb{X}_i^- = \{\mathbf{x}_{t_i}, \mathbf{x}_{t_{i-1}}, \dots, \mathbf{x}_{t_{i-k}}\}$ .

As with motion classification, we calculate the standardized maximal distances from the "starting" point of the forward and reverse trajectories:

$$A_i = \frac{\max_{j=1,\dots,k} \|\mathbf{x}_{t_{i+j}} - \mathbf{x}_{t_i}\|}{\sqrt{(t_{i+k} - t_i) \hat{\sigma}(t_i : t_{i+k})}},$$

$$B_i = \frac{\max_{j=1,\dots,k} \|\mathbf{x}_{t_{i-j}} - \mathbf{x}_{t_i}\|}{\sqrt{(t_{i+k} - t_i) \hat{\sigma}(t_{i-k} : t_i)}},$$

where  $\hat{\sigma}(t_i : t_{i+k})$  (respectively  $\hat{\sigma}(t_{i-k} : t_i)$ ) denotes the estimate of the diffusion coefficient from the forward trajectory  $\mathbb{X}_i^+$  (respectively the backward trajectory  $\mathbb{X}_i^-$ ).

To detect switching, we want to compare  $A_i$  and  $B_i$ : the idea is that if both values are in the same range, it's unlikely that the moment  $t_i$  is a change point. To do this, we

define the following function:

$$\phi(x; \eta_1, \eta_2) = \begin{cases} 1 & \text{if } x < \eta_1, \\ 2 & \text{if } x > \eta_2, \\ 0 & \text{otherwise,} \end{cases}$$

where  $\eta_1 \in \mathbb{R}^+$  and  $\eta_2 \in \mathbb{R}^+$  are cut-off values depending on the parameters of the procedure, with  $\eta_1 < \eta_2$ .

As explained before,  $\phi(A_i; \eta_1, \eta_2) = 0$  means that  $\mathbb{X}_i^+$  is Brownian,  $\phi(A_i; \eta_1, \eta_2) = 1$  means that  $\mathbb{X}_i^+$  is subdiffusive and  $\phi(A_i; \eta_1, \eta_2) = 2$  superdiffusive. Further, we compute

$$Q_i = \phi(A_i; \eta_1, \eta_2) - \phi(B_i; \eta_1, \eta_2), \quad i = k, \dots, n - k.$$

Here,  $Q_i = 0$  means that  $A_i$  and  $B_i$  belong to the same range of values defined by  $\eta_1$  and  $\eta_2$ . Then, both  $\mathbb{X}_i^+$  and  $\mathbb{X}_i^-$  are from the same type of diffusion and  $t_i$  is unlikely a change point. On the contrary, if  $Q_i \neq 0$  the subtrajectories  $\mathbb{X}_i^+$  and  $\mathbb{X}_i^-$  are not from the same type of diffusion and  $t_i$  is a potential change point.

### Grouping the potential change points into clusters

We assume that a subset of indexes where the concentration of potential change point is high (even if there are not connected) is likely to contain a true change point. We thus define a cluster of potential change points as a subset of index  $\mathcal{M} = i, \dots, i + l$  such that

$$\sum_{j=m}^{m+c-1} \mathbf{1}(Q_j \neq 0) \geq c^*, \quad \forall m = i, \dots, i + l - c + 1,$$

where  $c = k/2$  and  $c^* = 3k/8$  are tuning parameters, so that the cluster has a minimal size of  $k/2$  as in [H. Cao and Wu 2015], and a cluster is created if there is a concentration of at least 75% potential change points in a set of  $c$  successive points.

### Estimating a unique change point in each cluster

Let  $\mathcal{M}_j$  be the  $j^{\text{th}}$  cluster,  $\tau_j$  the true change point in  $\mathcal{M}_j$ , and  $r_j$  the index of  $\tau_j$  such that  $\tau_j = t_{r_j}$ . At  $\tau_j$ ,  $\mathbb{X}_{r_j}^-$  and  $\mathbb{X}_{r_j}^+$  are trajectories generated by different diffusion processes, such that  $B_{r_j}$  and  $A_{r_j}$  must be the most different. At a point  $t_i$  close to  $\tau_j$ , the subtrajectories  $\mathbb{X}_i^-$  and  $\mathbb{X}_i^+$  are composed of a diffusion mixture, which implies that

$|B_i - A_i| \leq |B_{r_j} - A_{r_j}|$ . Then, we estimate the most representative change point in the  $j^{\text{th}}$  cluster as

$$\hat{\tau}_j = t_{r_j}, \quad r_j = \max_{i \in \mathcal{M}_j} |B_i - A_i|.$$

### 1.5.3 Merging of motion classification and switching detection

In the end, it is possible to divide and classify a set of trajectories by combining the two aforementioned procedures. For the purposes of our work, we have merged motion classification and switch detection into a single software package (MATLAB). It works as follows: let  $\mathbb{X}_n$  be a trajectory of length  $n$ . If  $n > 25$ , we use the switching detection procedure to divide  $\mathbb{X}$  into several subtrajectories separated by their change points, and classify them separately. If  $10 < n < 25$ , then  $\mathbb{X}_n$  is too short to search for change points, and we classify it as a whole using the first procedure. For conservative reasons, if  $n < 10$ ,  $\mathbb{X}_n$  is classified as Brownian by default.

## 1.6 Simulating fluorescence microscopy particle dynamics with FLUOSIM simulator

In this section, we present FLUOSIM, a fluorescence microscopy simulation software developed by Matthieu Lagardère and Olivier Thoumine at IINS, University of Bordeaux [Lagardère et al. 2020]. It is a fast, robust and user-friendly computer program capable of simulating membrane protein dynamics for various live-cell and super-resolution imaging techniques, such as Single-Particle Tracking (SPT), PhotoActivation of Fluorescence (PAF), Fluorescence Recovery After Photobleaching (FRAP), Fluorescence Correlation Spectroscopy (FCS), PhotoActivated Localization Microscopy (PALM), direct Stochastic Optical Reconstruction Microscopy (dSTORM) and Point Accumulation In Nanoscopic Topography (PAINT). It allows the user to set diffusion coefficients, binding rates and fluorophore photophysics before simulating in real time the positions and intensities of thousands of independent molecules within 2D cell shapes, generating simulated image sequences that can be compared to real experiments. The program can model a wide range of dynamic 2D molecular systems with membrane diffusion and transient confinement. It has been used, for example, to characterize neurotransmitter receptors at synapses, or the trapping of membrane molecules by lipid rafts. In addition, FLUOSIM can be used to evaluate the performance of image processing algorithms, such as bioparticle detectors

and trackers, or motion classification methods. Although currently limited to Brownian motion simulation, it can potentially be used to simulate large datasets for deep learning algorithms [Lagardère et al. 2020].

Matthieu Lagardère provided us with the software, which we used as an experimentation tool in our work relating to Chapters 2 and 3 of this thesis. In what follows, we give a general presentation of the software principle and interface, before describing its mathematical framework in detail, and then reviewing its capabilities and performance.

### 1.6.1 General principle

After importing a 2D cell shape extracted from a microscopy image, the user sets various parameters such as the number of molecules, acquisition time and frame rate, local diffusion coefficients, trapping areas (un)binding rates and fluorescence switching rates (see Figure 1.19). Once the simulation is started, the software algorithm updates the positions and intensities of molecules at each time step, using random number generation to mimic stochastic motion and random photo-switching of fluorescent tags. The algorithm is capable of displaying the cell system in real time and providing post-processing information such as particle trajectories, image sequences and various graphs. The image sequences can be corrupted by Poisson or Gaussian noise, enabling the simulation of realistic image sequences. As FLUOSIM calculates particle positions with almost infinite precision, it is capable of generating super-resolution images similar to PALM or dSTORM, using a few additional parameters.

### 1.6.2 Mathematical framework

One particular feature of FLUOSIM that we use extensively in our research is the ability to create 2D subregions in which particles can be trapped (microdomains). In what follows, we define the underlying mathematical framework and notations of a FLUOSIM simulation, which we will use in Chapters 2 and 3.

First, we denote as  $\mathcal{D}$  the simulation domain corresponding to the the aforementioned 2D cell shape, whose boundaries normally reflect particles. Within  $\mathcal{D}$  can be defined a finite number of subregions with specific properties called trapping areas, that we denote as  $\{\mathcal{S}_1, \dots, \mathcal{S}_j, \dots, \mathcal{S}_k\} \subset \mathcal{D}$ . We also define  $\mathcal{S} = \bigcup_1^k \mathcal{S}_j$  and  $\bar{\mathcal{S}} = \mathcal{D}/\mathcal{S}$ .

Let  $\{p_1, \dots, p_l, \dots, p_N\} \in \mathcal{D}$  be  $N \in \mathbb{N}^*$  simulated particles. We denote the trajectory

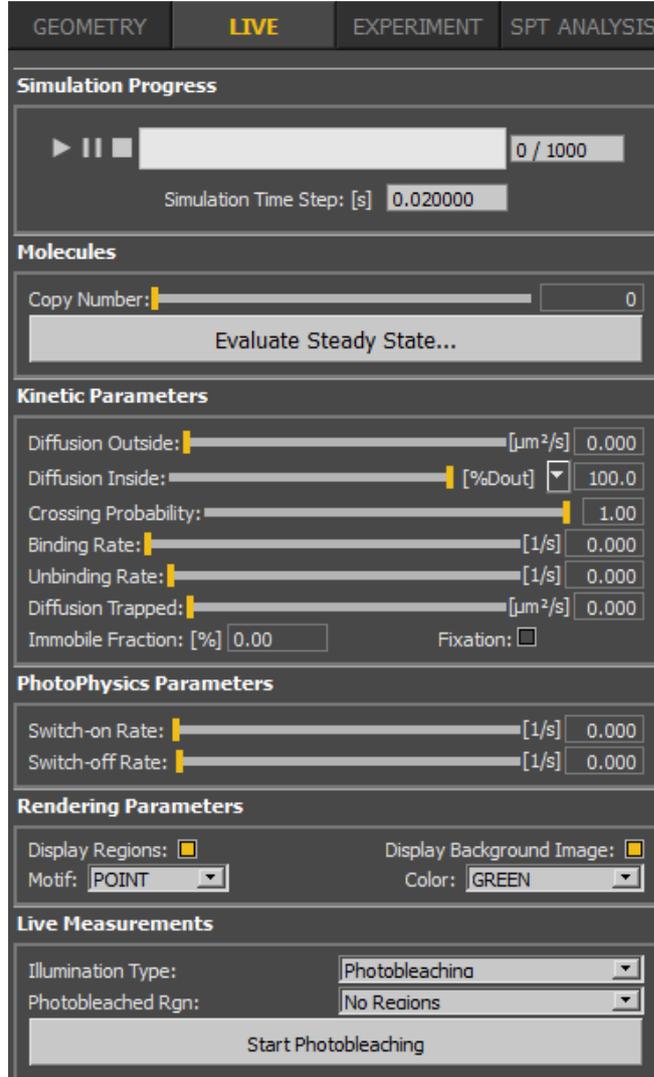


Figure 1.19 – The fluosim interface and its parameters.

of each particle  $p_l$  over time as

$$\mathbb{X}_l = \{ \mathbf{x}_{t_0}^{(l)}, \dots, \mathbf{x}_{t_i}^{(l)}, \dots, \mathbf{x}_{t_n}^{(l)} \},$$

where  $\mathbf{x}_{t_i}^{(l)} \in \mathbb{R}^2$  is the position of particle  $p_l$  at time  $t_i \in \mathbb{R}_+$ , and  $n \in \mathbb{N}^*$  is the length of the trajectory corresponding to the number of simulated frames. The time difference between two consecutive particle positions is a constant  $\Delta t$  corresponding to the simulation frame rate, with  $t_i = t_0 + i\Delta t$ .

If  $p_l \in \mathcal{S}$ , the particle may alternatively switch between two states: free or trapped.

If  $p_l$  is free, it can escape its trapping area  $\mathcal{S}_j$ . If it is trapped, the boundary of  $\mathcal{S}_j$  will normally reflect  $p_l$ . We define the trapping indicator as

$$T(\mathbf{x}_{t_i}^{(l)}) = \begin{cases} 1 & \text{if } p_l \text{ is trapped at time } t_i, \\ 0 & \text{if } p_l \text{ is free at time } t_i. \end{cases}$$

Each particle  $p_l \in \mathcal{D}$  undergoes Brownian motion with three possible diffusion coefficients  $\sigma_{\bar{\mathcal{S}}}$ ,  $\sigma_{\mathcal{S}}$  and  $\sigma_t \leq \sigma_{\bar{\mathcal{S}}}$  set by the user as follows:

- If  $p_l \in \bar{\mathcal{S}}$ ,  $\sigma(\mathbf{x}_{t_i}^{(l)}) = \sigma_{\bar{\mathcal{S}}}$ ,
- If  $p_l \in \mathcal{S}$  and  $p_l$  is free,  $\sigma(\mathbf{x}_{t_i}^{(l)}) = \sigma_{\mathcal{S}}$ ,
- If  $p_l \in \mathcal{S}$  and  $p_l$  is trapped,  $\sigma(\mathbf{x}_{t_i}^{(l)}) = \sigma_t$ .

We denote the binding and unbinding rates of the simulation respectively as  $\{k^+, k^-\} \in [0, 1/\Delta t]^2$  such that

$$\begin{aligned} P(T(\mathbf{x}_{t_{i+1}}) = 1 | T(\mathbf{x}_{t_i}) = 0, \mathbf{x}_{t_{i+1}} \in \mathcal{S}) &= k^+ \Delta t, \\ P(T(\mathbf{x}_{t_{i+1}}) = 0 | T(\mathbf{x}_{t_i}) = 1, \mathbf{x}_{t_{i+1}} \in \mathcal{S}) &= k^- \Delta t. \end{aligned}$$

Particles can also switch between on and off states, meaning they emit light or not, respectively. We define the switching indicator as

$$S(\mathbf{x}_{t_i}^{(l)}) = \begin{cases} 1 & \text{if } p_l \text{ is switched-on at time } t_i, \\ 0 & \text{if } p_l \text{ is switched-off at time } t_i. \end{cases}$$

We denote the switch-on and switch-off rates of the simulation as  $\{s^+, s^-\} \in [0, 1/\Delta t]^2$  such that

$$\begin{aligned} P(T(\mathbf{x}_{t_{i+1}}) = 1 | T(\mathbf{x}_{t_i}) = 0) &= s^+ \Delta t, \\ P(T(\mathbf{x}_{t_{i+1}}) = 0 | T(\mathbf{x}_{t_i}) = 1) &= s^- \Delta t. \end{aligned}$$

Note that  $k^+$ ,  $k^-$ ,  $s^+$  and  $s^-$  are space-independent constants that do not vary within trapping areas.

### 1.6.3 Performances

As illustrated in Figure 1.20, the performance of FLUOSIM was carefully tested in [Lagardère et al. 2020]. The simulated data it generated was compared with experimental

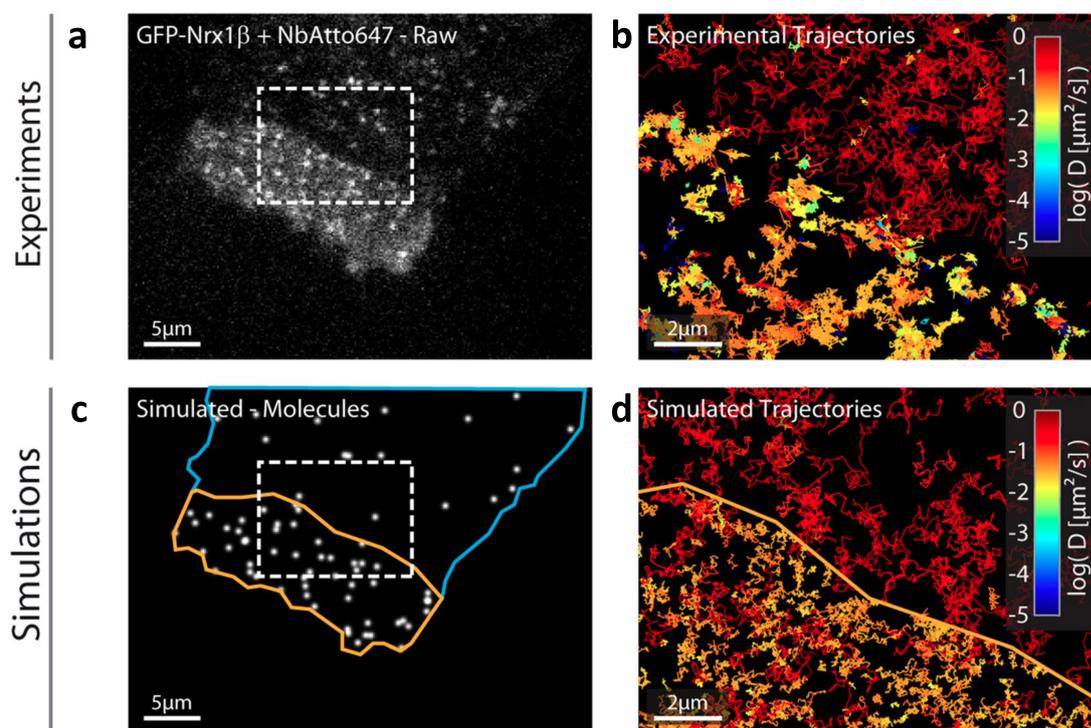


Figure 1.20 – Comparison between SPT experiments and their corresponding FLUOSIM simulations. **(a)** Image sequence showing a COS-7 cell expressing GFP-Nrx1 $\beta$  sparsely labeled with Atto647N-conjugated GFP nanobody. **(b)** Zoom on the ROI in **(a)** (white rectangle) showing the trajectories tracked after image acquisition. **(c)** FLUOSIM simulation of particles in a geometry similar to the experiment illustrated in **(a)**. **(d)** Zoom on the ROI in **(c)** representing trajectories simulated with FLUOSIM. (extracted and edited from [Lagardère et al. 2020])

data, in particular SPT, dSTORM, FRAP, and FCS experiments that were performed on the canonical neurexin-neurologin adhesion complex that mediates trans-synaptic adhesion in neurons. It successfully reproduced the experimental results using parameters extracted from published in vitro studies and/or measurements carried out by the authors.

FLUOSIM is a fast, robust program capable of simulating membrane protein dynamics in a wide range of fluorescence imaging modalities, with real-time display and precise control over kinetics, photo-physics and image acquisition parameters. For the time being, FLUOSIM can only simulate Brownian motion and simple molecular reactions, but the authors point out that subdiffusive or superdiffusive particle motion as well as complex reactions can be implemented, depending on user requirements on a case-by-case basis.

## 1.7 Conclusion

As our thesis work lies at the intersection of mathematics, optical microscopy and computer science, we introduced a whole set of key tools from these various disciplines in order to fully grasp our framework.

First, we introduced the statistical theory of stochastic processes, Brownian motion and diffusion. Second, we presented from a theoretical point of view several models and types of particle motion, such as free diffusion, subdiffusion and superdiffusion. Third, as the image sequences we worked on came from fluorescence microscopy imaging, we introduced various conventional, super-resolved, diffraction-unlimited super-resolved microscopy techniques such as Wide-Field, Confocal, TIRF, (Lattice) Light Sheet, Structured Illumination, Stimulated Emission Depletion and Single Molecule Localization. Fourth, we gave a brief summary of two popular particle trackers (Multiple Hypothesis Tracking [Chenouard, Bloch, and Olivo-Marin 2013], U-TRACK [Jaqaman et al. 2008]) that we used to extract localization information from biological image sequences. Finally, we presented two particle motion classification and segmentation methods [Briane, Kervrann, and Vimond 2018] [Briane, Vimond, Valades-Cruz, et al. 2020].

These methods and tools have been combined and integrated into the diffusion and drift estimation scheme that we present in the next chapter, and which constitutes the core of our work. In what follows, we propose a new algorithmic method that estimates diffusion and drift maps using particle tracking data. After presenting some related work and methods, we introduce our method from a theoretical and algorithmic point of view, before testing it on a extensive set of experiments.



# DENSE MAPPING OF INTRACELLULAR DIFFUSION AND DRIFT FROM MOLECULE TRACKS

---

The study of protein dynamics within cells is an important field of study in cell biology and biophysics, which is inherently linked to fluorescence microscopy. In general, fluorophores (fluorescent compounds) can be attached to proteins in order to observe and study intracellular dynamics. Consequently, there is a high demand for computational tools capable of automatically analyzing the microscopic data, and in particular the image sequences depicting the moving particles in the form of bright spots produced by fluorophores. In this chapter, we propose a method based on a spatio-temporal Gaussian kernel that estimates intracellular diffusion and drift maps among other outputs, using single-particle tracking data obtained through fluorescent microscopy image sequences.

First, we begin by reviewing the current state-of-the-art techniques involved in the estimation of intracellular diffusion and drift. In particular, we focus on the works of Hozé and Holcman [Hozé, Nair, et al. 2012] [Hozé and Holcman 2014] [Hozé and Holcman 2017] who proposed an Eulerian approach to solve this problem, and the works of El Beheiry and Masson [Beheiry, Dahan, and Masson 2015] [Masson, Casanova, et al. 2009], who proposed a Bayesian inference method to address the same problem. Second, we present the statistical and mathematical theory of our method in detail, especially the several spatiotemporal kernel estimators that are used to analyze the input SPT data, depending on the particle motion type. Third, we assess the quality and robustness of our method by reporting an extensive set of experiments, using both simulated, real, 2D and 3D data sets. We compared our method with the two aforementioned methods, before drawing a conclusion on its accuracy, robustness and advantages.

## 2.1 Related work

First, Hozé and Holcman consider a Eulerian approach to estimate diffusion and drift from tracklets in a series of papers [Hozé, Nair, et al. 2012] [Hozé and Holcman 2014] [Hozé and Holcman 2017]. Meanwhile, El Beheiry and Masson considered the Bayesian inference framework to achieve the same goal [Beheiry, Dahan, and Masson 2015] [Masson, Casanova, et al. 2009].

### 2.1.1 Eulerian method for diffusion estimation

In this subsection, we introduce the theory behind the Hozé and Holcman Eulerian approach. As a starting point, Langevin proposed a description of a stochastic particle trajectory following this equation [Langevin 1908]:

$$m\ddot{\mathbf{x}}(t) = -\gamma\dot{\mathbf{x}}(t) + \mathbf{f}_s(\mathbf{x}, t), \quad (2.1)$$

where  $m \in \mathbb{R}_+^*$  is the particle's mass,  $\mathbf{x} \in \mathbb{R}^d, d \in \{2, 3\}$  denotes the particle's spatial coordinates,  $\gamma \in \mathbb{R}_+^*$  is the particle's friction coefficient,  $\mathbf{f}_s : \mathbb{R}^d \times \mathbb{R}_+ \rightarrow \mathbb{R}^d$  is a stochastic, fluctuational force resulting from the particle's random collisions with the surrounding fluid molecules, and  $t \in \mathbb{R}_+$  designates time. If a deterministic force field  $\mathbf{f}_d : \mathbb{R}^d \times \mathbb{R}_+ \rightarrow \mathbb{R}^d$  is also applied to the particle, the equation (2.1) becomes

$$m\ddot{\mathbf{x}}(t) = -\gamma\dot{\mathbf{x}}(t) + \mathbf{f}_d(\mathbf{x}, t) + \mathbf{f}_s(\mathbf{x}, t).$$

The Smoluchowski's limit [Schuss 2010] [Holcman and Schuss 2017] of (2.1) shows that as  $\gamma \rightarrow \infty$ , the equation reduces to

$$\gamma\dot{\mathbf{x}}(t) = \mathbf{f}_d(\mathbf{x}, t) + \mathbf{f}_s(\mathbf{x}, t). \quad (2.2)$$

We assume that  $\mathbf{f}_s$  is a Gaussian process:

$$\mathbf{f}_s(\mathbf{x}, t) = f_s(\mathbf{x}, t)\dot{\mathbf{b}}(t),$$

where  $f_s \in \mathbb{R}_+$  and  $\mathbf{b}$  is standard Gaussian white noise. We can then rewrite (2.2) in the following form:

$$\dot{\mathbf{x}}(t) = \frac{\mathbf{f}_d(\mathbf{x}, t)}{\gamma} + \frac{f_s(\mathbf{x}, t)}{\gamma}\dot{\mathbf{b}}(t). \quad (2.3)$$

The source of the driving noise is the thermal agitation of the ambient lipid and membrane molecules. However, due to the acquisition timescale of empirically recorded trajectories, which is too low to follow the thermal fluctuations, rapid events are not resolved in the data, and at this coarser spatiotemporal scale, the motion is described by the effective stochastic equation [Hozé, Nair, et al. 2012] [Hozé and Holcman 2014]:

$$d\mathbf{x} = \boldsymbol{\mu}(\mathbf{x}, t)dt + \sigma(\mathbf{x}, t)d\mathbf{b},$$

where  $\boldsymbol{\mu} \in \mathbb{R}^d$  is the drift vector and  $\sigma \in \mathbb{R}^+$  is the diffusion coefficient.

This equation can be used to recover the drift and the diffusion from Single-Particle tracks acquired at any infinitesimal time step  $\Delta t$  [Schuss 2010] [Hozé and Holcman 2014] [Hozé and Holcman 2017] [Friedrich and Peinke 1997] [Siegert, Friedrich, and Peinke 1998]:

$$\boldsymbol{\mu}(\mathbf{x}, t) = \lim_{\Delta t \rightarrow 0} \frac{\mathbb{E}[\mathbf{x}(t + \Delta t) - \mathbf{x}(t) | \mathbf{x}(t) = \mathbf{x}]}{\Delta t}, \quad (2.4)$$

$$\sigma(\mathbf{x}, t) = \lim_{\Delta t \rightarrow 0} \frac{\mathbb{E}[|\mathbf{x}(t + \Delta t) - \mathbf{x}(t)|^2 | \mathbf{x}(t) = \mathbf{x}]}{\Delta t}, \quad (2.5)$$

where  $\mathbb{E}[\cdot | \mathbf{x}(t) = \mathbf{x}]$  denotes the averaging over all trajectories at point  $\mathbf{x}$  and time  $t$ . To estimate the local drift  $\boldsymbol{\mu}(\mathbf{x})$  and diffusion coefficients  $\sigma(\mathbf{x})$  at each point  $\mathbf{x}$ , Hozé et al. use a computational procedure summarized below [Hozé, Nair, et al. 2012].

In [Hozé, Nair, et al. 2012], the points of trajectories are first grouped within small square bins  $S(g_k, r)$  of size  $\Delta g$  and centered on lattice grid  $g_k$ . The local drift vector and diffusion coefficient are independently estimated in each square block. Let  $\{X_i(0), \dots, X_i(K\Delta t)\}$  be a set of  $N$  trajectories with  $i = 1, \dots, N$  of respective lengths  $l_1, \dots, l_N$  and  $\Delta t$  the sampling time. The discretization of equations (2.4) and (2.5) for the drift  $\boldsymbol{\mu}(\mathbf{x}_k)$  and diffusion  $\sigma(\mathbf{x}_k)$  in a bin centered at position  $\mathbf{x}_k$  are respectively

$$\boldsymbol{\mu}(\mathbf{x}_k) \approx \frac{1}{\Delta t \cdot n_k} \sum_{i=1}^N \sum_{j=0}^{l_i-1} \mathbf{1}_{[S(g_k, r)]}(\mathbf{x}_i(j\Delta t)) \cdot (\mathbf{x}_i((j+1)\Delta t) - \mathbf{x}_i(j\Delta t)),$$

$$\sigma(\mathbf{x}_k) \approx \frac{1}{2\Delta t \cdot n_k} \sum_{i=1}^N \sum_{j=0}^{l_i-1} \mathbf{1}_{[S(g_k, r)]}(\mathbf{x}_i(j\Delta t)) \cdot (\mathbf{x}_i((j+1)\Delta t) - \mathbf{x}_i(j\Delta t))^2,$$

where  $\mathbf{1}_{[S(g_k, r)]}$  is the indicator function of  $S(g_k, r)$  and  $n_k = \sum_{i=1}^N \sum_{j=0}^{l_i-1} \mathbf{1}_{[S(g_k, r)]}(\mathbf{x}_i(j\Delta t))$  is the number of trajectory points falling in the square block  $S(g_k, r)$ . More recently, Parutto et al. proposed to compute cosine-filtered maps to improve visualization. The role of cosine

weights is to decrease continuously the influence of the points but the spatial resolution is still crude when compared to the spatial resolution of input trajectory sets [Parutto et al. 2022].

This block-based approach has one severe limitation: the trade-out between resolution and input data quantity is often sub-optimal and can lead to poor quality in terms of results if the input data set is not dense enough. The method was actually designed to process very large SPT-PALM data sets. We propose an alternative approach based on the same framework to overcome these shortcomings.

### 2.1.2 Bayesian method for diffusion estimation

El Beheiry and Masson consider the Bayesian inference framework to address the same issues. INFERENCEMAP [Beheiry, Dahan, and Masson 2015] starts by building a mesh of the image domain from trajectory points. This can be achieved either by a basic Square Meshing, as in [Hozé, Nair, et al. 2012], a Voronoi tessellation [Voronoi 1908a] [Voronoi 1908b], or a Quad-Tree Meshing [Finkel and Bentley 1974]. The software then uses a Bayesian inference technique described in [Masson, Casanova, et al. 2009] to calculate the diffusion or the drift inside each meshing zone as follows.

The authors consider the overdamped Langevin equation (2.3) as a particle motion model, as well as the associated Fokker-Planck equation, describing the time-evolution of the particle transition probability  $P(\mathbf{x}_2, t_2 | \mathbf{x}_1, t_1)$ :

$$\frac{dP(\mathbf{x}_2, t_2 | \mathbf{x}_1, t_1)}{dt} = -\nabla \cdot \left( \frac{\mathbf{f}(\mathbf{x}, t)}{\gamma(\mathbf{x})} P(\mathbf{x}_2, t_2 | \mathbf{x}_1, t_1) - \nabla (\sigma(\mathbf{x}) P(\mathbf{x}_2, t_2 | \mathbf{x}_1, t_1)) \right). \quad (2.6)$$

This equation has no general analytic solution for arbitrary diffusion coefficient and force fields  $\sigma(\mathbf{x})$  and  $\mathbf{f}(\mathbf{x})$ . However, in a local area or region, it can be assumed that both  $\sigma$  and  $\mathbf{f}$  are constant, and the solution to (2.6) is of the following form:

$$P((\mathbf{x}_2, t_2 | \mathbf{x}_1, t_1) | \sigma_i, \mathbf{f}_i) = \frac{1}{4\pi \left( \sigma_i + \frac{\lambda^2}{t_2 - t_1} \right) (t_2 - t_1)} \exp \left( -\frac{\left( \mathbf{x}_2 - \mathbf{x}_1 + \frac{\mathbf{f}_i(t_2 - t_1)}{\gamma_i} \right)^2}{4 \left( \sigma_i + \frac{\lambda^2}{t_2 - t_1} \right) (t_2 - t_1)} \right),$$

where  $i$  denotes the mesh zone index and  $\lambda$  the experimental localization precision.

Furthermore, the overall probability of a trajectory  $T$  due to the spatially dependent variables  $\sigma$  and  $\mathbf{f}$  is obtained by multiplying the probabilities of all the individual

subdomains  $P(T|\sigma_i, \mathbf{f}_i)$ :

$$P(T|\sigma, \mathbf{f}) = \prod_i P(T|\sigma_i, \mathbf{f}_i). \quad (2.7)$$

According to Bayes' rule,

$$P(\sigma|\mathbf{f}, T) = \frac{P(T|\sigma, \mathbf{f})P(\sigma, \mathbf{f})}{P(T)},$$

where  $P(\sigma, \mathbf{f}|T)$  is the posterior probability,  $P(\sigma, \mathbf{f})$  is the prior probability, and  $P(T)$  is the evidence (which is treated as a normalization constant).

For each mesh zone, an optimization of the posterior probability  $P(\sigma, \mathbf{f}|T)$  is performed for the model parameters  $\sigma, \mathbf{f}$ . This is the core of the maximum a posteriori (MAP) Bayesian inference approach that gives its name to INFERENCEMAP. In particular, the estimation of  $\sigma$  is performed in each zone independently from the others, resulting in a rapid calculation as this mode consists of a single-variable optimization. The posterior probability used to infer the diffusion in a given zone is given by

$$P(\{\sigma_i\} | \{T_k\}) \propto \prod_k \prod_{\mu: \mathbf{x}_\mu^k \in S_i} \frac{P_J(\sigma_i) P_S(\sigma_i)}{4\pi \left(\sigma_i + \frac{\lambda^2}{\Delta t}\right) \Delta t} \exp\left(-\frac{(\mathbf{x}_{\mu+1}^k - \mathbf{x}_\mu^k)^2}{4 \left(\sigma_i + \frac{\lambda^2}{\Delta t}\right) \Delta t}\right),$$

where  $\sigma_i$  is the diffusion coefficient,  $\mu$  denotes the index for which the points  $\mathbf{x}_\mu^k$  of the  $k^{\text{th}}$  trajectory are in  $S_i$  (the current zone being analyzed), and  $\lambda$  is the experimental localization precision (30 nm by default). Here,  $P_J(\sigma_i)$  denotes the optimal Jeffrey's prior which may be activated and deactivated, and  $P_S(\sigma_i)$  is a diffusion smoothing prior.

Finally, the calculated values of  $\sigma$  are updated into the mesh and a mosaic-like map of diffusion is obtained (see example in Figure 2.5). It is worth noting that INFERENCEMAP also allows to calculate the drift within each mesh zone.

### 2.1.3 Conclusion

The main advantage of these two methods lie in their relative simplicity and fast processing time. Both methods however do not take into account particle motion types and act as if all particles undergo the same motion in a region of interest. In addition, they both necessitate a high and uniform density of particle positions to generate accurate results, performing poorly in low density regions as shown in Section 2.3. Finally, while the Hozé approach can easily be extended to work on 3D data sets, the El Beheiry algorithm

only works in 2D.

We propose to overcome several disadvantages of these two separate approaches by using motion classification and a particle position based approach in our method. To do so, we use a spatio-temporal kernel that calculates drift and diffusion values in each particle position of the data set, while adapting to the local data density. Additionally, classification allows one to use the most adapted formulas following particle motion type. Results can be displayed as point clouds and vector fields both in 2D and 3D.

## 2.2 Dense Mapping: a hybrid Lagrangian-Eulerian method for diffusion and drift estimation

### 2.2.1 Notations and definitions

#### Trajectories

Let  $d \in \{2, 3\}$ . In a  $d$ -dimensional space, we define a particle trajectory  $\mathbb{X}$  of length  $n \in \mathbb{N}^*$  as

$$\mathbb{X} = \{\mathbf{x}_{t_0}, \dots, \mathbf{x}_{t_i}, \dots, \mathbf{x}_{t_n}\},$$

where  $\mathbf{x}_{t_i} \in \mathbb{R}^d$  is the particle position at time  $t_i \in \mathbb{R}_+$ . We assume that the time difference between two consecutive particle positions is a constant  $\Delta t$  that we will call time resolution from now on, with  $t_i = t_0 + i\Delta t$ . For any particle position  $\mathbf{x} \in \mathbb{X}$ , we also denote  $t_{\mathbf{x}}$  the time associated to a given position.

Let  $\mathbb{X}$  be a particle trajectory. We define the trimmed trajectories  $\mathbb{X}'$  and  $\mathbb{X}''$  as

$$\begin{aligned}\mathbb{X}' &= \mathbb{X} \setminus \{\mathbf{x}_{t_n}\}, \\ \mathbb{X}'' &= \mathbb{X} \setminus \{\mathbf{x}_{t_0}, \mathbf{x}_{t_n}\}.\end{aligned}$$

#### Sets of particle positions

Let us consider a set of  $N \in \mathbb{N}^*$  trajectories  $\{\mathbb{X}_1, \dots, \mathbb{X}_N\}$ . We define the set  $\mathcal{X}$  of all particle positions from the  $N$  trajectories as

$$\mathcal{X} = \left\{ \mathbf{x} \in \mathbb{R}^d \mid \mathbf{x} \in \bigcup_{i=1}^N \mathbb{X}_i \right\}.$$

Based on the definition of  $\mathbb{X}'$  and  $\mathbb{X}''$ , we define  $\mathcal{X}'$  and  $\mathcal{X}''$  as:

$$\mathcal{X}' = \left\{ \mathbf{x} \in \mathbb{R}^d \mid \mathbf{x} \in \bigcup_{i=1}^N \mathbb{X}'_i \right\},$$

$$\mathcal{X}'' = \left\{ \mathbf{x} \in \mathbb{R}^d \mid \mathbf{x} \in \bigcup_{i=1}^N \mathbb{X}''_i \right\}.$$

### Motion type set notation

Let  $X$  be a generic set of particle positions, we define  $X_{free}$ ,  $X_{sub}$  and  $X_{sup}$  as

$$X_{free} = \{ \mathbf{x} \in X \mid \mathbf{x} \text{ is classified as Brownian} \},$$

$$X_{sub} = \{ \mathbf{x} \in X \mid \mathbf{x} \text{ is classified as subdiffusive} \},$$

$$X_{sup} = \{ \mathbf{x} \in X \mid \mathbf{x} \text{ is classified as superdiffusive} \}.$$

In what follows, this notation will typically concern the set  $\mathcal{X}$  with  $\mathcal{X}_{free}$ ,  $\mathcal{X}_{sub}$  and  $\mathcal{X}_{sup}$ .

### First and second order differences

Let  $\mathbb{X}$  be a trajectory of length  $n$ . For all  $\mathbf{x}_{t_i} \in \mathbb{X}'$ , we define

$$\Delta^{(1)} \mathbf{x}_{t_i} = \mathbf{x}_{t_{i+1}} - \mathbf{x}_{t_i} \quad (2.8)$$

as the first order difference of  $\mathbf{x}_{t_i}$ . Similarly, we define for all  $\mathbf{x}_{t_i} \in \mathbb{X}''$  the second order difference of  $\mathbf{x}_{t_i}$  as

$$\Delta^{(2)} \mathbf{x}_{t_i} = \mathbf{x}_{t_{i+1}} - 2\mathbf{x}_{t_i} + \mathbf{x}_{t_{i-1}}. \quad (2.9)$$

We denote  $\Delta^{(1)}\mathbb{X}$  and  $\Delta^{(2)}\mathbb{X}$  the sets of first and second order differences associated to the trajectory  $\mathbb{X}$ , respectively:

$$\Delta^{(1)}\mathbb{X} = \left\{ \Delta^{(1)} \mathbf{x}_{t_0}, \dots, \Delta^{(1)} \mathbf{x}_{t_{n-1}} \right\},$$

$$\Delta^{(2)}\mathbb{X} = \left\{ \Delta^{(2)} \mathbf{x}_{t_1}, \dots, \Delta^{(2)} \mathbf{x}_{t_{n-1}} \right\}.$$

Let us consider a set of  $N \in \mathbb{N}^*$  trajectories  $\{\mathbb{X}_1, \dots, \mathbb{X}_N\}$  of respective lengths  $\{n_1, \dots, n_N\}$  and  $\mathbf{x} \in \mathcal{X}'$ . By definition,

$$\exists!(i, j) \in \{1, \dots, N\} \times \{0, \dots, n_i - 1\}, \quad \mathbf{x} = \mathbf{x}_{t_j} \in \mathbb{X}'_i.$$

We define the first order difference of  $\mathbf{x}$  as the first order difference of its corresponding  $\mathbf{x}_{t_j}$ :

$$\Delta^{(1)}\mathbf{x} = \Delta^{(1)}\mathbf{x}_{t_j} \in \Delta^{(1)}\mathbb{X}_i.$$

Similarly, let  $\mathbf{x} \in \mathcal{X}''$ . By definition again,

$$\exists!(i, j) \in \{1, \dots, N\} \times \{1, \dots, n_i - 1\}, \quad \mathbf{x} = \mathbf{x}_{t_j} \in \mathbb{X}_i''.$$

The second order difference of  $\mathbf{x}$  is similarly defined as the second order difference of its corresponding  $\mathbf{x}_{t_j}$ :

$$\Delta^{(2)}\mathbf{x} = \Delta^{(2)}\mathbf{x}_{t_j} \in \Delta^{(2)}\mathbb{X}_i.$$

Given a set of trajectories  $\{\mathbb{X}_1, \dots, \mathbb{X}_N\}$ , our objective will be to estimate a drift vector and a diffusion coefficient on each particle position  $\mathbf{x} \in \mathcal{X}$ , using different formulas following the motion type of said particle (free diffusion, subdiffusion, or superdiffusion).

## 2.2.2 Dense diffusion and drift estimation

We now present the details of our estimation method that uses SPT data as an input. Nevertheless, is also capable to process image sequences directly, by subjecting them to a tracking process. The data consists of a set of  $N \in \mathbb{N}^*$  trajectories  $\{\mathbb{X}_1, \dots, \mathbb{X}_N\}$ . The principle relies on a spatio-temporal kernel to estimate diffusion and drift values on each point  $\mathbf{x} \in \mathcal{X}$ , that we call point of calculation. The kernel accounts for trajectories in the neighborhood of  $\mathbf{x}$ , and uses their own particle positions in a weighted fashion to estimate diffusion and drift at position  $\mathbf{x}$ . Depending on the motion type associated to  $\mathbf{x}$ , distinct calculation formulas are used.

### Definition of weights

Below, we detail the weighting process. The weighting function we use consists of a product between a spatial Gaussian and a temporal Gaussian centered around point of calculation  $\mathbf{x}$ . We call  $r_S$  and  $r_T$  the radiuses of the spatial Gaussian and the temporal Gaussian, respectively. The radius  $r_S$  is defined as the  $m^{th}$  minimum of the set of all distances between  $\mathbf{x}$  and other particle positions of  $\mathcal{X}$ :

$$r_S(\mathbf{x}, m) = \min^{(m)}(\{\|\mathbf{x}_k - \mathbf{x}\|, \mathbf{x}_k \in \mathcal{X}\}).$$

In practice,  $m \in \mathbb{N}^*$  is a parameter set by the user of the method. Let us define  $\mathcal{R}_S(\mathbf{x}, m)$  as the set of all particle positions surrounding  $\mathbf{x}$  within the spatial radius  $r_S(\mathbf{x}, m)$ :

$$\mathcal{R}_S(\mathbf{x}, m) = \{\mathbf{x}', \|\mathbf{x}' - \mathbf{x}\| \leq r_S(\mathbf{x}, m)\}.$$

We define the radius of the temporal Gaussian  $r_T(\mathbf{x}, m)$  as the maximum time difference between  $\mathbf{x}$  and elements of  $\mathcal{R}_S$ :

$$r_T(\mathbf{x}, m) = \max_{\mathbf{x}' \in \mathcal{R}_S(\mathbf{x}, m)} (|t_{\mathbf{x}'} - t_{\mathbf{x}}|).$$

Let us now denote  $\mathcal{X}_{class} \in \{\mathcal{X}_{free}, \mathcal{X}_{sub}, \mathcal{X}_{sup}\}$ . The weight assigned to each particle position  $\mathbf{x}_k \in \mathcal{X}_{class}$  relatively to point of calculation  $\mathbf{x}$ , depending on choice of motion type class and parameter  $m$  is defined as

$$w(\mathbf{x}_k, \mathbf{x}, m, \mathcal{X}_{class}) = \frac{\exp\left[-2\left(\left(\frac{\|\mathbf{x}_k - \mathbf{x}\|}{r_S(\mathbf{x}, m)}\right)^2 + \left(\frac{|t_{\mathbf{x}_k} - t_{\mathbf{x}}|}{r_T(\mathbf{x}, m)}\right)^2\right)\right]}{\sum_{\mathbf{x}_l \in \mathcal{X}_{class}} \exp\left[-2\left(\left(\frac{\|\mathbf{x}_l - \mathbf{x}\|}{r_S(\mathbf{x}, m)}\right)^2 + \left(\frac{|t_{\mathbf{x}_l} - t_{\mathbf{x}}|}{r_T(\mathbf{x}, m)}\right)^2\right)\right]}. \quad (2.10)$$

This function determines the importance of each  $\mathbf{x}_k$  for the estimation functions. It has the advantage of automatically adapting to the data quantity surrounding point of calculation  $\mathbf{x}$  instead of relying on a fixed estimation zone. The closer the particle positions are to  $\mathbf{x}$ , the more importance they are given in the estimation process, avoiding the imprecision that a block-based approach can produce by giving equal importance to every data point located inside the same block. The parameter  $m$  allows the user to indirectly control the kernel radius depending on the size of studied phenomena, and thus to have a single, simple control option over the method.

Now, we can move on to the core of our method: the diffusion and drift estimators. Different estimation formulas are used depending on motion type associated to point of calculation  $\mathbf{x}$ .

**Estimators in the case of free diffusion (or Brownian motion)**

In the case of Brownian motion, we assume that the drift component is null, i.e.  $\boldsymbol{\mu}(\mathbf{x}) = \mathbf{0}$ . The diffusion coefficient for the point  $\mathbf{x} \in \mathcal{X}_{free}$  is given by

$$\hat{\sigma}(\mathbf{x}, m) = \frac{\sum_{\mathbf{x}_k \in \mathcal{X}_{free}} w(\mathbf{x}_k, \mathbf{x}, m, \mathcal{X}_{free}) \|\Delta^{(2)} \mathbf{x}_k\|^2}{2d\Delta t}. \quad (2.11)$$

**Estimators in the case of subdiffusion**

In the case of subdiffusive motion, the drift and diffusion estimators for the point  $\mathbf{x} \in \mathcal{X}_{sub}$  are respectively given by

$$\hat{\boldsymbol{\mu}}(\mathbf{x}, m) = \frac{\sum_{\mathbf{x}_k \in \mathcal{X}_{sub}} w(\mathbf{x}_k, \mathbf{x}, m, \mathcal{X}_{sub}) \Delta^{(1)} \mathbf{x}_k}{\Delta t}, \quad (2.12)$$

$$\hat{\sigma}(\mathbf{x}, m) = \frac{\sum_{\mathbf{x}_k \in \mathcal{X}_{sub}} w(\mathbf{x}_k, \mathbf{x}, m, \mathcal{X}_{sub}) \|\Delta^{(1)} \mathbf{x}_k - \hat{\boldsymbol{\mu}}(\mathbf{x}) \Delta t\|^2}{d\Delta t}. \quad (2.13)$$

**Estimators in the case of superdiffusion**

In the case of superdiffusive motion, the drift and diffusion estimators for the point  $\mathbf{x} \in \mathcal{X}_{sub}$  are respectively given by

$$\hat{\boldsymbol{\mu}}(\mathbf{x}, m) = \frac{\sum_{\mathbf{x}_k \in \mathcal{X}_{sup}} w(\mathbf{x}_k, \mathbf{x}, m, \mathcal{X}_{sup}) \Delta^{(1)} \mathbf{x}_k}{\Delta t}, \quad (2.14)$$

$$\hat{\sigma}(\mathbf{x}, m) = \frac{\sum_{\mathbf{x}_k \in \mathcal{X}_{sup}} w(\mathbf{x}_k, \mathbf{x}, m, \mathcal{X}_{sup}) \|\Delta^{(2)} \mathbf{x}_k\|^2}{2d\Delta t}. \quad (2.15)$$

In the end, all the calculated estimators can be put together on a graph to form a pseudo-map of drift and diffusion consisting of a point cloud and a vector field. The parameter  $m$  allows the user to adjust the size of the kernel (i.e. the sizes of the weighting function's spatial and temporal radiuses) with one single parameter.

**2.2.3 Pseudo-code of the method**

We provide in this section the algorithmic implementation of the method. The algorithm typically takes a single plain text tracking file as input, and a calibration file as an

option. Alternatively, the user can directly use an image sequence that will be tracked by the algorithm.

The algorithm generates the following outputs:

- an illustration of classified input tracks,
- a histogram of input track lengths,
- two maps depicting diffusion and drift in the form of point clouds,
- an interpolated diffusion map,
- a diffusion histogram.

Optionally, it is also possible to obtain two additional diffusion maps where diffusion is calculated by subtrajectories with first and second order estimators. If the user wants time-evolving outputs, two images sequences depicting the evolution of diffusion and drift point clouds with time are computed.

Regarding parameters, the most essential ones are:

- the full filename of the input tracking file (and optionally calibration file), or alternatively image sequence,
- the spatial resolution in x-y axes (in micrometers),
- the spatial resolution in z axis (in micrometers as well),
- the time resolution (in seconds).

If the input of the algorithm is in the form of an image sequence, the user must choose a number of luminescent spots to track. If the tracking data presents aberrant jumps, they have the possibility to use a pre-cutting tool that removes jumps by dividing affected tracks. They also have the possibility to choose between several pre-sorting and classification parameters, but these have default values that can easily be left as is. Next, the kernel size parameter is the aforementioned parameter  $m$ . It is chosen as a number of points and has a default value (e.g.  $m = 600$ , which has empirically proven to be suitable for most data sets). Last, in the display parameters, the user can set the diffusion scale limit, choose to remove outliers in the maps, and optionally display the two additional aforementioned diffusion maps. If the user chose time-evolving outputs, they also have to choose the number of frames, frame rate and overlapping rate of the output image sequence.

---

**Algorithm 1** Dense Diffusion and Drift Mapping Method

---

```

1: procedure D3M2
2:   set(parameters)      ▷ filename; x, y, z, t resolution; pre-cutting, pre-sorting,
   classification, kernel size k and display options
3:   data = load(filename)
4:   Pre-format Data:
5:   data = pre-cut(data) ▷ if chosen by user, this option allows to remove aberrant
   jumps in trajectories
6:   data = pre-sort(data)    ▷ sorts trajectories by lengths, in order to use the
   appropriate classifying method
7:   data = classify(data)    ▷ classifies (sub)trajectories by motion type
8:   data = convert(data)    ▷ makes appropriate units conversions from pixels and
   frames to micrometers and seconds
9:   data = augment(data)    ▷ adds information to the data such as first and second
   order differences (2.8) and (2.9)
10:  Diffusion and Drift Core Calculation:
11:  InitializeVariables
12:  for each trajectory point  $P_i$  do
13:    Calculate all spatial distances between  $P_i$  and the other points
14:    Keep only the ensemble  $K_i$  of the k closest points to  $P_i$ 
15:    Calculate  $r_S(\mathbf{x}, m)$ 
16:    Calculate all temporal distances between  $P_i$  and the other points
17:    Calculate  $r_T(\mathbf{x}, m)$ 
18:    Remove from  $K_i$  the points that do not have the same motion type than  $P_i$ 
19:    Calculate weights with (2.10)
20:    Calculate drift and diffusion through estimator depending on the motion type
   with either (2.11), (2.12,2.13) or (2.14,2.15)
21:  results = clean(output)    ▷ removes redundant or contradictory estimated
   points/vectors
22:  display(results)

```

---

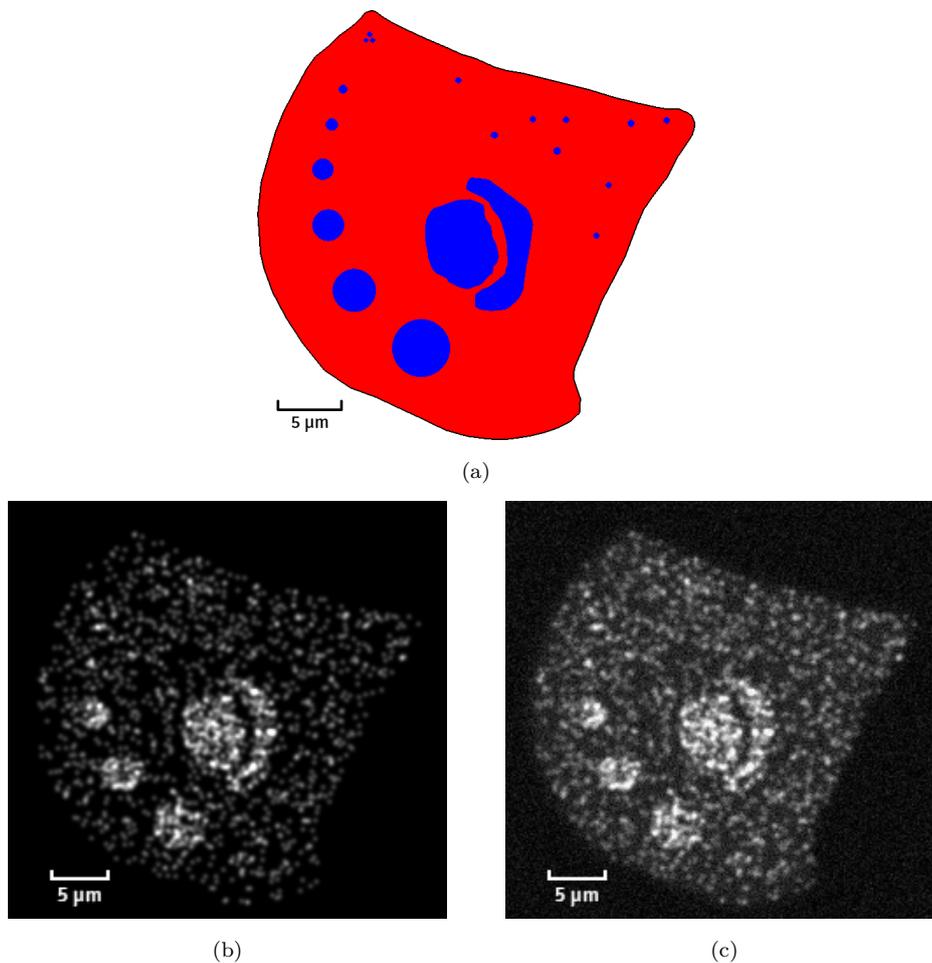


Figure 2.1 – The 2D simulated fluorescence microscopy video is generated with FLUOSIM. Real microscopy data are used to shape the cell into a crossbow. The sequence has a length of 400 frames, a time step of 0.005 s, a spatial resolution of 0.16  $\mu\text{m}$  and includes 2500 moving particles. **(a)** Ground truth diffusion map is shown. The cell contains 20 trapping regions shown in blue where the diffusion coefficient is set to 0.04  $\mu\text{m}^2/\text{s}$ . The rest of the cell has a diffusion coefficient set to 0.16  $\mu\text{m}^2/\text{s}$ . **(b)** Raw video output is obtained from FLUOSIM. **(c)** Augmented video output is derived from (b). A fake grey background was added to the cell, in addition to Gaussian noise. A linear decrease of brightness was added to simulate photobleaching.

## 2.3 Experimental evaluation of the algorithm

To assess the reliability of our method, we conducted several experiments on simulated and real fluorescence microscopy data sets described in this section.

### 2.3.1 Evaluation on simulated data sets

First, we used FLUOSIM [Lagardère et al. 2020] to generate simulated fluorescence microscopy image sequence and tracking data. FLUOSIM is an interactive and unified computer program that simulates protein dynamics for different live-cell imaging methods including SPT. The possibilities of particle motion type are limited to Brownian in its present version, but the program supports the creation of trapping areas from where the particle can exit with more or less probability. In such a region, a particle undergoes a fully stochastic dynamic, but often get out of it with difficulty, depending on the chosen parameters. Even though it does not exactly undergo an Ornstein-Uhlenbeck process, it is then possible to simulate subdiffusive motion by creating small, circular trapping areas.

In Figure 2.1, we illustrate the output of FLUOSIM in the case of a 2D crossbow-shaped cell. We specified 20 trapping areas of different sizes and shapes in which the diffusion is lower than in the cytosol (background). The moving particles were represented as white Gaussian spots as if it was recorded through a diffraction-limited system. The diffusion coefficient was set to  $0.16 \mu\text{m}^2/\text{s}$  inside the cytosol of the cell, except in the trapping areas where it was set to  $0.04 \mu\text{m}^2/\text{s}$ . The sequence shown in Figure 2.1 has a length of 400 frames, a time step of 5 ms, a spatial resolution of  $0.16 \mu\text{m}$  and includes 2500 moving particles. The binding rate of the trapping areas was set to  $1 \text{s}^{-1}$  while the unbinding rate was set to  $3 \text{s}^{-1}$ . The simulation does not feature any kind of superdiffusive movement. After the generation of the image sequence, we added a cell background, Gaussian noise, and a linear brightness fading to simulate photobleaching. The resulting image is shown in Figure 2.1 (b) and (c).

The tracking data were extracted from FLUOSIM in two different ways: first, we saved the ground truth tracks (shown in Figure 2.2(a)) generated by the simulation itself without using any intermediary tracking software on the simulated image sequence. These tracks are not corrupted by noise. Second, we saved the image sequence generated with FLUOSIM, but augmented with a fake cell background and a linear decrease of brightness to simulate photobleaching, and corrupted by artificial Gaussian noise. We used the augmented image sequence as an input for the software ICY, and tracked the particles with its embedded tracking module based on Multiple Hypothesis Tracking (MHT) [Blackman 2004] [Chenouard, Bloch, and Olivo-Marin 2013]. The tracks are shown in Figure 2.6(a).

In the two next paragraphs, we compare and discuss the outputs generated from these two versions of simulated tracking data with our method, but also with [Hozé, Nair, et al. 2012] whose MATLAB code is available on [bionewmetrics.org](http://bionewmetrics.org), and [Beheiry, Dahan, and

Masson 2015] whose software is available on [research.pasteur.fr](http://research.pasteur.fr).

### Experiments performed on ground truth tracks

We present here the results obtained with the different methods on uncorrupted ground truth tracks, directly extracted from FLUOSIM. Our method produced the outputs shown in Figure 2.2 regarding tracking data analysis, and in Figure 2.3 regarding diffusion maps. Figure 2.4 and Figure 2.5 illustrate the output obtained with [Hozé, Nair, et al. 2012] and [Beheiry, Dahan, and Masson 2015], respectively.

Figure 2.2(a) shows the entire input tracks on a single image while Figure 2.2(b) illustrates the motion classification performed prior to the actual map calculation thanks to [Briane, Kervrann, and Vimond 2018] [Briane, Vimond, Valades-Cruz, et al. 2020]. The data consists of 5640 trajectories totalizing 752 552 particle positions, which makes it a large data set. In this experiment, 94.35 % of the positions were classified as Brownian, while 2.73 % and 2.92 % were classified as subdiffusive and superdiffusive, respectively. This means that roughly 5% of the trajectories were wrongly classified either as superdiffusive, or as subdiffusive while they are outside the small confinement regions. We might as well notice the resemblance of the simulated cell shape with a real-life crossbow-shaped cell. Figure 2.2(c) for its part displays the track length histogram. We can notice that although the image sequence has a length of 400 frames, not all tracks are that long. FLUOSIM is indeed able to simulate the activation and deactivation of fluorophores: when the fluorophores are "turned-off", the generated tracks are interrupted and cut in several parts of inferior lengths.

The mapping results obtained with our method are illustrated in Figure 2.3. In particular, Figure 2.3(a) displays the diffusion point cloud output, while Figure 2.3(b) displays the interpolated diffusion bitmap inferred from the point cloud and Figure 2.3(c) represents the diffusion histogram corresponding to the point cloud. Our method took about 90 min processing time to generate the outputs. Albeit slow because of the file size, it provided an accurate and precise reconstruction of diffusion in comparison to simulation ground truth presented in Figure 2.1(a), both in terms of spatial localization and diffusion values, as shown on the diffusion map and its associated histogram.

Figure 2.4 shows the outputs obtained with [Hozé, Nair, et al. 2012] on the same data set: the first column illustrates to the raw output obtained with the method with different resolutions (1.2, 0.7 and 0.2  $\mu\text{m}$ ), while the second column corresponds to the same outputs with Gaussian blurring ( $\sigma = 0.8$ ) applied, and the third column displays the

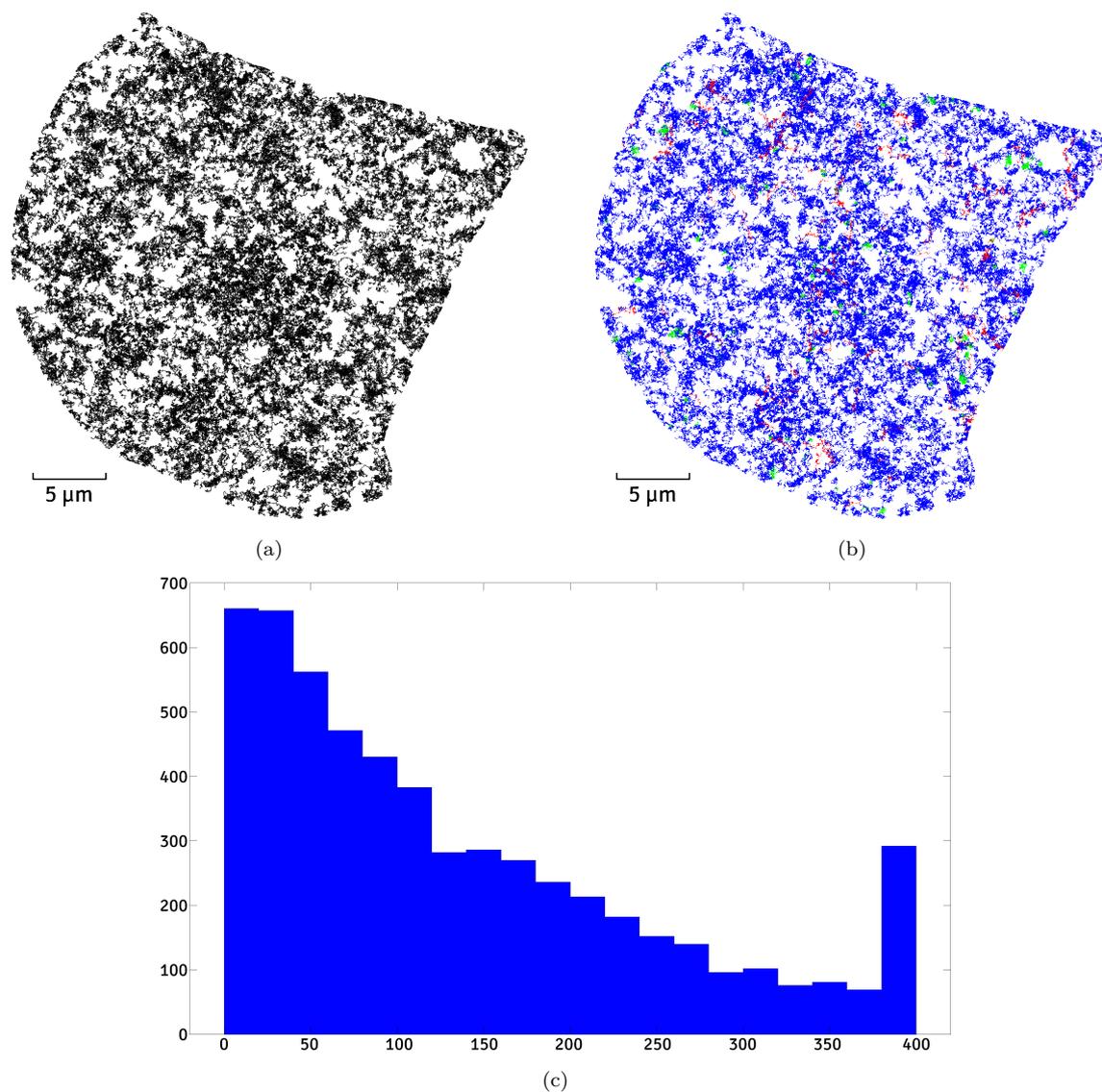


Figure 2.2 – The ground truth tracks are directly extracted from the FLUOSIM-simulated data set and analyzed. **(a)** The set of all simulated particle trajectories is shown. **(b)** Particle trajectories are classified by motion type. Free diffusion is represented in blue, subdiffusion is represented in green and superdiffusion is represented in red. **(c)** The histogram of the input trajectory lengths is shown.

interpolated blurred maps of the second column. The method [Hozé, Nair, et al. 2012] was very fast with a processing time of less than 1 s, and was able to provide an accurate and precise diffusion map, especially at the finest resolution of 0.2  $\mu\text{m}$  with Gaussian blurring and interpolation. It generated results similar to ours, even though the small confinement

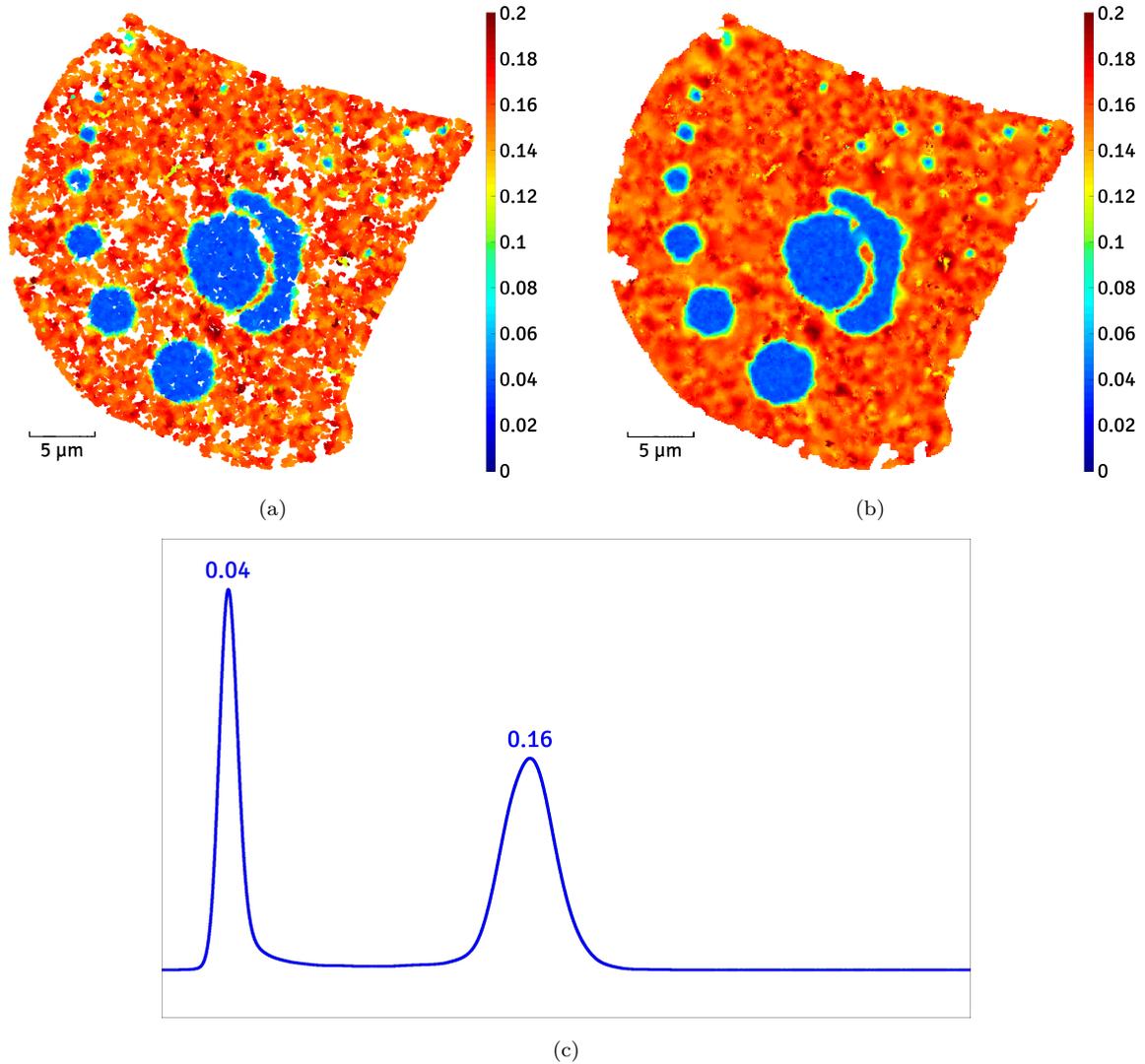


Figure 2.3 – The results obtained with our method on simulated ground truth tracks are shown. **(a)** The point cloud output represents the spatial distribution of the diffusion coefficient inside the cell, in  $\mu\text{m}^2/\text{s}$ . **(b)** The point cloud is derived into a map by interpolation, with the same scale units. **(c)** The point cloud is derived into a diffusion histogram. The values displayed on top of the modes are also in  $\mu\text{m}^2/\text{s}$

regions might not be as sharply represented. We can however notice the presence of stains in the largest confinement regions of the interpolated map, due to side effects produced by the lack of tracking data in some calculation blocks.

Figure 2.5 illustrates the outputs obtained with the INFERENCENMAP [Beheiry, Dahan, and Masson 2015]. The first three images show diffusion maps generated from Voronoi tessellations with different resolutions with a range of 20 to 35 min processing time, while

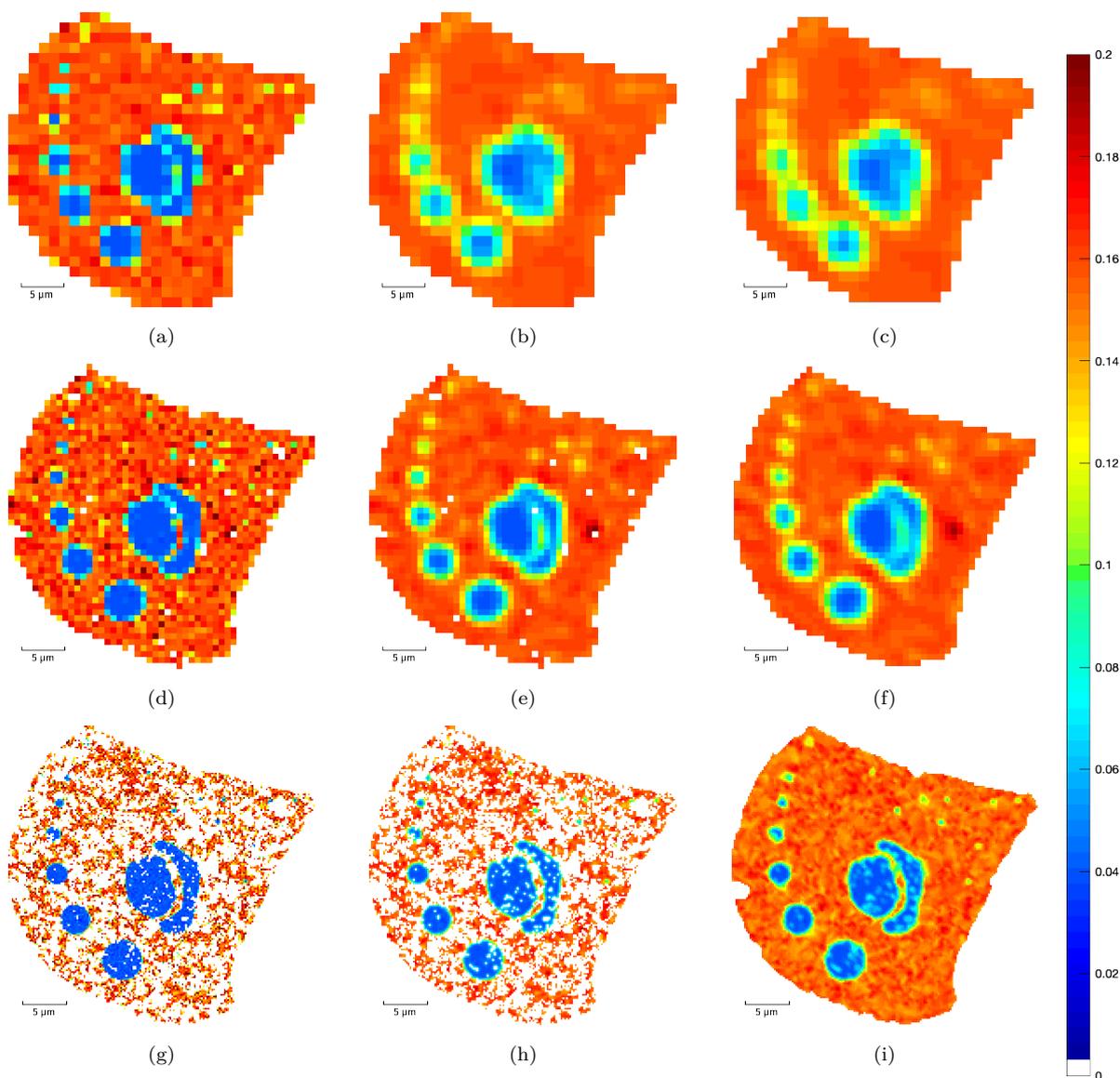


Figure 2.4 – The results obtained with method [Hozé, Nair, et al. 2012] on simulated ground truth tracks are shown. **(a-c)** Diffusion maps are generated using a coarse resolution parameter of  $1.2\ \mu\text{m}$ , with from left to right: the raw map, the Gaussian-filtered ( $\sigma = 0.8$ ) map, the Gaussian-filtered and interpolated diffusion map. **(d-f)** Diffusion maps are generated using a medium resolution parameter of  $0.7\ \mu\text{m}$ , and shown in the same order. **(g-i)** Diffusion maps are generated using a fine resolution parameter of  $0.2\ \mu\text{m}$ , and shown in the same order. The diffusion scale units are in  $\mu\text{m}^2/\text{s}$

the fourth one corresponds to a diffusion map generated from a quadtree in about 1 to 2 min. This method also produced good results, but the standard deviation between each

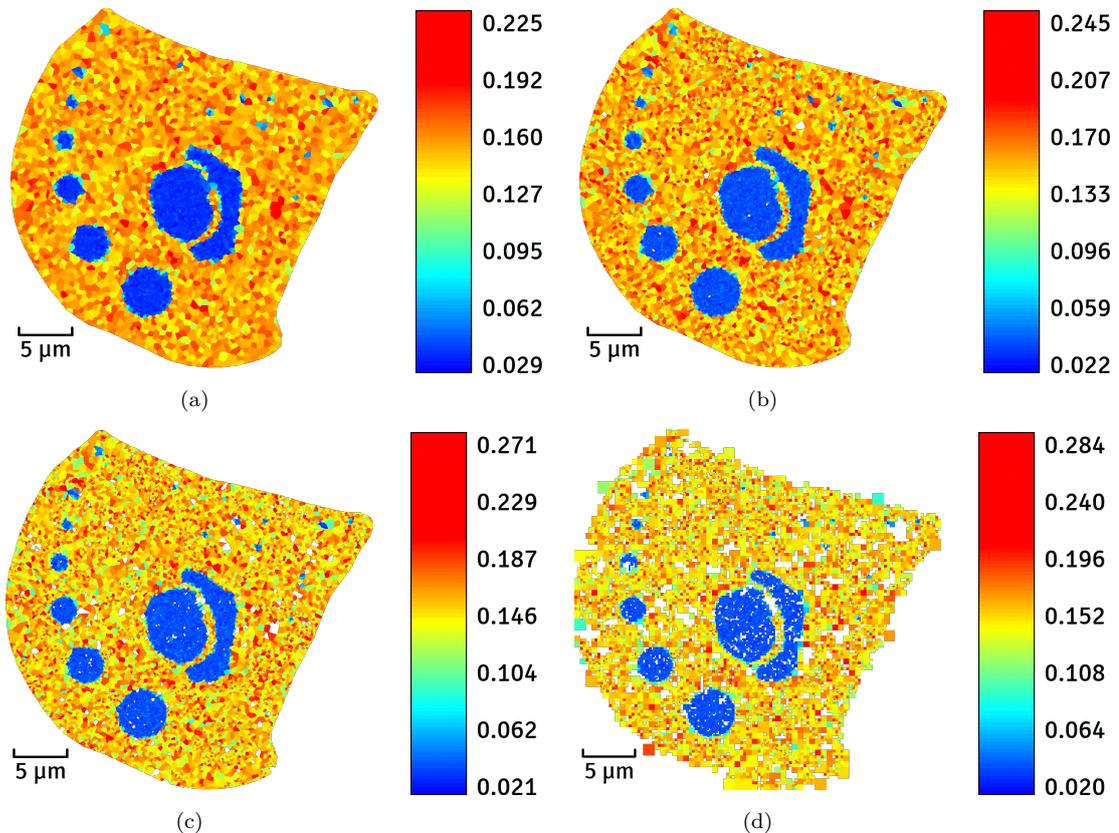


Figure 2.5 – The results obtained with method [Beheiry, Dahan, and Masson 2015] on simulated ground truth tracks are shown. **(a-c)** Diffusion maps are generated with the Voronoi tessellation mode, using respectively from left to right: the minimum number of zones (4700), the default number of zones (9401), and the maximum number of zones (14102). **(d)** A diffusion map is generated with the Quad-Tree mesh mode. The diffusion scale units are in  $\mu\text{m}^2/\text{s}$ .

tile of the tessellation is high and can be misleading. Also, the smaller confinement regions are not as sharply represented as with the other two methods. The same conclusions can be applied to the method using a quadtree instead of a Voronoi tessellation, as seen in Figure 2.5(d).

### Experiments on tracking data sets

As explained above, the first simulation experiment allowed us to generate additional tracking data by re-tracking an augmented version of the simulated image sequence with the MHT ICY plugin. Similarly to the run performed on ground truth tracks, our method produced an analysis of tracking data and diffusion maps, as shown in Figure 2.6 and

Figure 2.7, respectively. We included the results obtained with [Hozé, Nair, et al. 2012] and [Beheiry, Dahan, and Masson 2015] as well, as shown in Figure 2.8 and Figure 2.9.

Similarly to Figure 2.2, Figure 2.6 illustrates the data analysis preliminarily done by our method on the ICY-tracked data. More specifically, Figures 2.6(a), (b) and (c) respectively illustrate the input tracking data set, the motion classification results, and the track length histogram. The number of trajectories detected in this data set was higher than ground truth with 6496 tracks, but the overall number of particle positions is significantly lower, with 95 350 points. This implies that the obtained tracks are drastically shorter than ground truth tracks. In this experiment, 87.16 % of particle positions were classified as Brownian while 12.81 % were classified as subdiffusive. Rightfully no position was classified as superdiffusive, but this seems to be more related to tracking flaws than to good accuracy. The density of ICY-obtained tracks clearly shows that tracking the bright spots on the simulated image sequence (as it would be done on real biological data) significantly reduces the amount of available input data compared to using the tracking ground truth.

As shown in Figure 2.3, Figure 2.7 displays the mapping results obtained with our method: Figures 2.7(a), (b) and (c) respectively illustrate the diffusion point cloud output, the interpolated diffusion bitmap inferred from the point cloud, and the diffusion histogram corresponding to the point cloud. Processing was significantly faster with about 3 min of calculation time. Our method overestimated the diffusion coefficient, but still partially preserve the structure of the cell in comparison to the ground truth, with the largest confinement zones still visible, as shown on the diffusion map and its associated histogram.

Figure 2.8 displays the outputs obtained with [Hozé, Nair, et al. 2012] in the same way as in Figure 2.4, that is with block resolutions of 1.2, 0.7 and 0.2  $\mu\text{m}$  for the three lines, respectively, and raw, Gaussian-filtered (with  $\sigma = 0.8$ ), interpolated outputs for the three columns, respectively. Similarly as in the ground truth case, the method was very fast with less than 1 s processing time. However, it was not able to provide any accurate information in the diffusion map, no matter the resolution, filtering and interpolation used. In particular, the cases with smaller resolutions can be highly misleading as they appear to give some information corresponding to nothing that was simulated.

Figure 2.9 is the analogue of Figure 2.5, that is results obtained with INFERENCEMAP [Beheiry, Dahan, and Masson 2015], respectively generated from a Voronoi tessellation of three different resolutions (in less than 1 min) as well as one generated from a quadtree

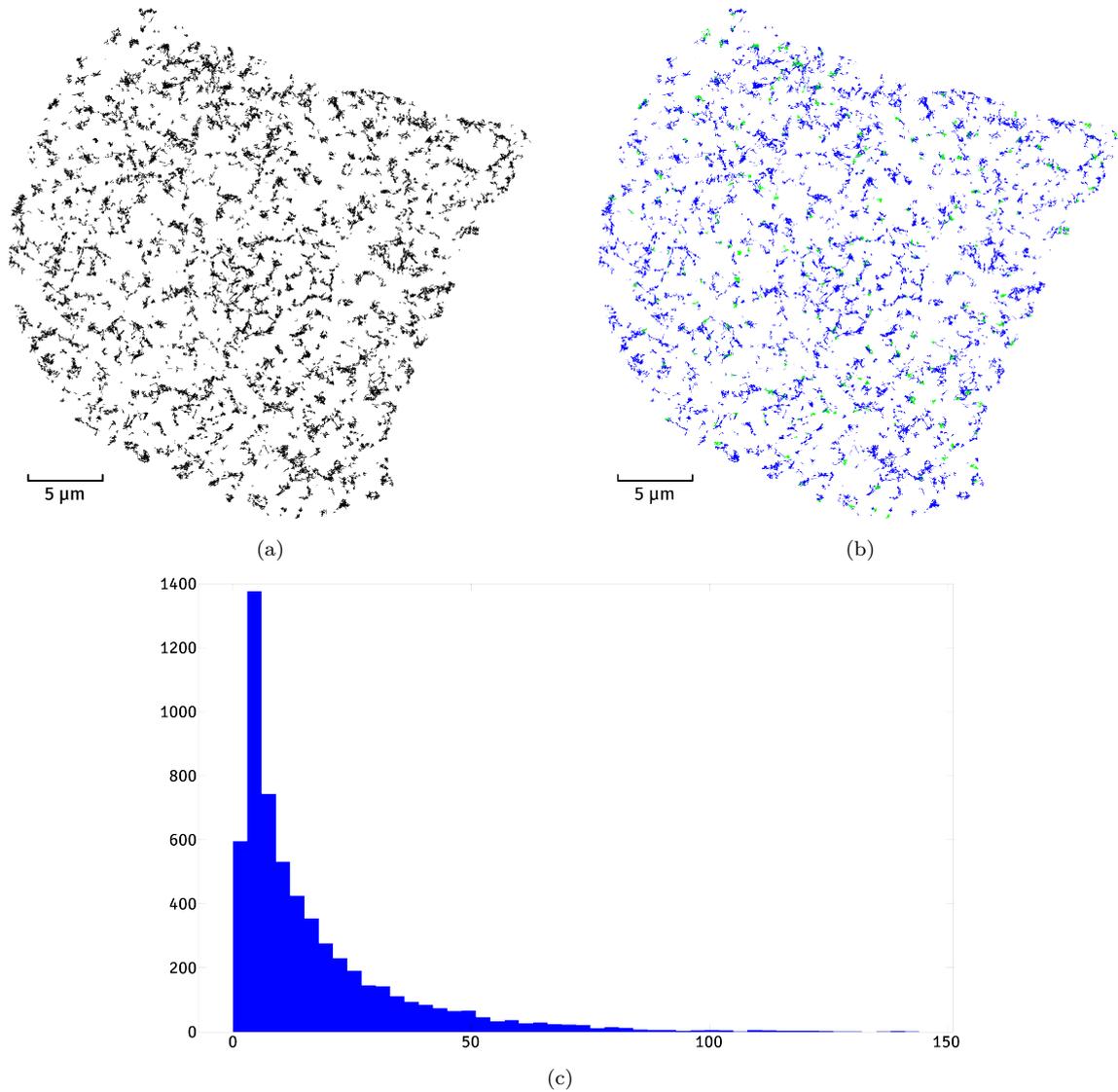


Figure 2.6 – The particles are tracked with ICY from the simulated video and analyzed. **(a)** The set of all simulated particle trajectories is shown. **(b)** Particle trajectories are classified by motion type. Free diffusion is represented in blue, subdiffusion is represented in green and superdiffusion is represented in red. **(c)** The histogram of the input trajectory lengths is shown.

(in a few seconds). This method was also unable to provide any exploitable results, no matter the mode or number of Voronoi tessellation tiles.

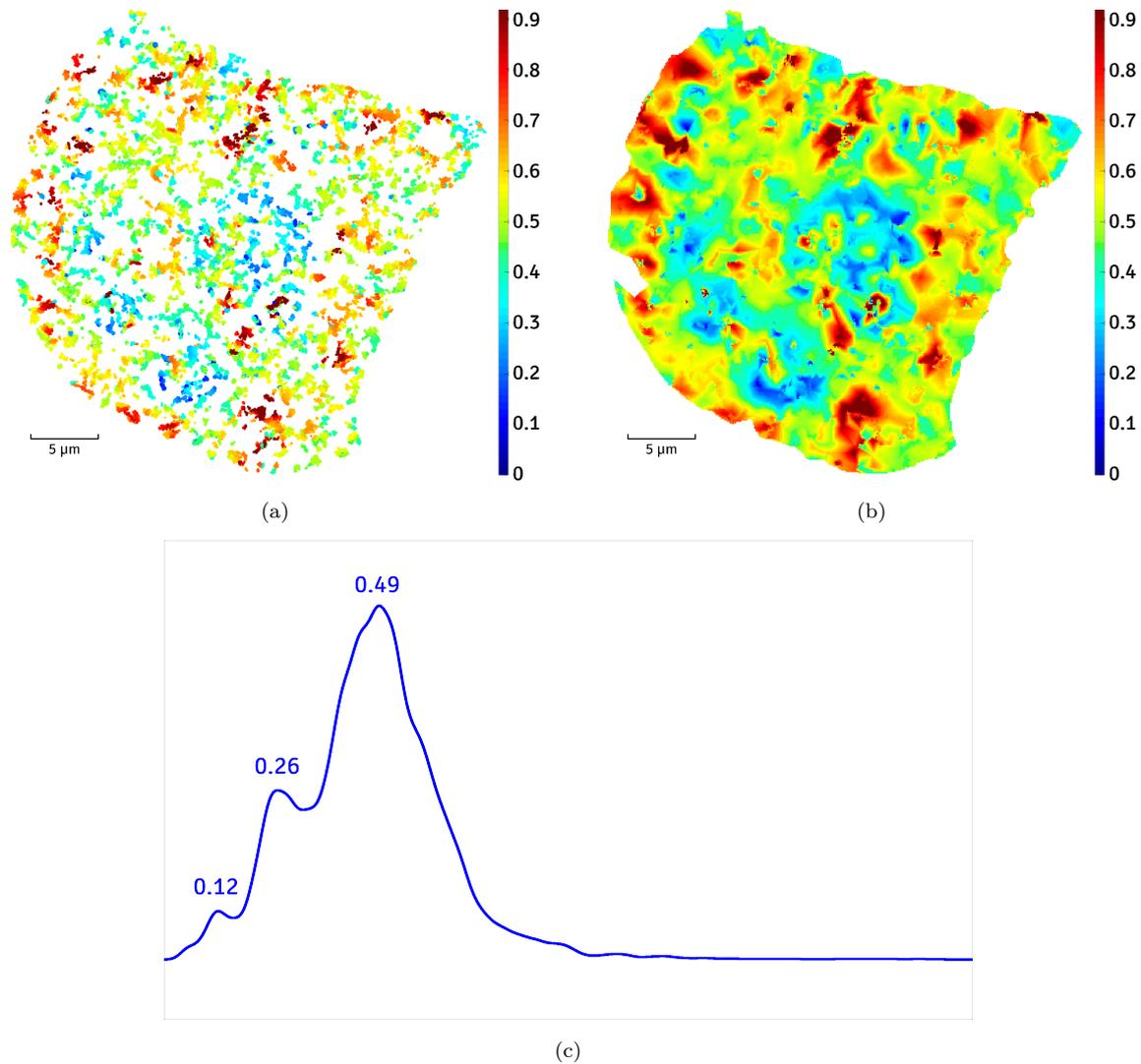


Figure 2.7 – The results obtained with our method on the data tracked with ICY from the FLUOSIM-simulated video, are shown. **(a)** The point cloud output represents the spatial distribution of the diffusion coefficient inside the cell, in  $\mu\text{m}^2/\text{s}$ . **(b)** The point cloud is derived into a map by interpolation, with the same scale units. **(c)** The point cloud is derived into a diffusion histogram. The values displayed on top of the modes are also in  $\mu\text{m}^2/\text{s}$

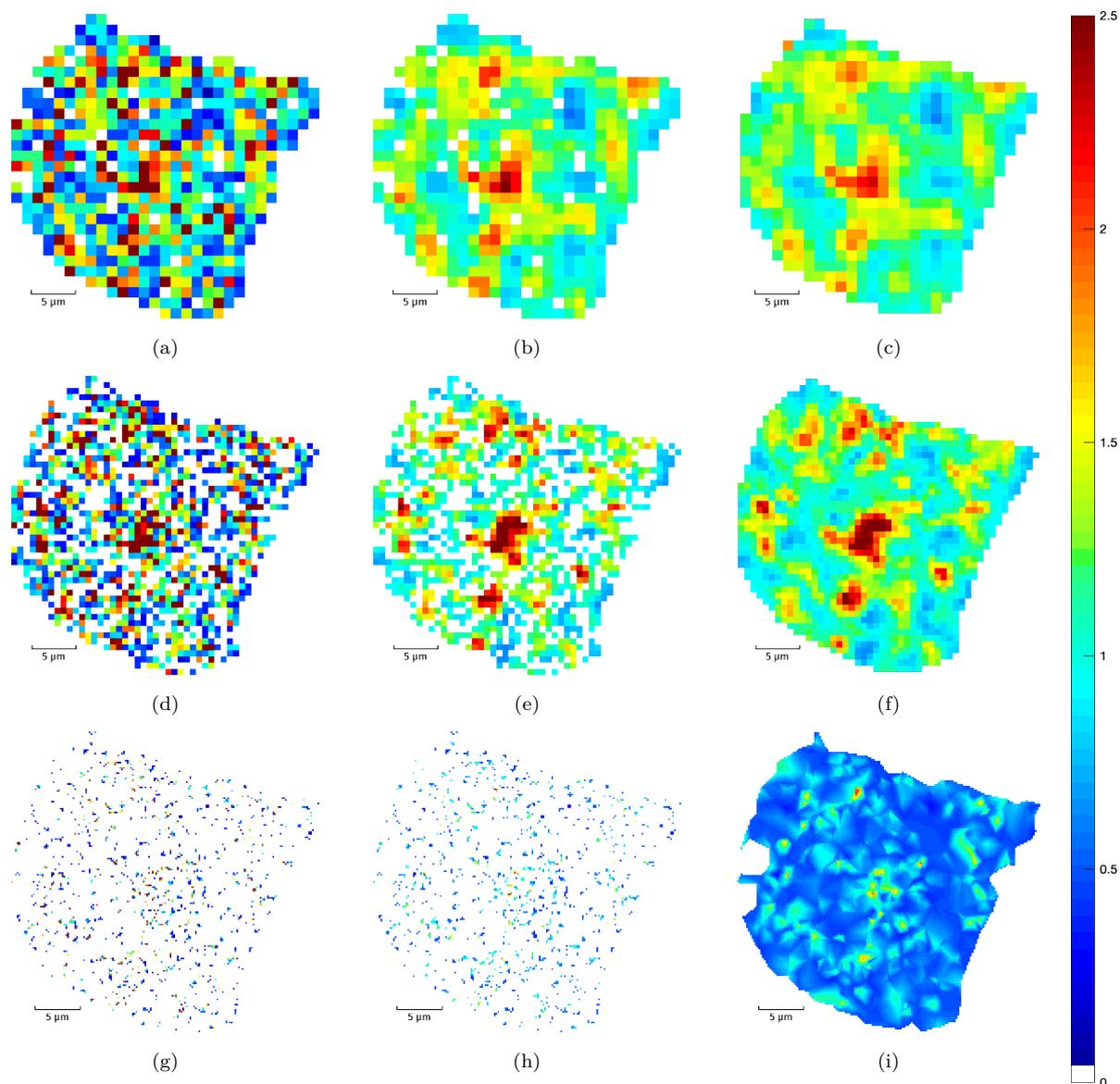


Figure 2.8 – The results obtained with method [Hozé, Nair, et al. 2012] on the data tracked with ICY from the FLUOSIM-simulated video, are shown. **(a-c)** Diffusion maps are generated using a coarse resolution parameter of 1.2  $\mu\text{m}$ , with from left to right: the raw map, the Gaussian-filtered ( $\sigma = 0.8$ ) map, the Gaussian-filtered and interpolated diffusion map. **(d-f)** Diffusion maps are generated using a medium resolution parameter of 0.7  $\mu\text{m}$ , and shown in the same order. **(g-i)** Diffusion maps are generated using a fine resolution parameter of 0.2  $\mu\text{m}$ , and shown in the same order. The diffusion scale units are in  $\mu\text{m}^2/\text{s}$

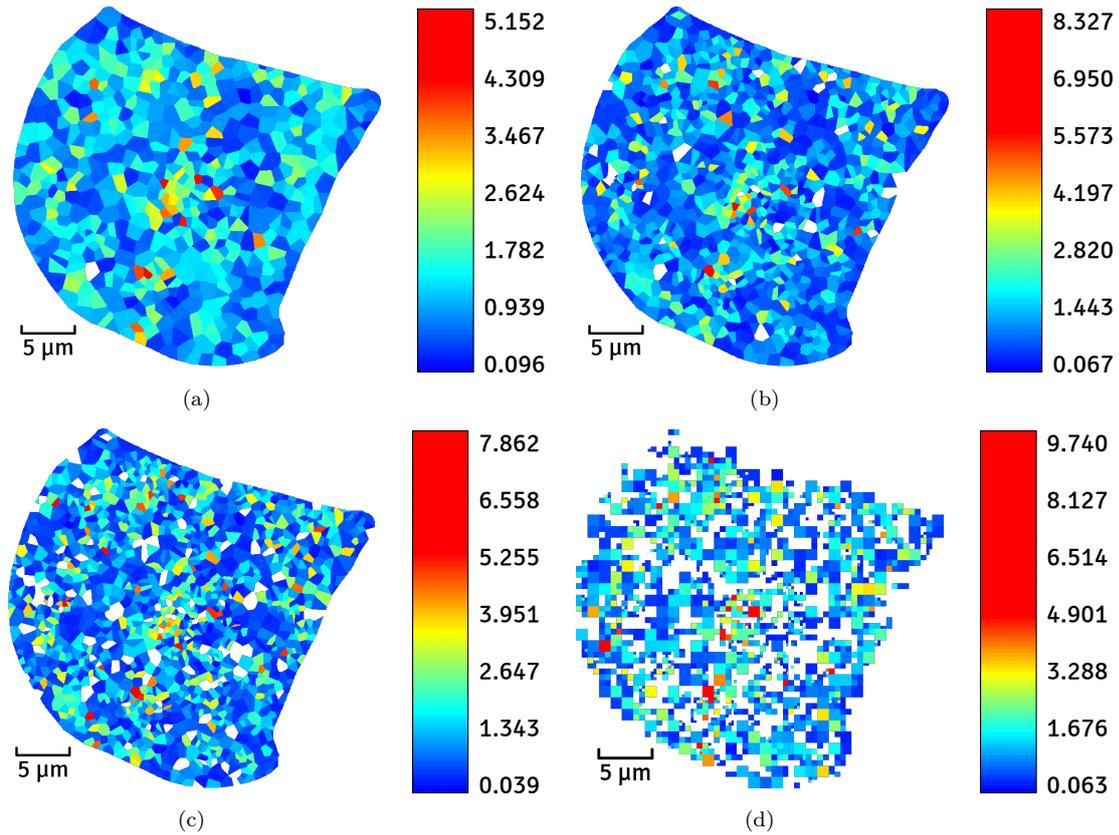


Figure 2.9 – The results obtained with method [Beheiry, Dahan, and Masson 2015] on the data tracked with ICY from the FLUOSIM-simulated video, are shown. **(a-c)** Diffusion maps are generated with the Voronoi tessellation mode, using respectively from left to right: the minimum number of zones (611), the default number of zones (1223), and the maximum number of zones (1834). **(d)** A diffusion map is generated with the Quad-Tree mesh mode. The diffusion scale units are in  $\mu\text{m}^2/\text{s}$ .

### 2.3.2 Evaluation on real data sets

The method has been assessed on real fluorescence microscopy data generated with various microscopy technologies such as SPT-PALM, LLSM and TIRF. Due to the absence of ground truth, we can only give a qualitative assessment of the results, based on what is expected to occur in the specimen.

#### Analysis of dynamics of Su(H) transcription factors

For this first real data experiment, we used tracking data kindly provided by the Bray Lab, Department of Physiology, Development and Neuroscience at the University of Cambridge, which were originally used in [Gomez Lamarca et al. 2018]. It was generated by the MHT plugin in ICY from a PALM image sequence depicting Su(H) transcription factors attached to mEOS3.2 fluorescent proteins in *Drosophila* salivary gland cells. Our method produced the outputs shown in Figure 2.10 regarding tracking data analysis and in Figure 2.11 regarding diffusion and drift maps. Figure 2.12 and Figure 2.13 illustrate the outputs obtained with [Hozé, Nair, et al. 2012] and [Beheiry, Dahan, and Masson 2015], respectively.

The data analysis (see Figure 2.10) shows that this data set is very small, and that the track density is dissimilar between the center of the specimen and the rest of the data set, as shown in Figure 2.10(a). The data set consists here of 2937 trajectories and 13 410 particle positions. Furthermore, the vast majority of tracks are shorter than 10 frames as shown in Figure 2.10(c). As such, most tracks (90.90 %) were classified as Brownian due to insufficient length, while the rest (9.10 %) was classified as subdiffusive, as shown in Figure 2.10(b).

Similarly to experiments on simulated data, Figure 2.11 displays the mapping results obtained with our method: (a) shows the diffusion point cloud output, (b) represents the interpolated diffusion bitmap inferred from the point cloud, and (c) is the diffusion histogram corresponding to the point cloud. Our method generated the outputs in about 30 s. Even though the number of tracks is small, it seems that our method was able to recover some interesting information from this data set. In particular, we can notice a region of lower diffusion in the center of the specimen, corresponding to the region of higher point density. If we examine the drift map in Figure 2.11(d), we can also notice a circular drift centered in the area. This might correspond to the position of the locus.

Figure 2.12 displays the outputs obtained with [Hozé, Nair, et al. 2012]. The first

column corresponds to the raw output from the method with different resolutions (1.2, 0.7 and 0.2  $\mu\text{m}$ ), while the second column corresponds to the same outputs with Gaussian blurring ( $\sigma = 0.8$ ) applied. The third column corresponds to the interpolated blurred maps of the second column. This method was very fast (less than 1 s processing time) but not able to provide any accurate information in the diffusion map, no matter the resolution, filtering and interpolation used. In particular, the cases with smaller resolutions can be highly misleading as they appear to highlight imaginary features in the specimen.

Figure 2.13 shows the outputs obtained with the software INFERENCEMAP [Beheiry, Dahan, and Masson 2015]. The three first images illustrate diffusion maps generated from a Voronoi tessellation (processing time in a range of 15 to 30 s) with different resolutions and the fourth one corresponds to a diffusion map generated from a quadtree (processing time in a range of a few seconds). It was also unable to provide and exploitable results, no matter the mode or number of Voronoi tessellation tiles.

### **Analysis of dynamics of calnexin plasmids in HEK-293 cells**

In this experiment, we used tracking data extracted from a HILO (pseudo-TIRF) image sequence acquired with a Zeiss Elyra microscope at Cambridge Institute for medical Research, showing mEOS2-calnexin plasmids in HEK-293 cells, originally used in [Holcman, Parutto, et al. 2018]. Tracking was performed using the particle tracking module of the microscope software (Elyra Zen edition, Zeiss), and the data is available on [bionewmetrics.org](http://bionewmetrics.org). Our method produced the outputs shown in Figure 2.14 regarding tracking data analysis and in Figure 2.15 for the diffusion maps. Figure 2.16 illustrates the outputs obtained with [Hozé, Nair, et al. 2012], as in [Holcman, Parutto, et al. 2018].

Regarding the analysis conducted on the data, Figure 2.14(a), (b) and (c) display tracks, motion classification and track length distribution, respectively. The data set includes a total of 8180 trajectories and 43 088 particle positions. Once again, the trajectories were very short, which led to the vast majority of them being classified as Brownian (99.91 %). Only a very small fraction was classified as subdiffusive (0.09 %), and no position was classified as superdiffusive.

Figure 2.16 shows the results obtained with [Hozé, Nair, et al. 2012]. As mentioned earlier, this method was fast with less than 1 s processing time. We used a single spatial resolution of 0.4  $\mu\text{m}$  in Figure 2.16(a) to mimic the original Figure in [Holcman, Parutto, et al. 2018], while Figure 2.16(b) shows the same map with additional Gaussian blurring ( $\sigma = 0.8$ ). Figure 2.16(c) on its part displays the interpolated version of Figure 2.16(b).

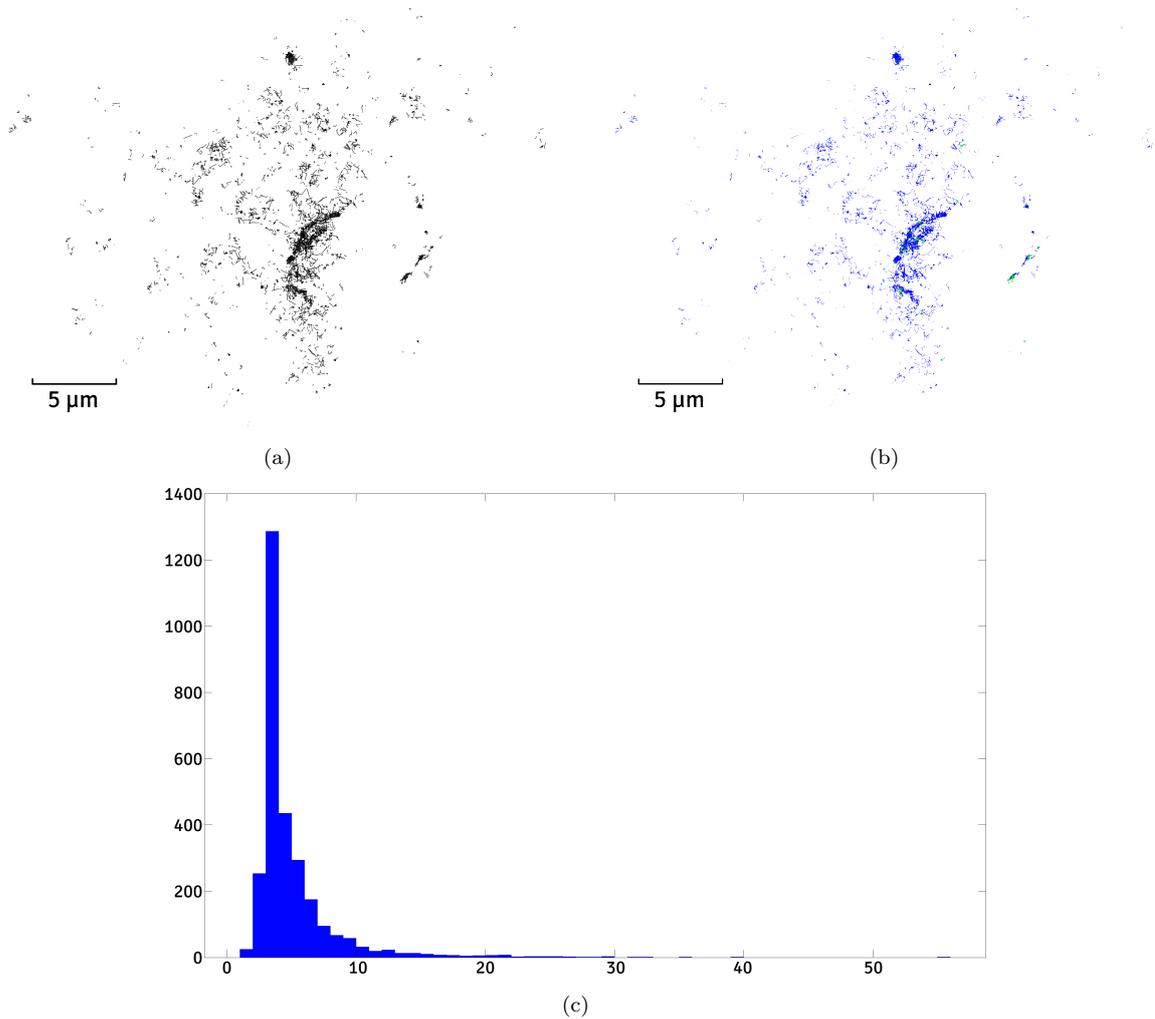


Figure 2.10 – The tracking data obtained from the transcription factors data set is analyzed. **(a)** The set of all simulated particle trajectories is shown. **(b)** Particle trajectories are classified by motion type. **Free diffusion** is represented in **blue**, **subdiffusion** is represented in **green** and **superdiffusion** is represented in **red**. **(c)** The histogram of the input trajectory lengths is shown.

Regarding our results, Figure 2.15(a) shows the diffusion point cloud output and Figure 2.15(b) displays the interpolated diffusion bitmap inferred from the point cloud. Figure 2.15(c) on is part illustrates the diffusion histogram corresponding to the point cloud. Our method took 2 to 3 min to process the data. It is interesting to note that our output is consistent with the output obtained with [Hozé, Nair, et al. 2012]: the diffusion regions spatially correspond with similar values. That said, our method provides more details.

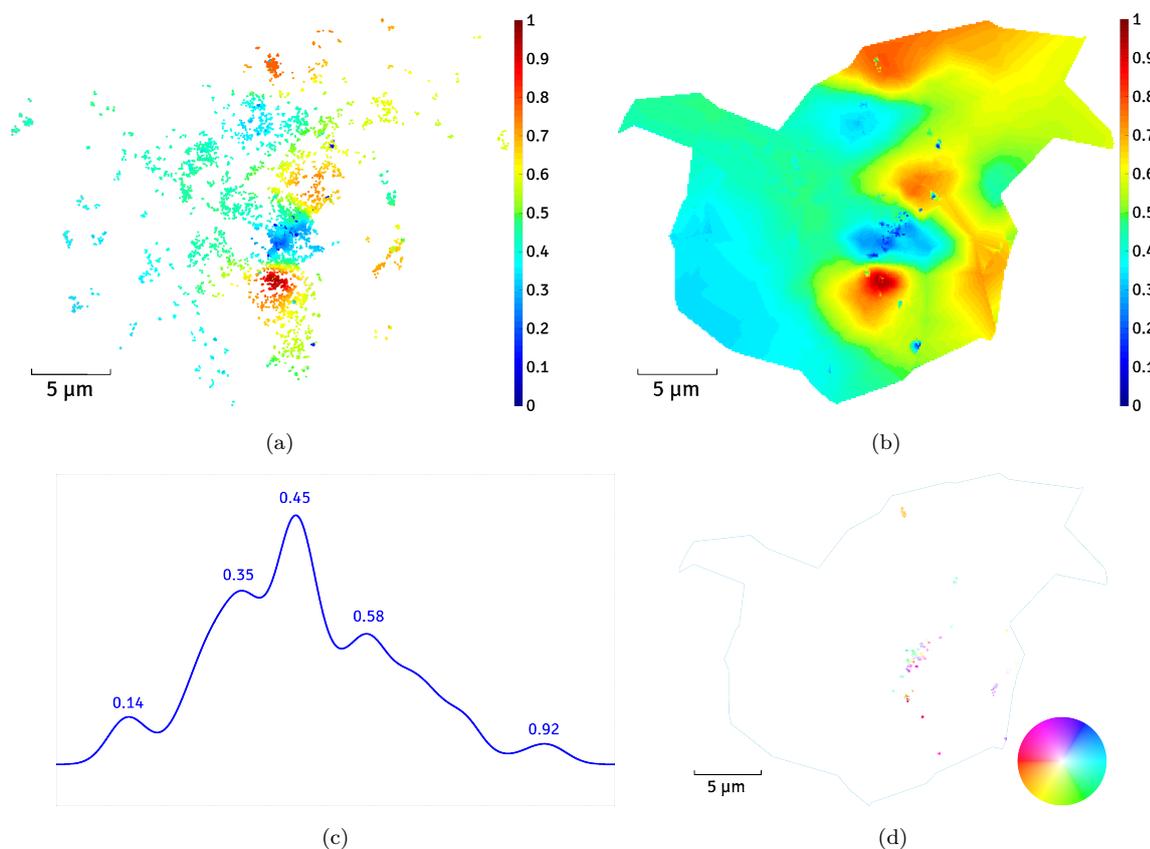


Figure 2.11 – The results obtained with our method on real data (transcription factors), are shown. **(a)** The point cloud output represents the spatial distribution of the diffusion coefficient inside the cell, in  $\mu\text{m}^2/\text{s}$ . **(b)** The point cloud is derived into a map by interpolation, with the same scale units. **(c)** The point cloud is derived into a diffusion histogram. The values displayed on top of the modes are also in  $\mu\text{m}^2/\text{s}$ . **(d)** This second point cloud represents the spatial distribution of the drift inside the cell. Each point represents a drift vector. The wheel indicates both the direction and the amplitude of the vectors.

### Analysis of dynamics of HaloTagged CRT in HEK-293 cells

Similarly to the previous experiment, the image sequence used here was downloaded from [bionewmetrics.org](http://bionewmetrics.org), depicts HEK-293 cells and was originally used in [Holcman, Parutto, et al. 2018]. Acquisition and tracking were performed at Cambridge Institute for medical Research with a Zeiss Elyra microscope in TIRF/HILO mode and the microscope software, respectively. This time however, the tracked molecule is TMR-labelled Halo-tagged CRT. As previously, Figure 2.17, Figure 2.18 and Figure 2.19 illustrate our tracking data analysis, our diffusion maps and the output obtained with [Hozé, Nair, et al.

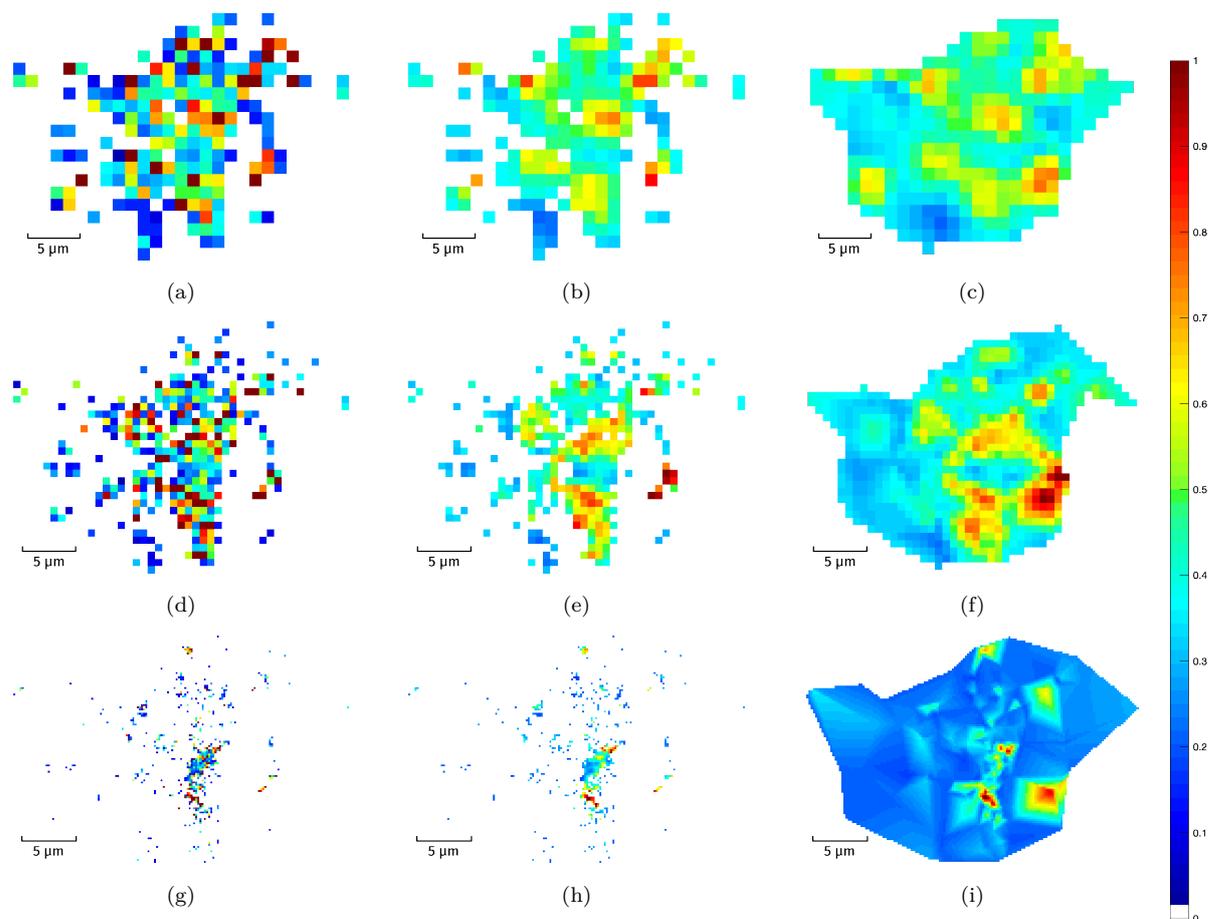


Figure 2.12 – The results obtained with method [Hozé, Nair, et al. 2012] on real data (transcription factors), are shown. **(a-c)** Diffusion maps are generated using a coarse resolution parameter of  $1.2\ \mu\text{m}$ , with from left to right: the raw map, the Gaussian-filtered ( $\sigma = 0.8$ ) map, the Gaussian-filtered and interpolated diffusion map. **(d-f)** Diffusion maps are generated using a medium resolution parameter of  $0.7\ \mu\text{m}$ , and shown in the same order. **(g-i)** Diffusion maps are generated using a fine resolution parameter of  $0.2\ \mu\text{m}$ , and shown in the same order. The diffusion scale units are in  $\mu\text{m}^2/\text{s}$

2012], respectively.

Figure 2.17 illustrates the data analysis conducted on the data, with the same subfigures as in the previous experiment and the same short trajectory problem leading to most of the particle positions being classified as Brownian. Of all 107 306 particle positions distributed in 21 708 trajectories, 93.17% were classified as Brownian while 6.83% were classified as subdiffusive.

Concerning the outputs obtained with [Hozé, Nair, et al. 2012], we used a spatial

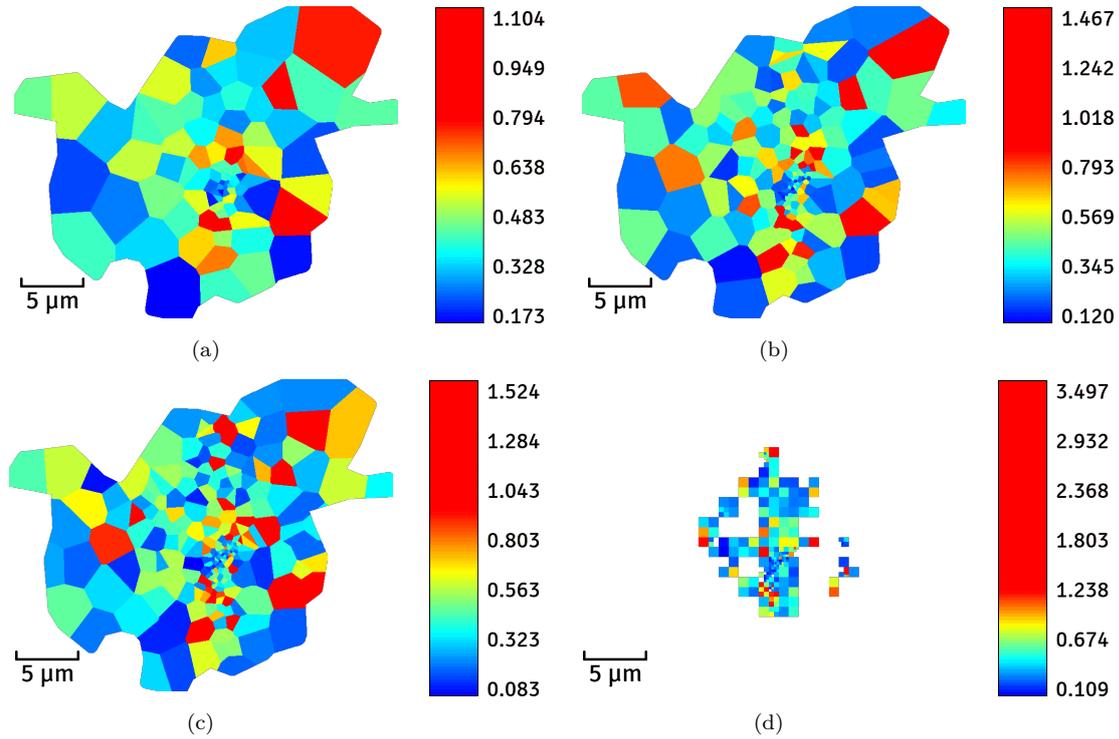


Figure 2.13 – The results obtained with method [Beheiry, Dahan, and Masson 2015] on real data (transcription factors), are shown. **(a-c)** Diffusion maps are generated with the Voronoi tessellation mode, using respectively from left to right: the minimum number of zones (86), the default number of zones (172), and the maximum number of zones (258). **(d)** A diffusion map is generated with the Quad-Tree mesh mode. The diffusion scale units are in  $\mu\text{m}^2/\text{s}$ .

resolution of  $0.2\mu\text{m}$  in Figure 2.19 to mimic the original Figure in [Holcman, Parutto, et al. 2018]. Similarly as in Figure 2.16, Figure 2.19(b) and (c) depict the blurred and interpolated versions of Figure 2.19(a). As usual, less than 1s was necessary to process the outputs.

Figure 2.18 illustrates the results obtained with our method in the same way as explained above. Our method took approximately 6 min to process the data. As in the previous experiment, our output is consistent with the output obtained with [Hozé, Nair, et al. 2012], with diffusion regions once again spatially corresponding with similar values. Nevertheless, our method gives a more direct, sharp and informative map than [Hozé, Nair, et al. 2012], as there is no need for any Gaussian blur.

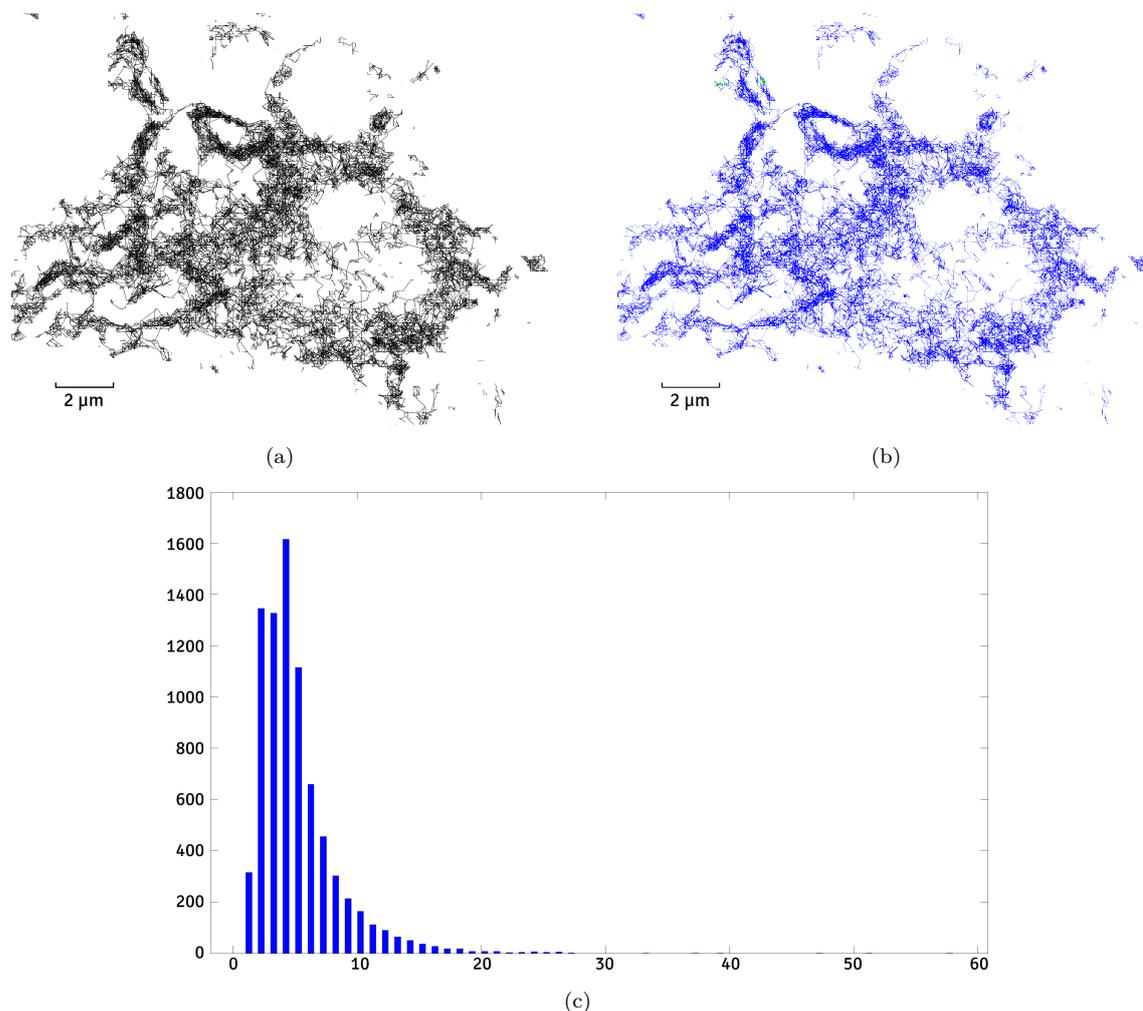


Figure 2.14 – The tracking data obtained from the mEOS2-calnexin plasmids in HEK-293 cells data set is analyzed. **(a)** The set of all simulated particle trajectories is shown. **(b)** Particle trajectories are classified by motion type. **Free diffusion** is represented in **blue**, **subdiffusion** is represented in **green** and **superdiffusion** is represented in **red**. **(c)** The histogram of the input trajectory lengths is shown.

### Analysis of 3D dynamics of galectin-3 in SUM159 cell

To illustrate the possibility of using 3D data sets, we used 3D tracking data extracted from a LLSM fluorescence microscopy image sequence showing Atto647-N-attached Galectin-3 in a SUM159 cell. The images were acquired by the joint SERPICO/STED team from the Inria Center at the University of Rennes and the Curie Institute, and tracking was carried out using U-TRACK [Jaqaman et al. 2008] [Roudot, Legant, et al. 2022].

Our method produced the outputs shown in Figure 2.20 for the tracking data classifi-

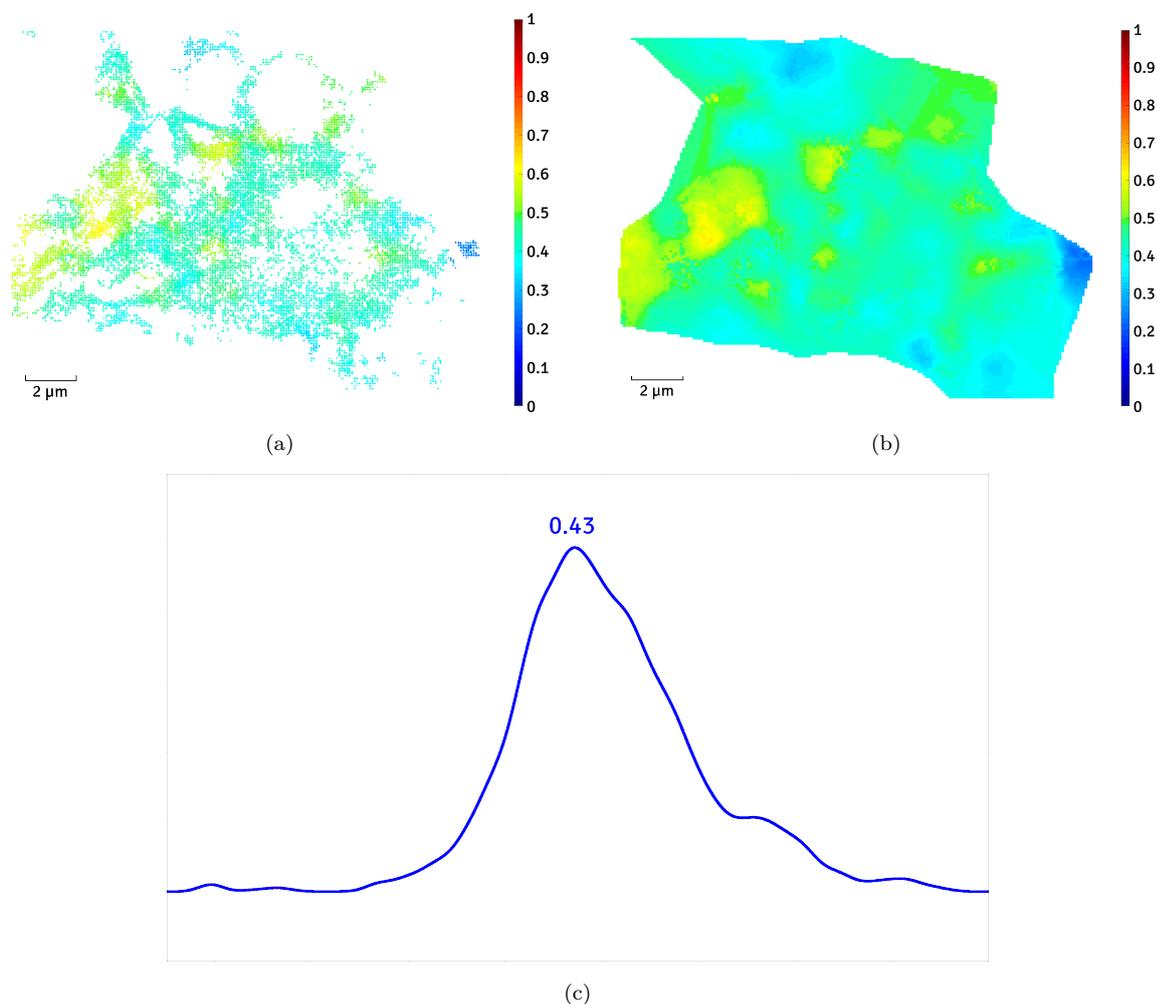


Figure 2.15 – The results obtained with our method on real data (mEOS2-calnexin plasmids in HEK-293 cells), are shown. **(a)** The point cloud output represents the spatial distribution of the diffusion coefficient inside the cell, in  $\mu\text{m}^2/\text{s}$ . **(b)** The point cloud is derived into a map by interpolation, with the same scale units. **(c)** The point cloud is derived into a diffusion histogram. The values displayed on top of the modes are also in  $\mu\text{m}^2/\text{s}$

cation, in Figure 2.21 for the trajectory lengths histogram, in Figure 2.22 for the diffusion and drift maps, and in Figure 2.23 for the diffusion histogram, respectively.

Figure 2.22(a) shows the diffusion point cloud. It is possible with MATLAB to spin and zoom around the results. Interpolation methods are not relevant in this case since they would hide the points inside the 3D volume. The drift vector field is shown in Figure 2.22(b) in 3D mode in the form of arrows (as opposed to color-coded points in 2D mode).

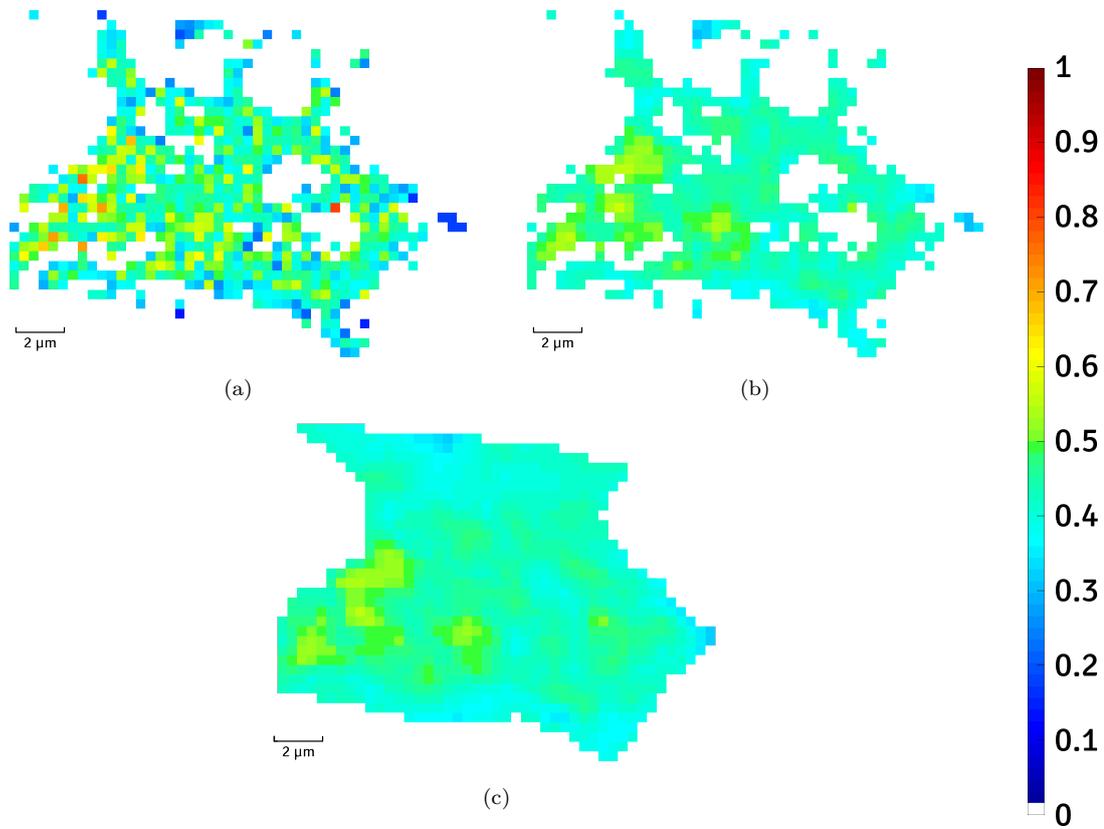


Figure 2.16 – The results obtained with method [Hozé, Nair, et al. 2012] on real data (mEOS2-calnexin plasmids in HEK-293 cells), are shown. Diffusion maps are generated using a resolution parameter of  $0.4\mu\text{m}$ , with from left to right: **(a)** the raw map, **(b)** the Gaussian-filtered ( $\sigma = 0.8$ ) map, **(c)** the Gaussian-filtered and interpolated diffusion map. The diffusion scale units are in  $\mu\text{m}^2/\text{s}$

The data set boundary is represented here as a grey 3D surface. In this case, it allows the shape of the cell to be displayed, making it easy to locate the drift vectors. Finally, Figure 2.23 displays the diffusion histogram corresponding to the point cloud.

The other two methods which with we compared our results on the different 2D data set do not allow for a 3D analysis. This functionality is a novelty in the domain.

### Analysis of 3D dynamics of CD-mannose-6 phosphate receptor in HeLa cell

Our method was tested by a third-party researcher (Cesar Valades-Cruz) on 3D real data depicting CD-Mannose-6 Phosphate (eGFP-CD-M6PR) receptors in a HeLa cell. Images were acquired using LLSM microscopy by the joint SERPICO/STED team from the Inria Center at the University of Rennes and the Curie Institute, and tracking was

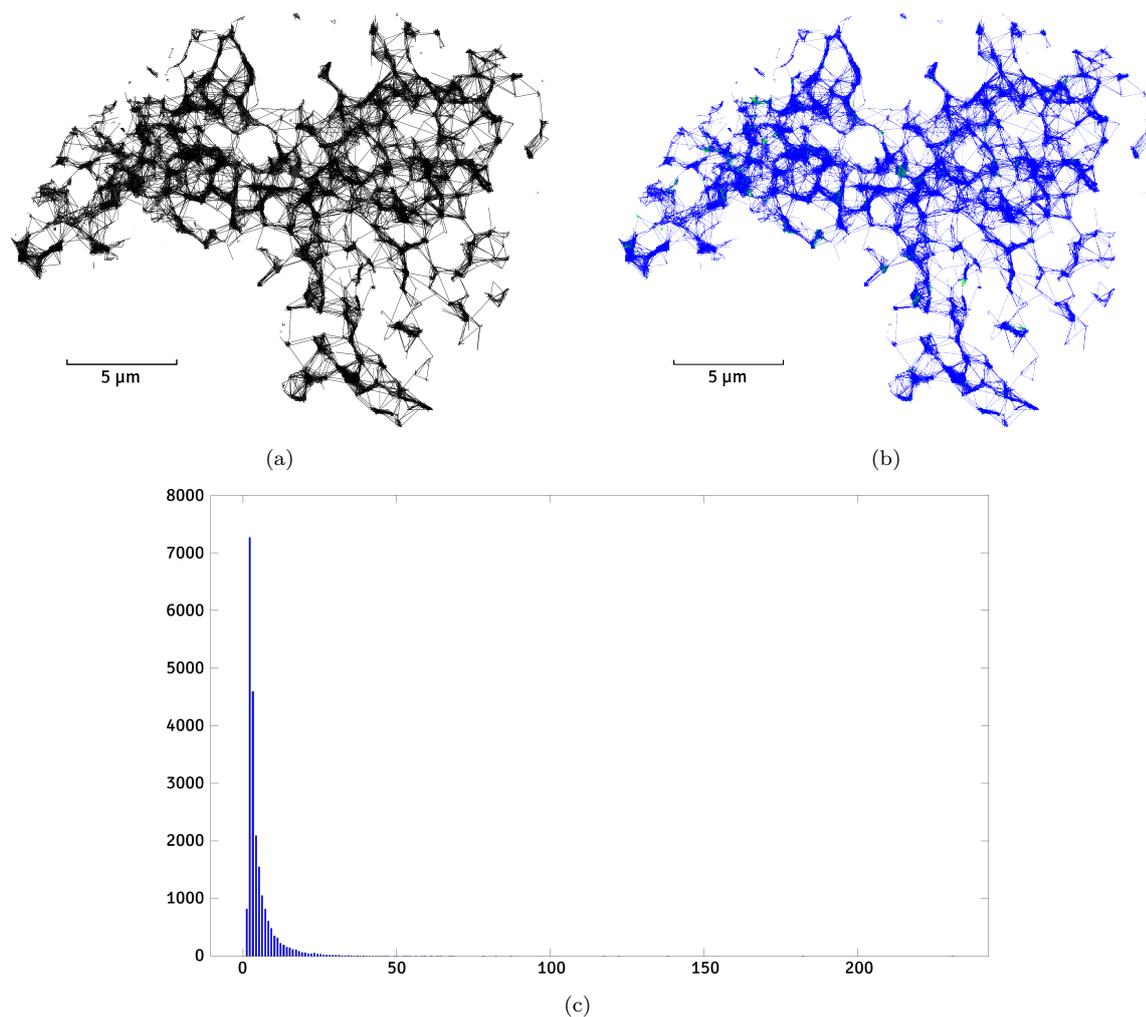


Figure 2.17 – The tracking data obtained from the TMR-labelled Halo-tagged CRT in HEK-293 cells data set is analyzed. **(a)** The set of all simulated particle trajectories is shown. **(b)** Particle trajectories are classified by motion type. **Free diffusion** is represented in **blue**, **subdiffusion** is represented in **green** and **superdiffusion** is represented in **red**. **(c)** The histogram of the input trajectory lengths is shown.

carried out using U-TRACK. Our method produced the outputs shown in Figure 2.24.

In Figure 2.24(a), the trajectories are represented inside a 3D view of deconvoluted fluorescence microscopy data. In Figure 2.24(b), the 3D point cloud output represents the spatial distribution of the diffusion coefficient inside the cell, in  $\mu\text{m}^2/\text{s}$ . In Figure 2.24(c), particle trajectories are classified by motion type. **Free diffusion** is represented in **blue**, **subdiffusion** is represented in **green** and **superdiffusion** is represented in **red**. Finally, the histogram of input trajectory lengths is shown in Figure 2.24(d).

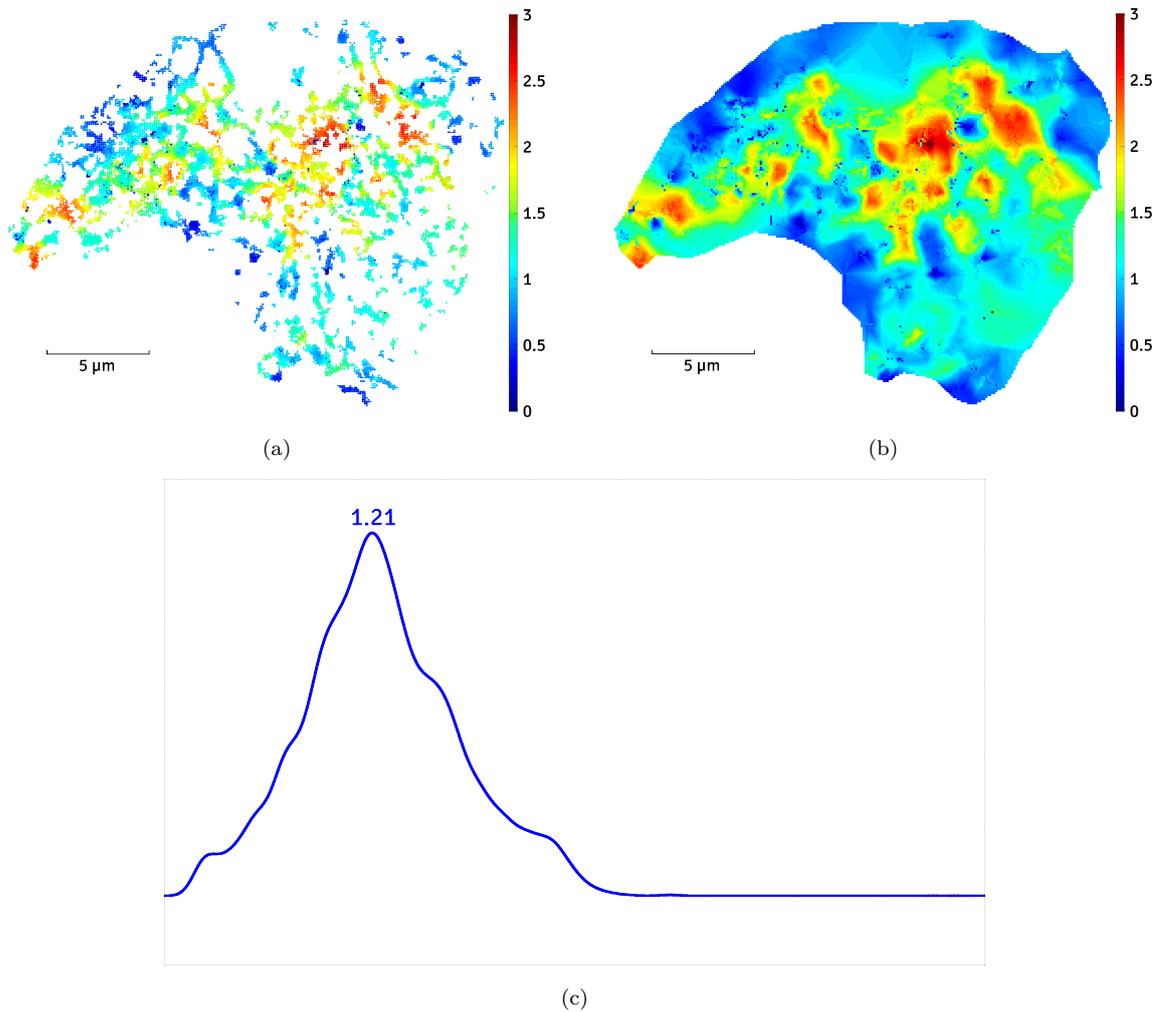


Figure 2.18 – The results obtained with our method on real data (TMR-labelled Halo-tagged CRT in HEK-293 cells), are shown. **(a)** The point cloud output represents the spatial distribution of the diffusion coefficient inside the cell, in  $\mu\text{m}^2/\text{s}$ . **(b)** The point cloud is derived into a map by interpolation, with the same scale units. **(c)** The point cloud is derived into a diffusion histogram. The values displayed on top of the modes are also in  $\mu\text{m}^2/\text{s}$

### Time-evolving analysis feature for diffusion and drift

Our method also allows for an time-evolving display of diffusion and drift, in the form of 2 videomaps outputs.

This feature allows for a biologist to have access to a time-evolving analysis of a given data set, instead of a single steady map in which time-evolving particle behavior could be hidden. In the parameters, the user can choose the number of frames and framerate of

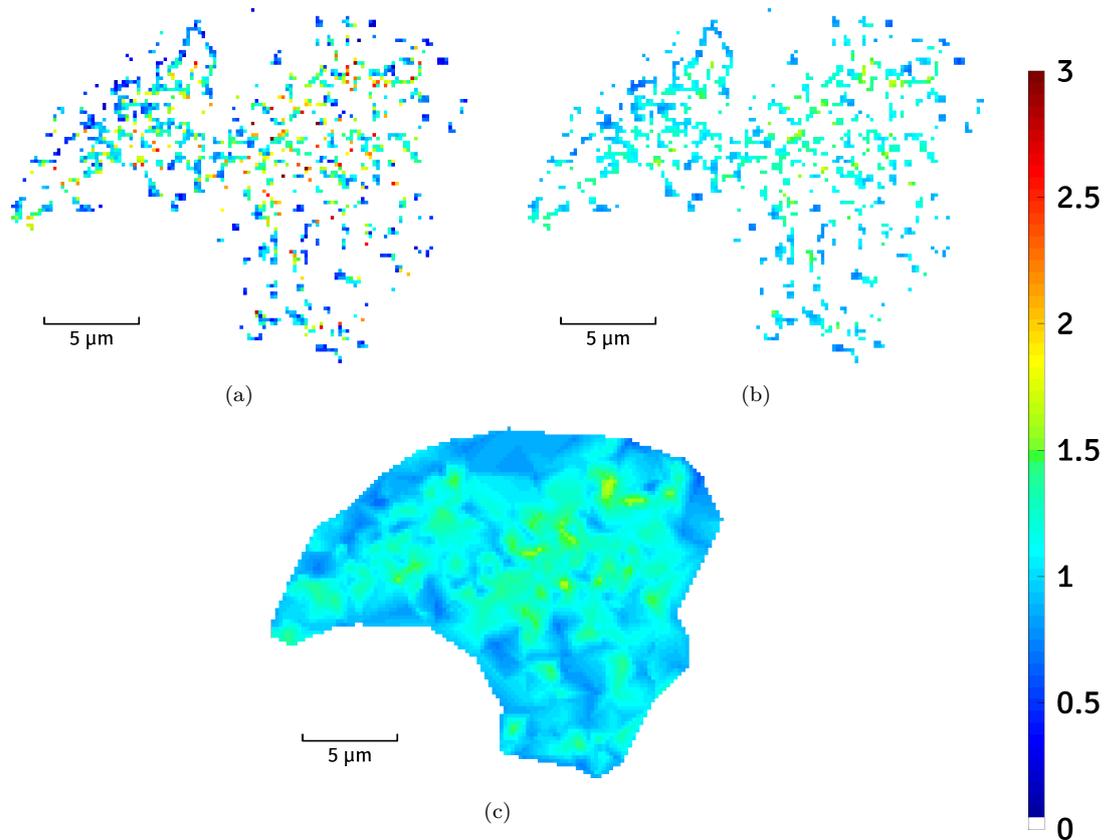


Figure 2.19 – The results obtained with method [Hozé, Nair, et al. 2012] on real data (TMR-labelled Halo-tagged CRT in HEK-2993 cells), are shown. Diffusion maps are generated using a resolution parameter of  $0.2\ \mu\text{m}$ , with from left to right: **(a)** the raw map, **(b)** the Gaussian-filtered ( $\sigma = 0.8$ ) map, **(c)** the Gaussian-filtered and interpolated diffusion map. The diffusion scale units are in  $\mu\text{m}^2/\text{s}$

the output video. In order to generate a smooth depiction of both quantities’ evolutions, our method allows to choose an temporal overlapping rate between each calculated frame, set between 0 and 1. In addition to the diffusion and drift videomaps, the algorithm also returns the time interval corresponding to each frame, in seconds.

In Figures 2.25 and 2.26, we used tracking data depicting the motion of HIV-1 mEOS2-tagged Gag proteins inside a CD4 T cell as an input to generate diffusion and drift videomaps. The data were originally used in [Floderer et al. 2018]. Acquisitions were made on a TIRF-PALM Nikon Ti microscope supplied by MRI and CEMIPAI platforms in Montpellier, France, and tracking was carried out with a plug-in from METAMORPH (Molecular Devices).

The chosen framerate of the output video is of 5 frames per second, with a total of 9

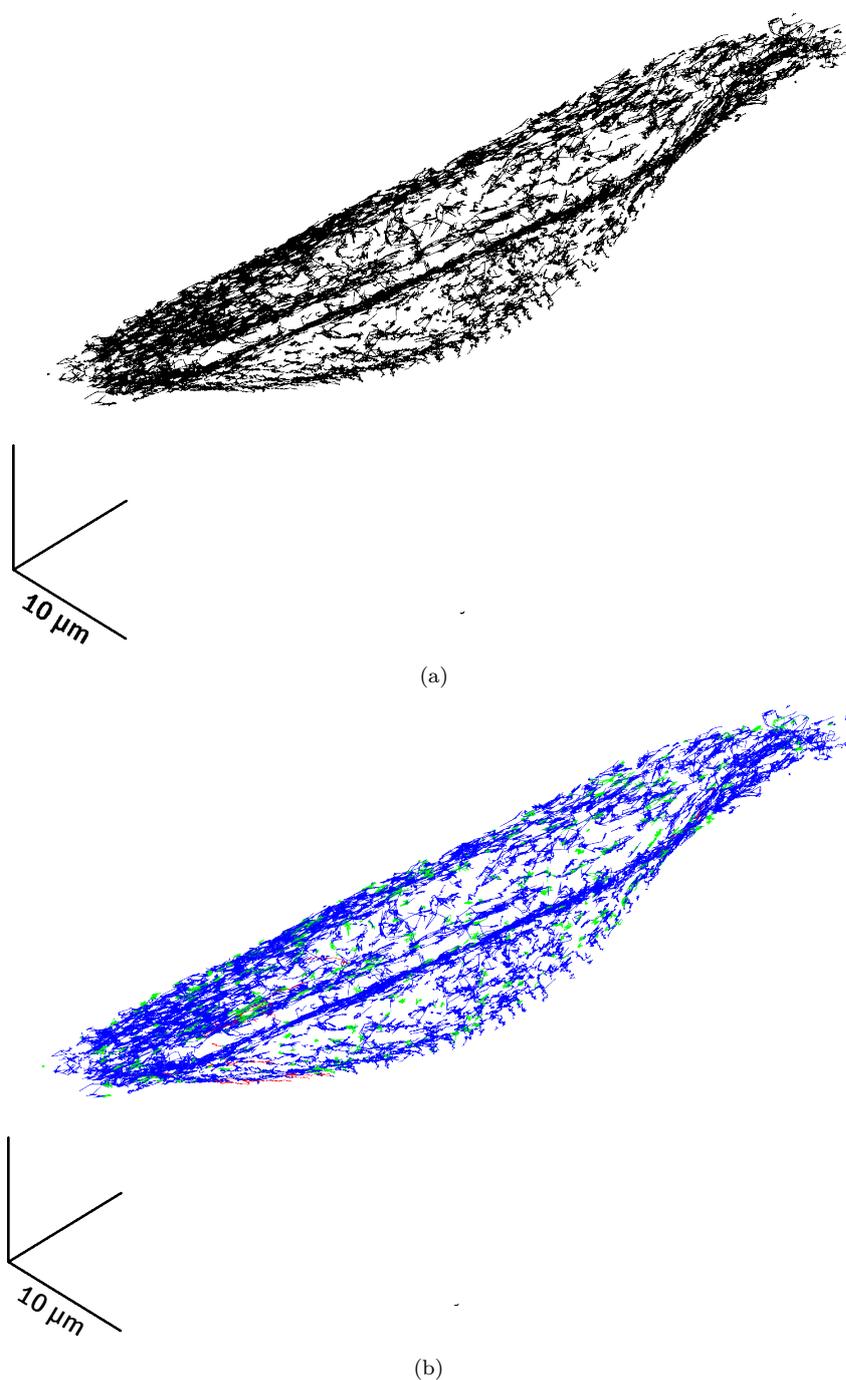


Figure 2.20 – The 3D tracking data obtained from the Galectin-3 in SUM159 cell data set is analyzed. **(a)** The set of all simulated particle trajectories is shown. **(b)** Particle trajectories are classified by motion type. Free diffusion is represented in blue, subdiffusion is represented in green and superdiffusion is represented in red.

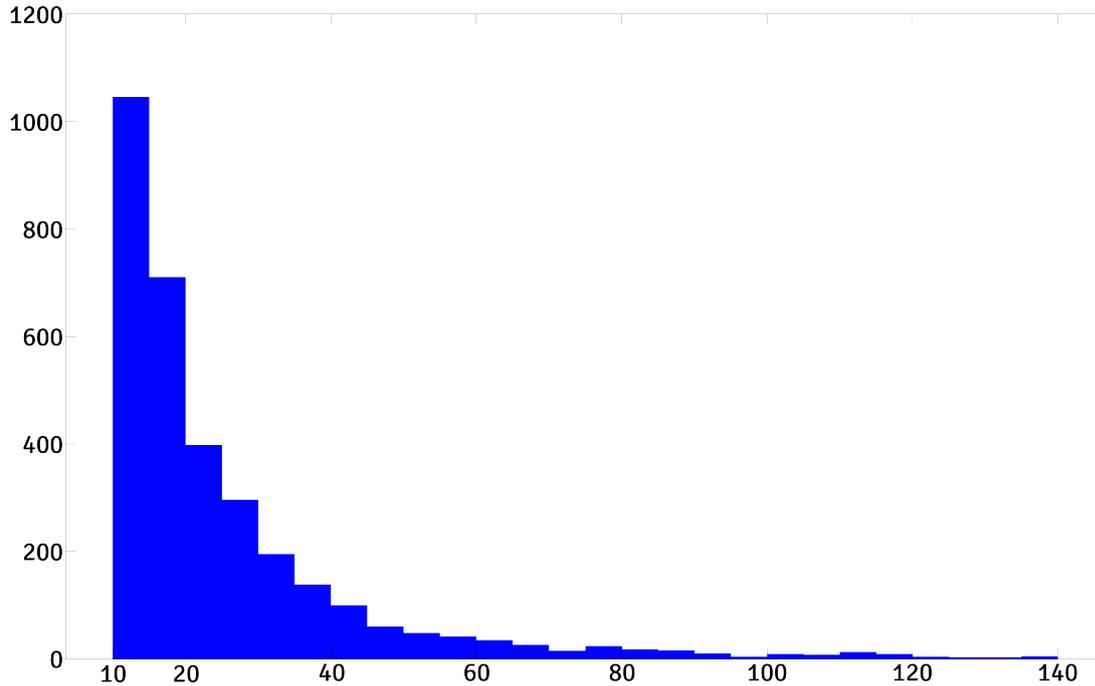


Figure 2.21 – The length histogram derived from the trajectories displayed in Figure 2.20(a) is shown.

frames and an overlapping rate of  $2/3$ . Each frame corresponds to a time interval of 21.89 s for a total of 80 s. This demonstrates that DenseMapping is versatile and can adapt to various biological studies, including spatiotemporal estimation (see another recent example in Appendix A.2).

## 2.4 Conclusion

We have proposed a new mapping method to robustly estimate the diffusion and drift maps from particle tracks. We first presented the theory behind the two state-of-the-art methods linked to this very specific purpose. Then we detailed the background behind our method, that uses a spatiotemporal kernel estimator guided by particle motion classification and following the particle tracks. The particle motion classification is performed through the previous works of Vincent Briane [Briane, Kervrann, and Vimond 2018] [Briane, Vimond, Valades-Cruz, et al. 2020]. We gave a quick peak at the computer implementation of our method via a compact and summarized pseudo-code of the method, which has been developed in MATLAB.

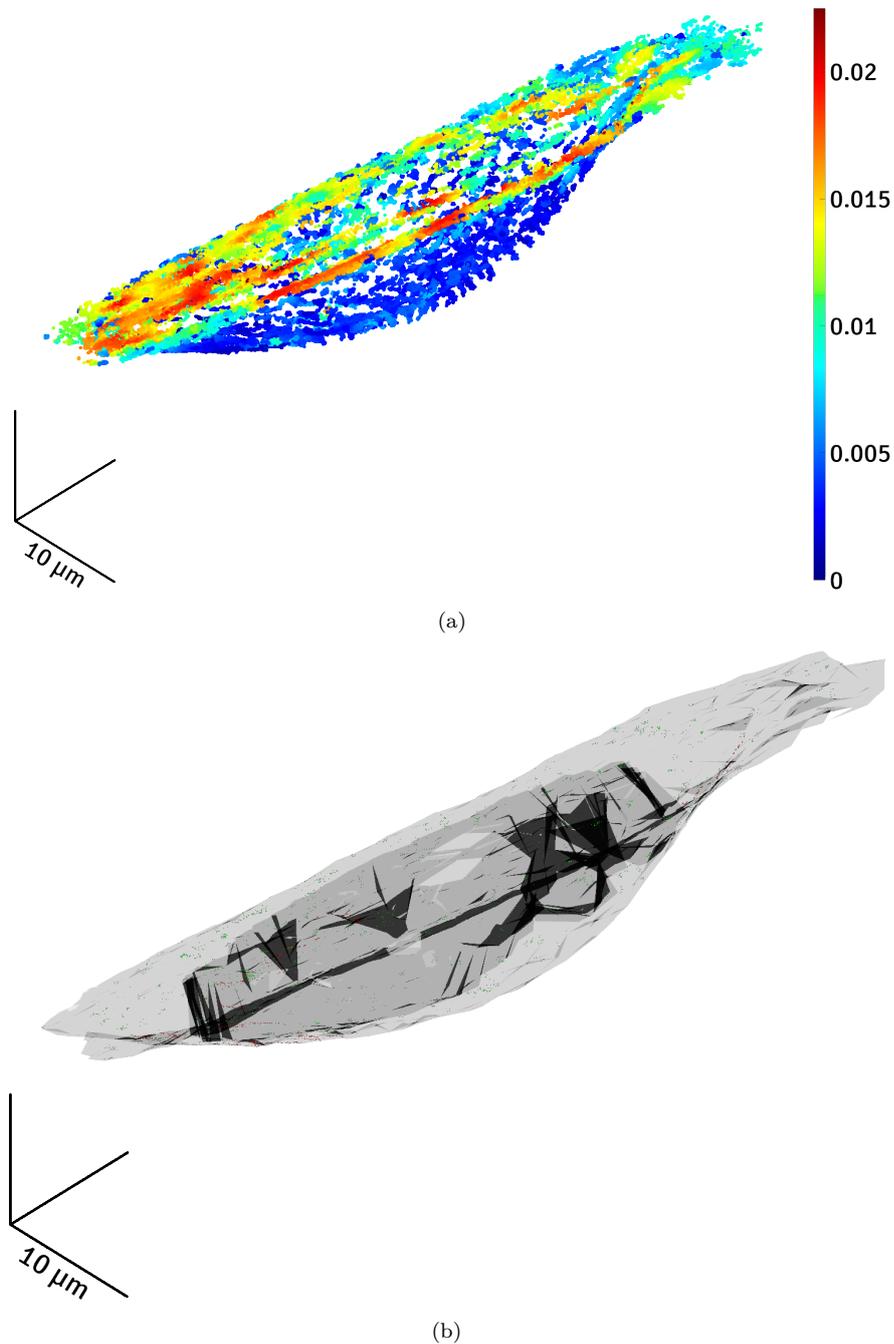


Figure 2.22 – The results obtained with our method on real 3D data (Galectin-3 in SUM159 cell), are shown. **(a)** The 3D point cloud output represents the spatial distribution of the diffusion coefficient inside the cell, in  $\mu\text{m}^2/\text{s}$ . **(b)** This 3D vector field represents the spatial distribution of the drift inside the cell. **Green vectors** correspond to **subdiffusion-related** drift whereas **red vectors** correspond to **superdiffusion-related** drift.

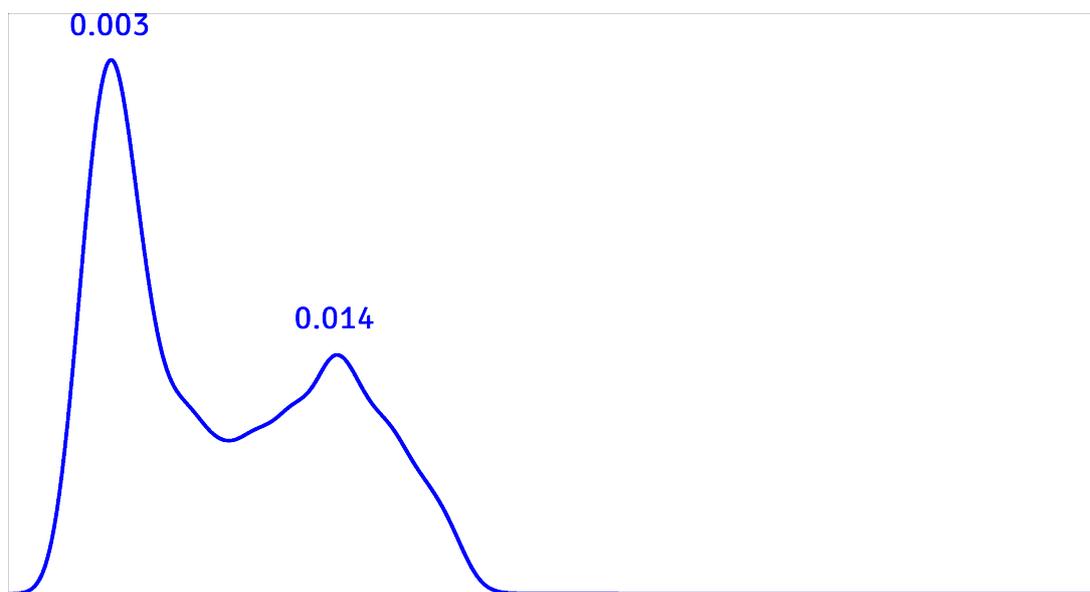
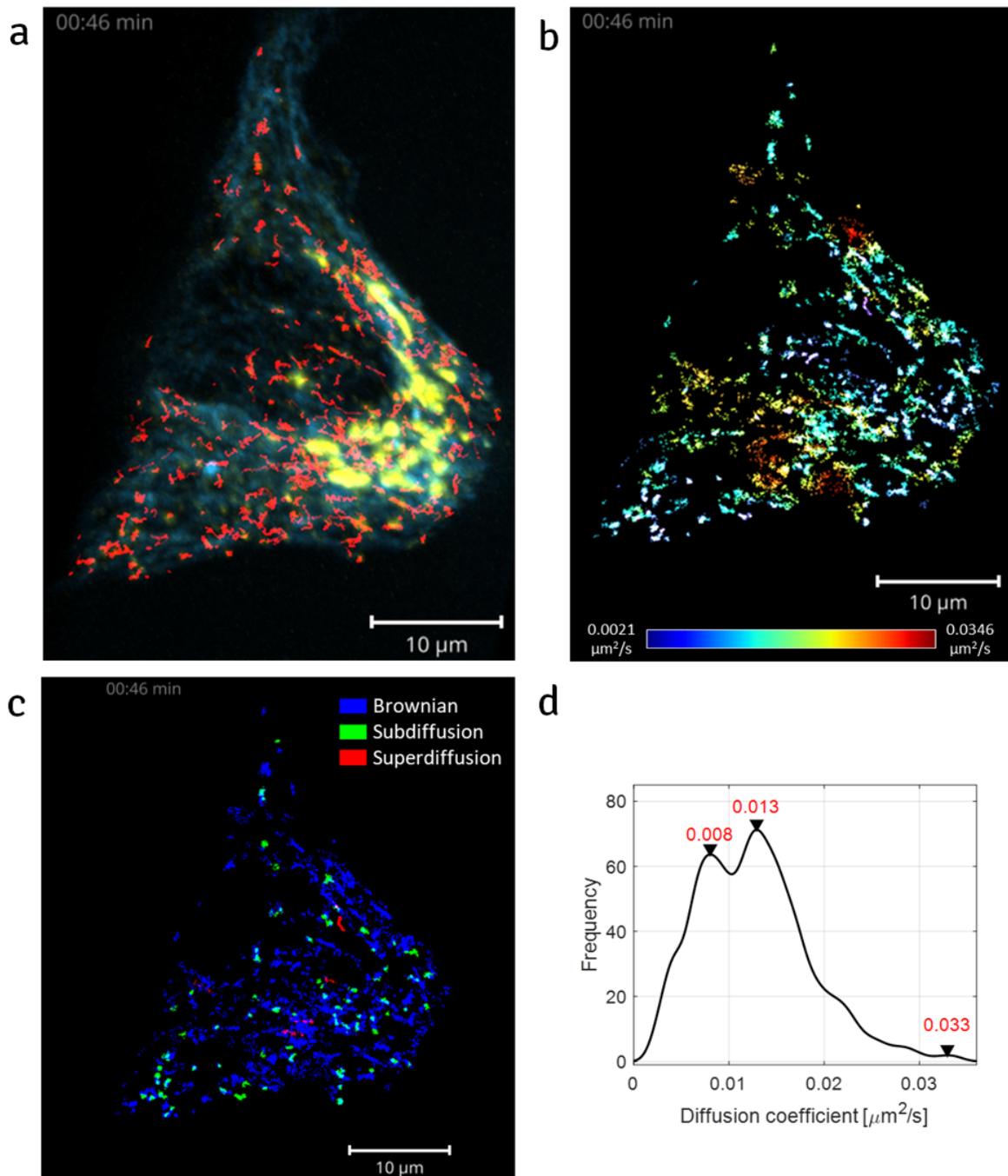


Figure 2.23 – The diffusion histogram derived from the point cloud displayed in Figure 2.22(a) is shown. The values displayed on top of the modes are in  $\mu\text{m}^2/\text{s}$ .

From an experimental point of view, our approach gives satisfactory results on both simulated and real data, providing high-resolution estimation maps even if the number of input data points is low. The method can handle both 2D and 3D tracking data without significant loss in computation time, thanks to the fact that the computation is point-based (instead of block-based). We showed that our method is not only effective by itself, but also surpasses state-of-the-art methods in the sense that it is more robust when facing a lack of data quantity. The obtained results are helpful to detect subtle biological events in image sequences.

In summary, the advantages of our method reside in its ability to adapt to the density of tracks and particle motion. It performs well in small or low-density data sets, both in 2D and 3D, and is able to do time-evolving analyses. This makes the method adapted to a large range of microscopy techniques, as well as diverse biological subjects of study such as viruses, molecular transport, transcription factors, etc. DenseMapping was transferred to Myriade company ([www.myriadelab.com](http://www.myriadelab.com)) to infer particle size distributions, and was applied to analyze the dynamics of extracellular vesicles, viruses and phages.



(a)

Figure 2.24 – The results obtained with our method on real 3D data (CD-Mannose-6 Phosphate (CD-M6PR) receptor in HeLa cell), are shown. (a) The trajectories are represented inside a 3D view of deconvoluted fluorescence microscopy data. (b) The 3D point cloud output represents the spatial distribution of the diffusion coefficient inside the cell, in  $\mu\text{m}^2/\text{s}$ . (c) Particle trajectories are classified by motion type. Free diffusion is represented in blue, subdiffusion is represented in green and superdiffusion is represented in red. (d) The histogram of the input trajectory lengths is shown.

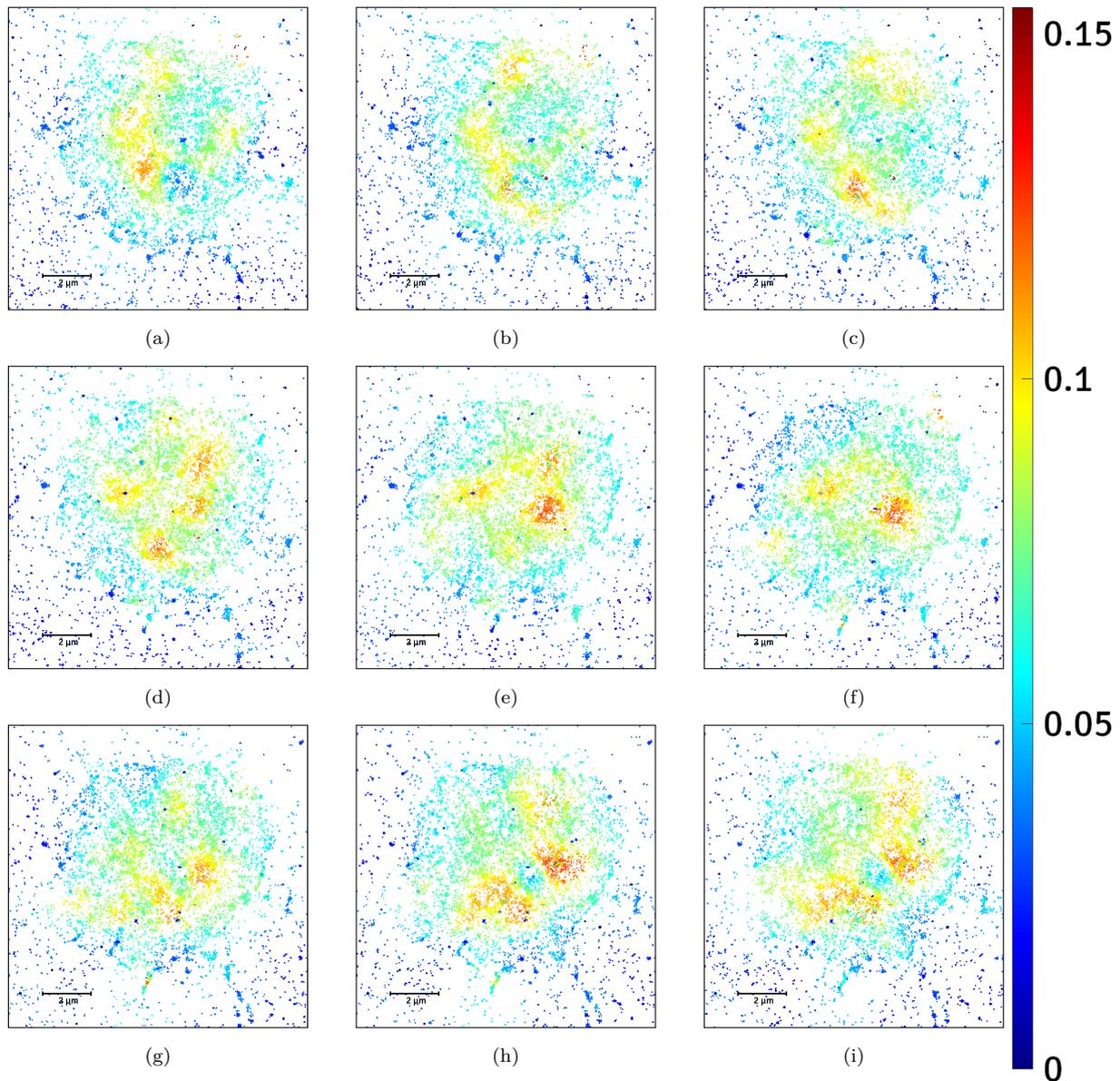


Figure 2.25 – The results obtained with our method on real data (HIV-1 Gag in CD4 T cell), are shown as an evolving diffusion videomap of 9 frames, with a time interval overlapping of 2/3 between each frame. (a) to (i) The point cloud output represents the evolution of the spatial distribution of the diffusion coefficient inside the cell, in  $\mu\text{m}^2/\text{s}$ , from  $t = 0\text{ s}$  to  $t = 80\text{ s}$ .

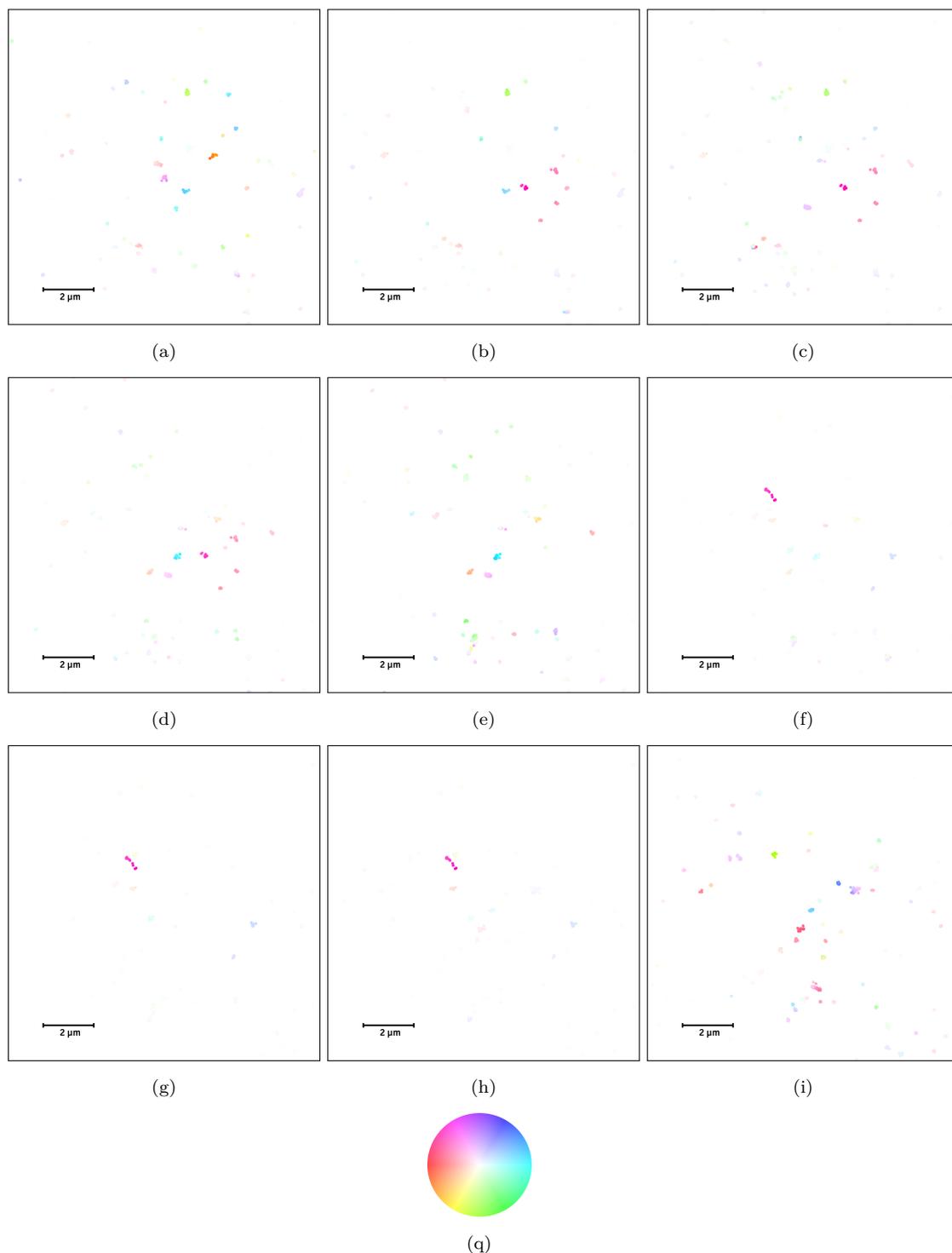


Figure 2.26 – The results obtained with our method on real data (HIV-1 Gag in CD4 T cell), are shown as an evolving drift videomap of 9 frames, with a time interval overlapping of  $2/3$  between each frame. **(a)** to **(i)** This second point cloud represents the evolution of the spatial distribution of the drift inside the cell, from  $t = 0$  s to  $t = 80$  s. Each point represents a drift vector, whose direction and amplitude are respectively represented by hue and saturation. **(q)** The hue-saturation colorwheel is used to interpret the color representation of drift vectors.



# MICRO-DOMAINS AND CONFINEMENTS

## DOMAINS DETECTION

---

This chapter was published in collaboration with V. Briane in *Physical Biology* as "*A computational approach for detecting micro-domains and confinement domains in cells: a simulation study*" [Briane, Salomon, et al. 2020].

### 3.1 Introduction

As the interior of a living cell is a fluctuating environment, the trajectories of intracellular particles can be modeled with stochastic processes having continuous paths. Diffusions belong to this class of processes and can model a large range of intracellular movements. They are widely used in the biophysical literature [H. Qian, Sheetz, and Elson 1991] [Saxton and Jacobson 1997]. Biophysicists distinguish four main types of diffusion, namely free diffusion (also referred to as Brownian motion), superdiffusion, confined diffusion and anomalous diffusion. These different diffusion types correspond to specific biological scenarios as briefly described in [Briane, Vimond, and Kervrann 2020]. A particle evolving freely inside the cytosol or along the plasma membrane is modeled by free diffusion. Its motion is due to constant collisions with smaller particles undergoing thermal fluctuations. Thus, the particle does not travel along any particular direction and can take a very long time to go to a precise area in the cell. Active intracellular transport suggests that particle motion is direct. The particles (called in this context cargos) are carried by molecular motors along microtubules and actin filament networks. Superdiffusion models the motion of molecular motors and their cargos. Confined or restricted diffusion [Metzler and Klafter 2000] [Hozé, Nair, et al. 2012] is characteristic of trapped particles: the particle encounters a binding site, then it pauses for a while before dissociating and moving away. Anomalous diffusion includes particles that encounters dynamics or fixed obstacles [Saxton 1994] [Berry and Chaté 2014], or particles slowed down by the contrary current

due to the viscoelastic properties of the cytoplasm. Very recently, [Smith, Cianci, and Grima 2017] studied the effect of heterogeneous crowding on the motion of intracellular particles. They assumed that the motion of the particle is hindered by spherical objects, and then derived an advection-diffusion equation to describe the motion of a small particle traveling across an arbitrary distribution of large crowder molecules. Finally, we emphasize that particles can switch from one motion type to another over time; for instance, [Fusco et al. 2003] showed experimentally that mRNA molecules undergo cycles between anchoring, free diffusion and superdiffusion (active transport).

An amount of work has been dedicated to the description of motions of individual particles in the cell. The second step consists in understanding the spatial distribution of these motions within the cell and interpreting the underlying biological mechanisms. As a pioneer work, [Hummer 2005] studied the spatial mapping of diffusion coefficients using Bayesian inference. This chapter inspired other mapping methods (e.g. see [Masson, Dionne, et al. 2014] and [Hozé, Nair, et al. 2012]). In this chapter, we are interested in localizing areas where subdiffusion occurs as in [Hozé, Nair, et al. 2012]. Regulation of cellular physiological processes relies on molecular interactions (binding and unbinding) at specific places and involves trafficking in confined local micro-domains. Our objective is to automatically detect these micro-domains from groups of particle trajectories. [Hozé, Nair, et al. 2012] model these micro-domains as potential wells that attract intracellular particles. The authors use an Eulerian method based on a non-parametric estimation of the drift parameter of the underlying diffusion process. This method assumes a high concentration of particles over the spatial domain of interest to be meaningful. Moreover, this Eulerian approach is not designed to handle a mixture of different diffusion types observed in a small region. All the trajectories in a small region are uniformly averaged, potentially leading to wrong conclusions. To overcome this problem, we propose an original procedure to detect micro-domains.

Our method is based on the combination of the clustering algorithm DBSCAN ([Ester et al. 1996]) and a three-label classification procedure ([Briane, Kervrann, and Vimond 2018]) which assigns a label (free diffusion, subdiffusion and superdiffusion) to each trajectory. Our results on simulations mainly demonstrate that:

1. our method detects satisfyingly the micro-domain shapes;
2. our method is able to handle a mixture of several diffusion types.

The remainder of this chapter is organized as follows. In Section 3.2 we present the mathematical models and the notations used in the chapter. In Section 3.3, we present

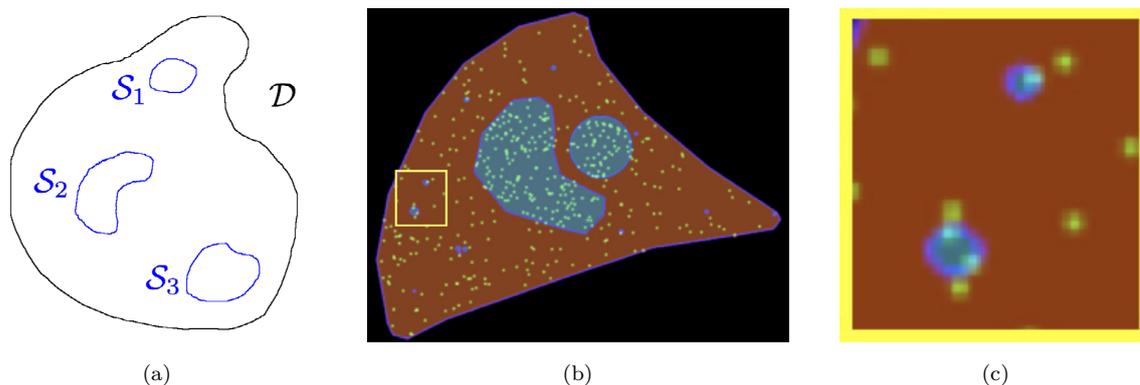


Figure 3.1 – Illustrations of configurations with trapping areas. (a) Sketch with three trapping areas ( $\mathcal{S} = \mathcal{S}_1 \cup \mathcal{S}_2 \cup \mathcal{S}_3$ ); (b) Example of a simulated image at a given time  $t$ , depicting a crossbow-shaped cell and green particles in small trapping (blue) areas and in the cytosol  $\bar{\mathcal{S}}$  (orange). (c) ROI of (b)

related works and focus on the Eulerian method proposed by [Hozé, Nair, et al. 2012]; we emphasize the underlying assumptions, the key parameters, and the properties of this approach used to detect potential wells within cells. In Section 3.4, we describe the simulation framework corresponding to the FLUOSIM simulator ([Lagardère et al. 2020]) used in our experiments. In Section 3.5, we describe the procedure that combines the DBSCAN clustering algorithm ([Ester et al. 1996]) with the test procedure described in [Briane, Kervrann, and Vimond 2018]. In Section 3.6, we evaluate the performance of the computational method on simulated sequences, and we illustrate its potential on real TIRF microscopy image data depicting small zones with a high concentration of subdiffusive vesicles. In Section 3.7, we discuss the main features of our approach.

## 3.2 Notations and mathematical framework

In the sequel, we assume a population of  $m$  independent trajectories, with  $m$  of order of magnitude of  $10^3$  particles. We denote  $\mathbf{x}_t^{(i)} \in \mathcal{D}$  the position of the  $i$ th particle at time  $t$ , where  $\mathcal{D} \subset \mathbb{R}^2$  denotes a spatial bounded domain. Let us define  $P$  subregions  $\mathcal{S}_1, \dots, \mathcal{S}_p$  such that  $\mathcal{S}_p \subset \mathcal{D}$  in which the particles can be trapped (see Figure 3.1(a)). The union of subdomains is denoted  $\mathcal{S} = \cup_{p=1}^p \mathcal{S}_p$ , and  $\bar{\mathcal{S}} = \mathcal{D} \setminus \mathcal{S}$  is defined as the complementary domain where the particles undergo free diffusion with diffusion coefficient  $\sigma_{\bar{\mathcal{S}}}$ . We impose that the particles are normally reflected at the boundaries of  $\mathcal{D} = \mathcal{S} \cup \bar{\mathcal{S}}$ .

Inside a subregion  $\mathcal{S}_p \subset \mathcal{D}$ , a particle can switch between two states: trapped or

non-trapped (also denoted as free):

1. the particle is trapped in  $\mathcal{S}_p$ : the particle motion is a confined free diffusion normally reflected at the boundary  $\partial\mathcal{S}_p$  of  $\mathcal{S}_p$ , with a diffusion coefficient  $\sigma_T \leq \sigma_{\bar{\mathcal{S}}}$ ;
2. the particle is not trapped in  $\mathcal{S}_p$ : the particle motion is a free diffusion with diffusion coefficient  $\sigma_S$ .

Formally, we introduce the following indicator variable:

$$\phi(\mathbf{x}_t^{(i)}) = \begin{cases} 1 & \text{if the particle } i \text{ is trapped at time } t, \\ 0 & \text{if the particle } i \text{ is not trapped at time } t. \end{cases}$$

Given the binding rate  $k^+$  and the unbinding rate  $k^-$ , the probabilities of a particle belonging to  $\mathcal{S}$  to be trapped are defined as follows:

$$\begin{aligned} P(\phi(\mathbf{x}_{t+h}) = 1 \mid \phi(\mathbf{x}_t) = 0, \mathbf{x}_{t+h} \in \mathcal{S}) &= k^+h + o(h), \\ P(\phi(\mathbf{x}_{t+h}) = 0 \mid \phi(\mathbf{x}_t) = 1, \mathbf{x}_{t+h} \in \mathcal{S}) &= k^-h + o(h), \end{aligned}$$

where  $h > 0$  and  $o(h)/h \rightarrow 0$  as  $h \rightarrow 0$ . In the sequel, the binding and unbinding rates  $k^+$  and  $k^-$  are two constants which do not depend on space. Hence, a trapped particle close to the boundaries  $\partial\mathcal{S}_p$  has the same probability to unbind to the trapping region  $\mathcal{S}_p$  than a trapped particle in the middle of the region  $\mathcal{S}_p$ . For the sake of simplicity, we assume that all the subregions  $\mathcal{S}_p$  share the same common parameters  $k^+$ ,  $k^-$ ,  $\sigma_T$  and  $\sigma_S$ .

In our estimation framework, we observe a collection of  $m$  two-dimensional trajectories  $\mathcal{X}_m$ . We denote  $\mathbb{X}_{n_i}^{(i)}$  the vector of  $n_i$  spatial positions along the trajectory associated to the  $i^{\text{th}}$  particle:

$$\begin{aligned} \mathbb{X}_{n_i}^{(i)} &= (\mathbf{x}_{t_0}^{(i)}, \dots, \mathbf{x}_{t_{n_i}}^{(i)}), \quad i = 1, \dots, m, \\ \mathcal{X}_m &= \{\mathbb{X}_{n_i}^{(i)}, i = 1, \dots, m\}. \end{aligned}$$

We assume that each discrete trajectory is generated by a stochastic process  $(\mathbf{x}_t^{(i)})$  with continuous path defined on the spatial domain  $\mathcal{D} \subset \mathbb{R}^2$  and which is a solution of the stochastic differential equation (SDE)

$$d\mathbf{x}_t^{(i)} = \boldsymbol{\mu}(\mathbf{x}_t^{(i)}) dt + \sigma^{(i)} d\mathbf{b}_t^{(i), h^{(i)}}, \quad t \in [t_0, t_{n_i}],$$

where  $\mathbf{b}_t^{(i), h^{(i)}}$  is a two-dimensional fractional Brownian motion of unknown Hurst param-

eter  $\mathfrak{h}^{(i)} \in (0, 1)$ ,  $\sigma^{(i)} > 0$  is the diffusion coefficient, and  $\boldsymbol{\mu}^{(i)} : D \rightarrow \mathbb{R}^2$  is the drift term. The different trajectories do not need to have the same drift parameter and the same diffusion coefficients, provided that the drift functions and diffusion coefficients do not depend on time.

### 3.3 Most related work

In this section, we focus on the two most related methods developed to detect 2D microdomains. First, we describe the Eulerian method proposed by [Hozé, Nair, et al. 2012], based on the two following steps:

1. non-parametric estimation of the drift vector field from multiple trajectories;
2. fitting to a parametric drift function.

As starting point, [Hozé, Nair, et al. 2012] model microdomains as potential wells that attract intracellular particles. A high density of particles within a local neighborhood is assumed. Formally, let us consider a collection of  $m$  independent trajectories  $\mathcal{X}_m$  such as all the trajectories  $\mathbb{X}_{n_i}^{(i)} = (\mathbf{x}_{t_0}^{(i)}, \dots, \mathbf{x}_{t_{n_i}}^{(i)})$  are generated from the common diffusion process  $(\mathbf{x}_t)$ , solution of the overdamped Langevin equation defined as

$$d\mathbf{x}_t = -\nabla U(\mathbf{x}_t) dt + \sigma d\mathbf{b}_t,$$

where  $\nabla$  is the gradient operator,  $U : \mathbb{R}^2 \rightarrow \mathbb{R}$  is the potential function, and  $\sigma > 0$  the diffusion coefficient. In the sequel, the two-dimensional drift function (or vector field) is defined as  $\boldsymbol{\mu}(\mathbf{x}) = -\nabla U(\mathbf{x})$ . The drift function is defined through the limit as

$$\boldsymbol{\mu}(\mathbf{x}) = \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} E(\mathbf{x}_{t+\Delta t} - \mathbf{x}_t \mid \mathbf{x}_t = \mathbf{x}). \quad (3.1)$$

In [Hozé, Nair, et al. 2012], an empirical estimator  $\hat{\boldsymbol{\mu}}(\mathbf{x})$  of the drift is computed over a square  $r \times r$  window  $W_r(\mathbf{x})$  centered at point  $\mathbf{x} \in \mathcal{D} \subset \mathbb{R}^2$ , in order to reflect the local conditioning  $\mathbf{x}_t = \mathbf{x}$  in equation (3.1). The estimation of  $\hat{\boldsymbol{\mu}}(\mathbf{x})$  amounts to computing the average of all displacements of particles passing through the window  $W_r(\mathbf{x})$ . The estimation of  $\hat{\boldsymbol{\mu}}(\mathbf{x})$  is spatially varying and its accuracy depends on the number of trajectory points in  $W_r(\mathbf{x})$ . Fortunately, we generally observe a high density of particles in trapping areas. In these microdomains, the drift vector field is thus reliably estimated but it is assumed that, locally in the  $r \times r$  neighborhood, all the trajectories undergo the same

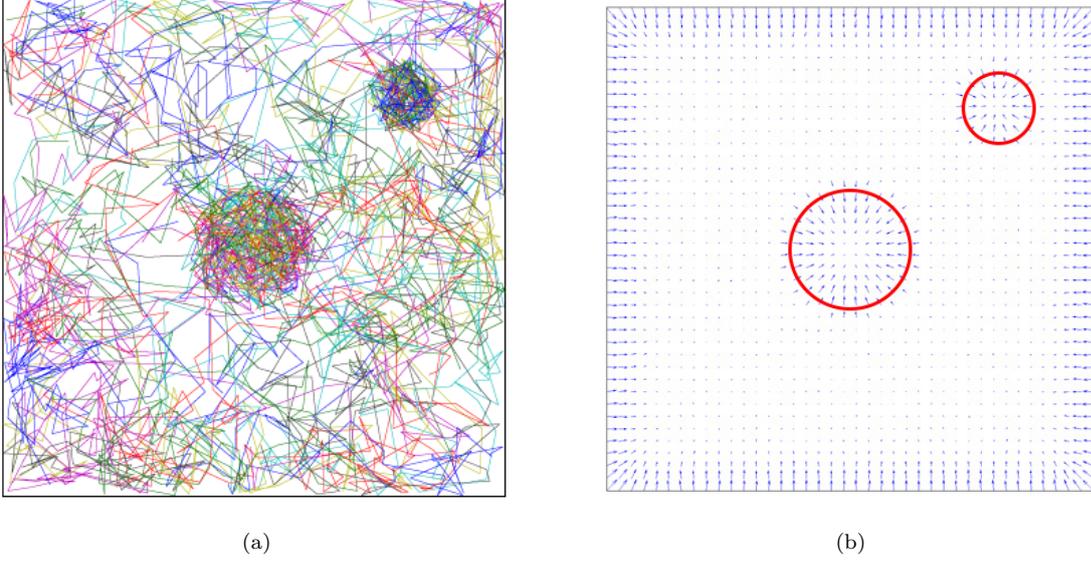


Figure 3.2 – Drift field computed with the method of [Hozé, Nair, et al. 2012]. The FLUOSIM simulation comprises 30 frames of  $200 \times 200$  px as the domain  $\mathcal{D}$  is a square of  $5 \mu\text{m}$  and the spatial resolution is  $\Delta x = 0.025 \mu\text{m}$  (see table 3.1). We compute  $\hat{\boldsymbol{\mu}}(x)$  on each point of this  $200 \times 200$  lattice from trajectories (a) ( $r = 5$  px). We can clearly see the two microdomains  $\mathcal{S}_1$  and  $\mathcal{S}_2$  (b): the vector field converges to the centers of these domains delineated by two red circles (b). We can see that the vector field is orthogonal to the boundaries of the square  $\mathcal{D}$ . It is due the normal boundary condition on  $\partial\mathcal{D}$ . Elsewhere, the  $\hat{\boldsymbol{\mu}}(x)$  is almost null, reflecting Brownian motion.

diffusion process with the same drift function. The parameter  $r$  is a free parameter of the algorithm, which can be interpreted as a bandwidth parameter in density estimation. In Figure 3.2, we illustrate the estimation of drift from trajectories on a artificial sequence depicting two circular trapping areas. A general view of trapping areas with variable sizes generated with FLUOSIM [Lagardère et al. 2020], is shown in Figure 3.1(b)).

In the second step, the authors assume that the potential  $U(\boldsymbol{x})$  has a parametric form and can be approximated by a truncated polynomial of second order, defined as

$$U(\boldsymbol{x}) = \begin{cases} A \left( \|\boldsymbol{x} - \boldsymbol{x}_0\|_{a,b}^2 - 1 \right) & \text{if } \|\boldsymbol{x} - \boldsymbol{x}_0\|_{a,b}^2 < 1, \\ 0 & \text{otherwise,} \end{cases}$$

where  $\boldsymbol{x}_0$  is the attractor, and  $A > 0$  is the depth of the potential modeling the strength of the attractive force toward  $\boldsymbol{x}_0$ , restricted to the underlying 2D ellipse. The parameters  $(a, b)$  are the axis lengths of the ellipse of center  $\boldsymbol{x}_0$  in which the particle are under the

action of an attractive force. In the computational scheme, the parameters  $a, b, \mathbf{x}_0$  and  $A$  are assumed partially known and/or estimated by minimizing a least-square error. To conclude, the approach of [Hozé, Nair, et al. 2012] is very appropriate to analyze a large number of tracks as computed in single-particle tracking PALM imaging. The motivation was to compute the residence time of receptors in dendritic spines using short receptor trajectories. The main advantage is the processing of very short tracks. Nevertheless, it cannot handle mixtures or changes of motion occurring in a local neighborhood.

Another Eulerian method was described in [Masson, Dionne, et al. 2014]. The authors propose a Bayesian framework to infer the drift vector field and diffusion coefficients. The idea is first to decompose the image into non overlapping blocks. All the sub-trajectories inside a block (or mesh element) are supposed to be driven by the same stochastic differential equations (SDE). Instead of using non-parametric estimates of drift and diffusion coefficients as in [Hozé, Nair, et al. 2012], the authors consider a Gaussian approximation of the SDE to model the likelihood function. Meanwhile, a Jeffreys prior is used as a default prior distribution for the drift and the diffusion coefficient, supposed to be constant in each block. It follows that the a posteriori distribution on the whole space is defined as the product of the a posteriori distributions of each block. An alternative prior is proposed to penalize gradients of the drift field (or the coefficient diffusion field) ([Beheiry, Türkcan, et al. 2016]); in that case, the maximum *a posteriori* (MAP) estimator is more computationally demanding. Finally, once the local drifts (and diffusion coefficients) are estimated, [Masson, Dionne, et al. 2014] propose to fit the estimated vector field to a parametric vector field as in [Hozé, Nair, et al. 2012]. They consider a penalized least square estimator of the vector field to impose a smooth solution. The main advantage of the Bayesian method is that prior information about the drift and diffusion coefficient can be easily incorporated, including errors due to misconnections of particles during tracking.

In this chapter, we are interested in the variation over space of particle motion. More specifically, different areas within the cell induce different dynamics. We assume that we can describe these dynamics with a finite collection of motions. We must note that there also exists a vast literature analyzing the switch of motion over time. Again, the motions belong to a finite collection of transport states. Then, [Monnier, Barry, et al. 2015] propose a Bayesian model selection to infer the transient transport states along a trajectory over time. The switches between the different transient states are modeled with a Markov process. In their case, the different states correspond to Brownian motion with different diffusion coefficients and Brownian motion with drift with different values for

both the drift and the diffusion coefficients.

### 3.4 Modeling and simulation of trapped particles

In this section, we propose a simulation framework based on general concepts in biophysics to mimic trapping areas. The artificial data sets will serve to evaluate the performance of detection and estimation methods described in Section 3.5. As said earlier in Section 3.1, we assume that the particles can switch between the trapped and the free states. When the particles are trapped, they undergo free diffusion in confined regions. When the particles are free, they are driven by free diffusion, only constrained by the boundary conditions on the limits of the domain  $\mathcal{D}$ . We model the proportions of trapped and free particles in the whole domain and in the trapping regions through two systems of differential equations.

#### 3.4.1 Modeling of particle dynamics

In the aforementioned situation, the dynamics of trapped and free particles are described by two connected processes:

1. the dynamics of individual particles corresponding to normally reflected free diffusion in  $\mathcal{D}$  and in  $\mathcal{S}$  when trapped;
2. the trapping process.

The trapping process induces a different motion for the trapped particles; inversely the particles can be trapped only if their motion drive them inside  $\mathcal{S}$ .

Formally, we assume  $m$  particles independently drawn from the uniform distribution over  $\mathcal{D}$  :  $\mathbf{x}_0^{(i)} \sim \mathcal{U}(\mathcal{D})$ ,  $i = 1, \dots, m$ . Note that  $n_i = n$  if the trajectory length is the same for all the particles. Instead of considering the exact dynamics of a particle, we simply model its probability to be in any trapping region  $\mathcal{S}_p$  (equivalently to be in  $\mathcal{S} = \cup_{p=1}^p \mathcal{S}_p$ ) by

$$p_{\mathcal{S}} := P(\mathbf{x}_t^{(i)} \in \mathcal{S}) = \frac{|\mathcal{S}|}{|\mathcal{D}|}, \quad i = 1, \dots, m.$$

Consequently, it is assumed that at every time  $t$ , the spatial point process  $(\mathbf{x}_t^{(1)}, \dots, \mathbf{x}_t^{(m)})$  is a binomial process over  $\mathcal{D}$  of parameter  $m$  and which density function is the uniform density ([Baddeley, Bárány, and Schneider 2007]). Furthermore, when the number  $m$  of

particles is large enough, the evolution of the two populations can be modeled by a system of differential equations depending on parameters  $p_S$ ,  $k^+$  and  $k^-$ .

Let us denote  $b(t) = P(\phi(\mathbf{x}_t) = 1)$  and  $f(t) = P(\phi(\mathbf{x}_t) = 0)$  the proportions of trapped particles and free particles, respectively. The functions  $t \rightarrow b(t)$  and  $t \rightarrow f(t)$  are solutions of

$$\begin{cases} \frac{db}{dt} = \gamma_1 f(t) - \gamma_2 b(t), \\ b(t) + f(t) = 1, \end{cases} \quad (3.2)$$

where  $\gamma_1$  and  $\gamma_2$  are, respectively, the global binding rate and global unbinding rate, such that

$$\begin{aligned} P(\phi(\mathbf{x}_{t+h}) = 1 \mid \phi(\mathbf{x}_t) = 0) &= \gamma_1 h + o(h), \\ P(\phi(\mathbf{x}_{t+h}) = 0 \mid \phi(\mathbf{x}_t) = 1) &= \gamma_2 h + o(h), \end{aligned}$$

with  $h > 0$  and  $h \rightarrow 0$ . Consequently,  $(\phi(\mathbf{x}_t))$  is a continuous-time homogeneous Markov chain with states 0, 1 with infinitesimal generator parameters  $k^+$  and  $k^-$  ([Brémaud 2020]). Finally, we can show that (see Appendix A.1)

$$\gamma_1 = k^+ p_S, \quad (3.3)$$

$$\gamma_2 = k^-. \quad (3.4)$$

The solution of the system (3.2) is as follows:

$$\begin{cases} b(t) = \left( b(0) - \frac{k^+ p_S}{k^+ p_S + k^-} \right) e^{-(k^+ p_S + k^-)t} + \frac{k^+ p_S}{k^+ p_S + k^-}, \\ b(t) + f(t) = 1. \end{cases} \quad (3.5)$$

As a uniform distribution is assumed, we have the constraint that  $b(0) \leq p_S$ . In other words, only particles inside  $\mathcal{S}$  can be trapped. The situation  $b(0) = p_S$  matches with the situation where initially all the particles inside  $\mathcal{S}$  are trapped. From equation (3.5), the transitory regime is exponential and it converges to an equilibrium proportion:

$$\lim_{t \rightarrow \infty} b(t) = \frac{k^+ p_S}{k^+ p_S + k^-}.$$

If  $b(0) < k^+ p_S / (k^+ p_S + k^-)$ , function  $b(t)$  is decreasing toward its equilibrium point  $k^+ p_S / (k^+ p_S + k^-)$ . If  $b(0) > k^+ p_S / (k^+ p_S + k^-)$ , function  $b(t)$  is increasing toward its

equilibrium point. If  $b(0) = k^+p_{\mathcal{S}}/(k^+p_{\mathcal{S}} + k^-)$ , function  $b(t)$  is constant and equal to its equilibrium point. The characteristic time is defined as  $\tau_c = 1/(k^+p_{\mathcal{S}} + k^-)$ . We can consider that the stationary regime is reached when  $t > 5\tau_c$ .

In the same way, we can model the proportions of trapped particles and free particles inside  $\mathcal{S}$ . If  $t \rightarrow b_{\mathcal{S}}(t)$  denotes the proportion of trapped particles inside  $\mathcal{S}$  and  $t \rightarrow f_{\mathcal{S}}(t)$  denotes the proportion of free particles inside  $\mathcal{S}$ , we have:

$$\begin{cases} \frac{db_{\mathcal{S}}}{dt} = k^+f_{\mathcal{S}}(t) - k^-b_{\mathcal{S}}(t), \\ b_{\mathcal{S}}(t) + f_{\mathcal{S}}(t) = 1. \end{cases} \quad (3.6)$$

As before, the solution has an exponential form. The equilibrium proportion of trapped particles inside any trapping regions  $\mathcal{S}_p$  (same for all regions) is as follows:

$$\lim_{t \rightarrow \infty} b_{\mathcal{S}}(t) = \frac{k^+}{k^+ + k^-}. \quad (3.7)$$

Below, we describe the simulation of microdomains generated by running the FLUOSIM software [Lagardère et al. 2020] based on the set of aforementioned equations. First, we evaluate the model (3.5) describing the evolution of the proportion of trapped and free particles in the whole domain  $\mathcal{S}$ . Second, we assess the model (3.6) describing the evolution of the proportion of trapped and free particles in the set of trapping areas  $\mathcal{S}$ .

### 3.4.2 Simulations of trajectories and microdomains

In our first experiment, we designed two trapping regions  $\mathcal{S}_1$  and  $\mathcal{S}_2$  in a square domain  $\mathcal{D} \subset \mathbb{R}^2$  with radius  $5 \mu\text{m}$  (origin  $(0,0)$  at the bottom left corner of  $\mathcal{D}$ ):

1.  $\mathcal{S}_1$  is a circle with radius  $r_1 = 0.65 \mu\text{m}$  and center  $(2.5, 2.5)$ ;
2.  $\mathcal{S}_2$  is a circle with radius  $r_2 = 0.39 \mu\text{m}$  and center  $(4, 4)$ .

This situation typically corresponds to the analysis of a region of interest as illustrated in Figure 3.1(b) (yellow square). The parameters of simulation are given in table 3.1. Here, 2500 particles are uniformly distributed over  $\mathcal{D}$ . FLUOSIM allows us to identify which particle is trapped at time  $t$ . In that case,  $\phi(\mathbf{x}_t^{(i)})$  is known, meaning that if the particle  $i$  is trapped or not trapped at time  $t$ . In Figure 3.3, we plot the positions of the particles labelled as free or trapped at time  $t = 10 \text{ s}$  (transitory regime) and  $t = 100 \text{ s}$  (stationary regime).

Type	Parameters	Value
Biophysics parameters	$\sigma^2$	$1 \mu\text{m}^2/\text{s}$
	$k^+$	$0.2 \text{s}^{-1}$
	$k^-$	$0.05 \text{s}^{-1}$
Microscopy parameters	$\Delta t$	0.1 s
	$\Delta x$	$0.025 \mu\text{m}$

Table 3.1 – Parameters of the first FLUOSIM simulation.

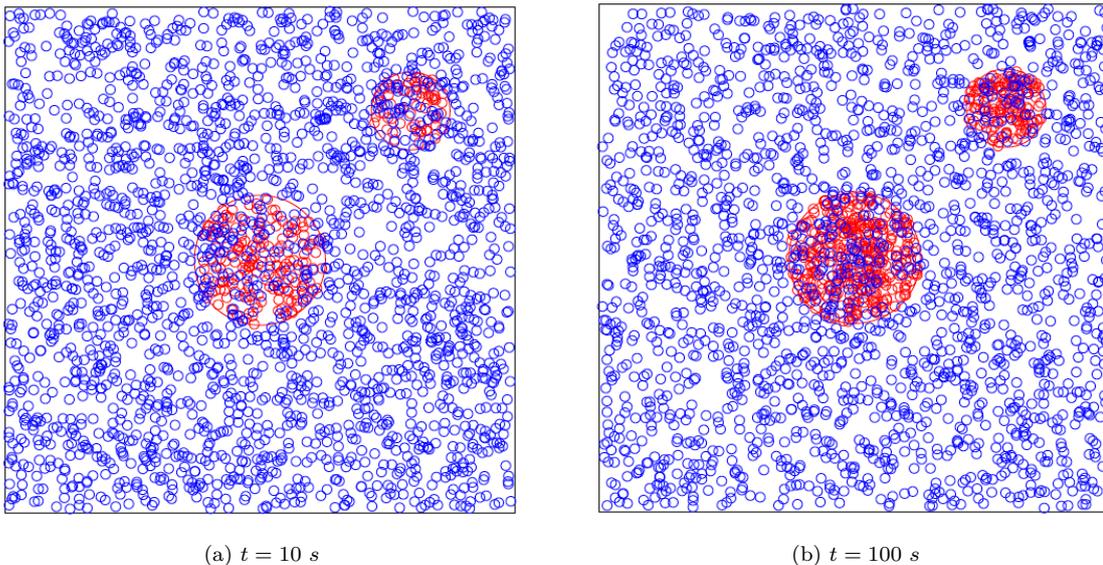


Figure 3.3 – Positions of the particles simulated with FLUOSIM at time  $t = 10 \text{ s}$  and  $t = 100 \text{ s}$  (free particles are represented in blue, trapped particles are represented in red). The red circles delineate the regions  $\mathcal{S}_1$  (large disk) and  $\mathcal{S}_2$  (small disk).

### Evaluation of proportions of trapped and free particles in the whole domain

We compare the proportion of trapped particles  $t \rightarrow b(t)$  corresponding to equation (3.5) to the true proportion of trapped particles. The curves are shown in Figure 3.4(a). The fit of the transitory regime ( $t < 5\tau_c = 77.5 \text{ s}$ ) is rather good even if the stationary regime is not exactly the same as the one predicted by the model. The predicted equilibrium is 22.53% while the mean true proportion of trapped particles computed on the last 100 steps of time is 21.35%. The relative error of the model is 5.24%, mainly due to the fact that the model oversimplifies the underlying trapping process; the dynamics of the particles are only controlled by the parameter  $p_S$ , and we assume that the particles are always in the stationary regime of a normally reflected free diffusion on  $\mathcal{D}$ . This assumption holds at the beginning, as the particles are initially drawn from the uniform distribution

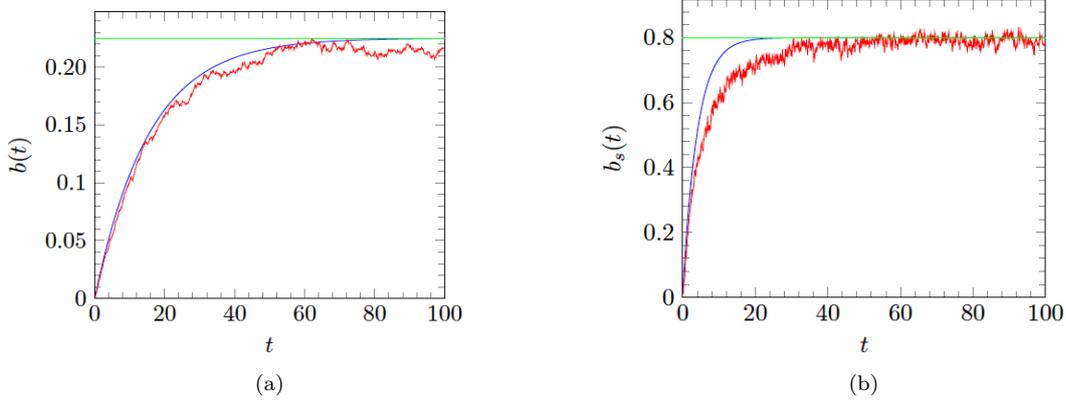


Figure 3.4 – Evolution of the proportions of trapped particles over time. (a) proportions of trapped particles in  $\mathcal{D}$ ; (b) proportions of trapped particles in the trapping region  $\mathcal{S}$ . The **true proportion** of trapped particles computed from the data, the **proportions computed** with model 3.5, and the **asymptote** of the model 3.5 are represented in **red**, **blue**, and **green**, respectively. We computed the red curve over  $m = 2500$  trajectories simulated with FLUOSIM (see table 3.1).

on  $\mathcal{D}$ . Nevertheless, as  $t$  increases, the trapping process makes this assumption invalid. This explains the relative good fit of the transitory phase and the relative lack of fit of the asymptotic phase. However, we can consider that it is a rather satisfying modeling, regarding the flexibility and the number of parameters.

### Evaluation of proportions of trapped and free particles in trapping areas

We compare the proportion of trapped particles  $t \rightarrow b_S(t)$  given by equation (3.6) to the true proportion of trapped particles inside  $\mathcal{S}$ . Interestingly, we observe the opposite behavior compared to the case of  $b(t)$  (proportion of trapped particles in the whole domain  $\mathcal{D}$ ). There is a clear lack of fit during the transitory regime but a perfect fit during the stationary regime (see Figure 3.4(b)). We can see that the transitory regime has the same duration for  $b(t)$  and  $b_S(t)$ . During this phase, the number of particles inside  $\mathcal{S}$  increases due to the process of Brownian particles entering in  $\mathcal{S}$  and getting trapped. However, it is explicitly assumed in equation (3.6) that the number of particles is constant (and large). This explains the lack of fit during the transitory phase which matches with a period during which the number of particles increases in  $\mathcal{S}$ . On the contrary, once  $b(t)$  has reached the stationary regime, the number of particles inside  $\mathcal{S}$  is approximately constant (even if it is not the same particles that remain in  $\mathcal{S}$  from one time to another). In the latter case, the model (3.5) is relevant and the stationary regime is well predicted by the

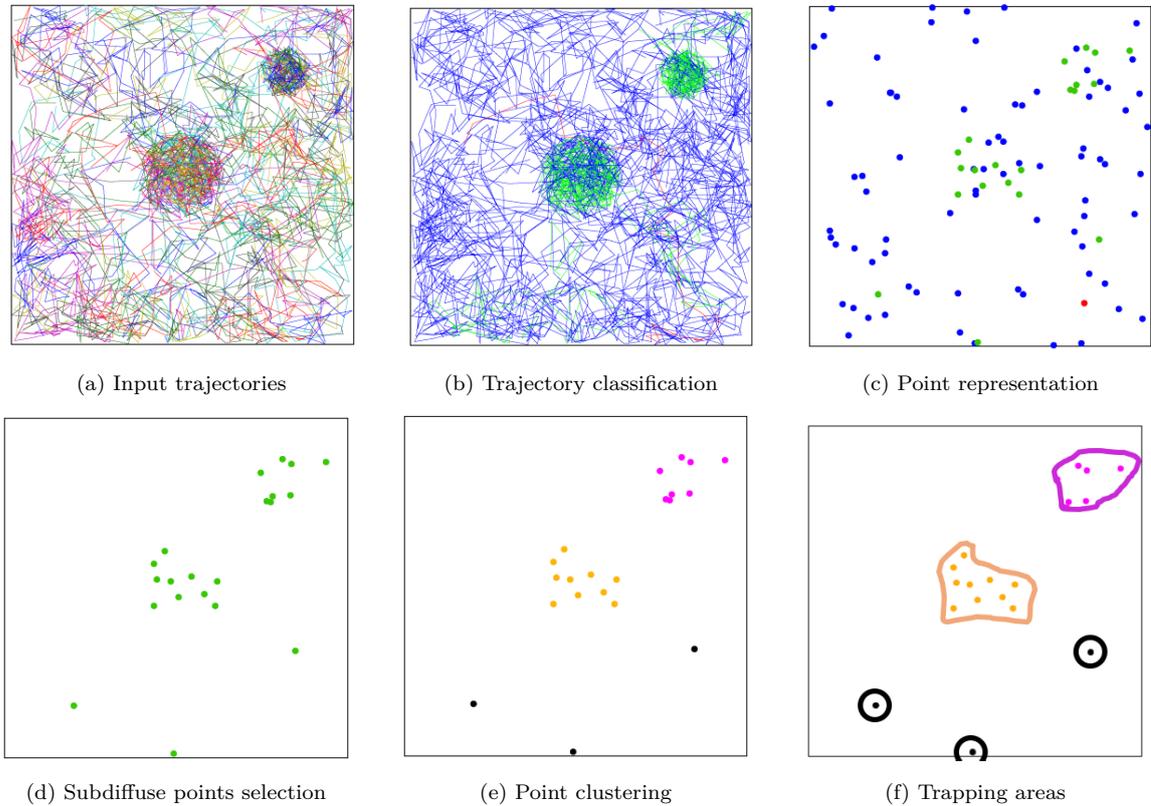


Figure 3.5 – Overview of our computational approach for detecting trapping areas. (a) input trajectories; (b) classification of trajectories into three diffusion groups (**superdiffusion**, **subdiffusion**, **free diffusion**); (c) representative points on trajectories; (d) display of subdiffusive representative points; (e) estimation of clusters associated to trapping areas; (f) detection of trapping areas.

model.

### 3.5 Clustering of trajectories for micro-domain detection

In this section, we describe the computational approach used to automatically detecting microdomains where particles undergo subdiffusion. The proposed approach is not based on the biophysical modeling and the FLUOSIM simulator presented earlier. FLUOSIM serves here to quantitatively demonstrate the performance of our approach on synthetic sequences (ground truth).

Our four-step detection procedure is based on the combination of the clustering algorithm DBSCAN [Ester et al. 1996] and the labeling of trajectories as explained in [Briane, Kervrann, and Vimond 2018]. The procedure can be used for a large range of application problems in live cell imaging, and is summarized as follows (see Figure 3.5):

1. Step #1: Classification of trajectories: we run our test procedure [Briane, Kervrann, and Vimond 2018] on the collection of trajectories  $\mathcal{X}_m$  (Figure 3.5(b)) to label the tracks into three diffusion groups: free diffusion, superdiffusion, and subdiffusion. We denote  $\Omega_1(\mathcal{X}_m)$  the set of trajectory indexes corresponding to the acceptance of hypothesis H1 (subdiffusion).
2. Step #2: Representative point selection: we choose a unique point  $\tilde{\mathbf{x}}_i$  to represent each trajectory  $\mathbb{X}_{n_i}^{(i)}$  (Figure 3.5(c)). We remove points associated to superdiffuse trajectories and Brownian trajectories (Figure 3.5(d)).
3. Step #3: Point clustering: we partition the set  $\Omega = \{\tilde{\mathbf{x}}_i | i \in \Omega_1(\mathcal{X}_m)\}$  into clusters (Figure 3.5(e)).
4. Step #4: Detection of trapping areas: we use the clusters to delineate confinement areas and microdomains (Figure 3.5(f)).

In the sequel, we successively present Steps #2, #3 and #4 of the method. As already explained, we will use the simulation framework presented in Section 3.3 to illustrate the different steps of our method.

### 3.5.1 Classification of trajectories

Step #1 is straightforward as it corresponds to the test procedure described in [Briane, Kervrann, and Vimond 2018]. This method is based on a non-parametric three-decision test to distinguish the three types of diffusion (superdiffusion, subdiffusion, free diffusion). The null hypothesis is that the observed trajectory made of  $n$  points is generated from a Brownian motion and the two other alternatives are subdiffusion and superdiffusion. The test statistic  $T_n$  is the standardized largest distance covered by the particle from its starting point. It can be interpreted as follows:

1. If the value of  $T_n$  is low, it means that the process stayed close to its initial position and can be trapped in a microdomain (subdiffusion);
2. If the value of  $T_n$  is high, the particle went far from its initial position and the particle is actively transported (e.g. via a molecular motor) (superdiffusion).

The test is appropriately normalized, asymptotically converges under the null hypothesis and under parametric alternatives which are currently considered in the biophysics literature. The test procedure, applied to subdiffusive trajectories, is easily controlled by a p-value.

Our procedure is applied to detect areas where subdiffusion occurs, but it can also be used to detect areas in which superdiffusion or free diffusion are the main dynamics. In that case, we just need to replace the set  $\Omega_1(\mathcal{X}_m)$  by, respectively,  $\Omega_0(\mathcal{X}_m)$  (set of trajectory indexes corresponding to free diffusion) or  $\Omega_2(\mathcal{X}_m)$  (set of trajectory indexes corresponding to superdiffusion).

Moreover, if we observe long trajectories, dynamics can vary, switching from one diffusion model to another. In this case, labeling a trajectory with a single diffusion class would be misleading. To overcome this problem, we can use the algorithm described in [Briane, Vimond, Valades-Cruz, et al. 2020] (also based on the classification proposed in [Briane, Kervrann, and Vimond 2018]) to detect the change points where the diffusion switches occur and segment the trajectories accordingly. Then we can use these sub-trajectories as inputs of the proposed algorithm.

Finally, it is possible that, due to the spatial extensions of the trapping regions, a trapped particle never hits the edges of the region during the period of observation. In this case, these particles have no chance to be detected as subdiffusive by the procedure described in [Briane, Kervrann, and Vimond 2018]. If we assume that the trapped particles have a significantly lower diffusion coefficient than the untrapped particles -a realistic biological assumption- we can use this supplementary assumption to detect the trapped particles even in this case. Consequently, we proceed as follows. We keep all the particles classified as subdiffusive as before. For all the particles classified as Brownian, we can compute an estimator of the diffusion coefficient as follows:

$$\hat{\sigma}_n^2 = 1/(2n\Delta t) \sum_{j=1}^n \|\mathbf{x}_{t_j} - \mathbf{x}_{t_{j-1}}\|_2^2.$$

In the next step, we classify the trajectories first detected as Brownian by using the method [Briane, Kervrann, and Vimond 2018] based on the estimated diffusion coefficient, and we re-label as subdiffusive all the trajectories detected as Brownian by the method [Briane, Kervrann, and Vimond 2018] for which the estimated diffusion coefficient is low enough. A threshold must be set by either a statistical procedure -for example based on the value of the diffusion coefficients of the trajectories detected as subdiffusive in

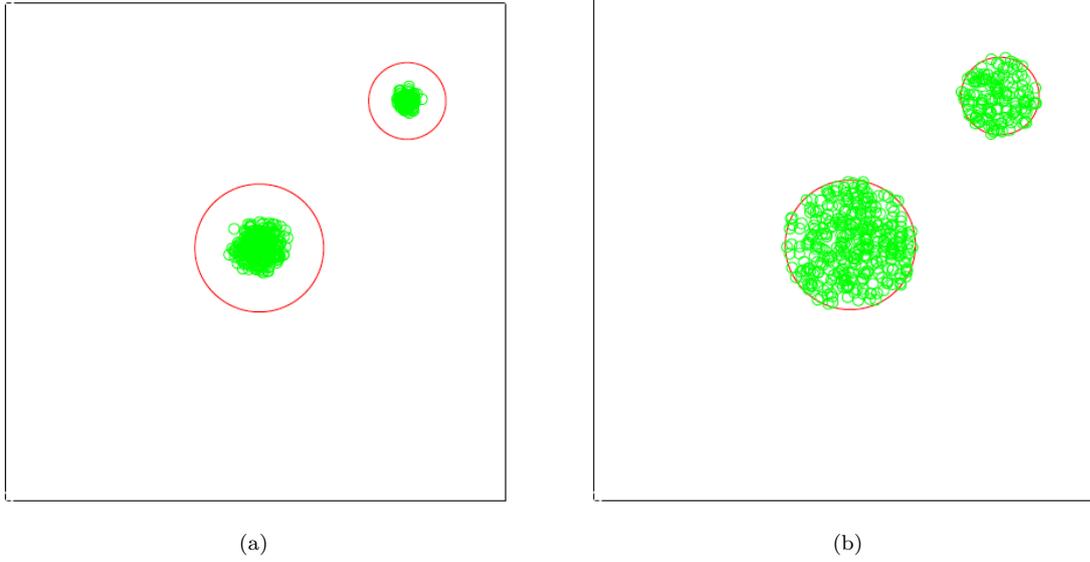


Figure 3.6 – Spatial distribution of the trapped particles in the FLUOSIM simulation; (a) the trajectories are represented by their mean point; (b) the trajectories are represented by  $\mathbf{x}_{\lfloor n/2 \rfloor}$ . A trajectory is considered to be trapped if it is trapped during the whole period of observation. The **red circles** represent the **boundaries** of the trapping regions  $\mathcal{S}_1$  (large circle) and  $\mathcal{S}_2$  (small circle).

the first place- or by an expert. These trajectories would correspond to the subdiffusive trajectories that never hit the edges of the subregions. We can then run the next steps of the procedure on this new set of subdiffusive trajectories.

### 3.5.2 Representative trajectory points and spatial distribution

In Step #2, a single point is required to represent a trajectory made of  $n$  points. In what follows, we consider mapping functions  $\mathbf{f} : \mathbb{R}^{2 \times n} \rightarrow \mathbb{R}^2$  of the following form:

- Mean point:  $\mathbf{f}(\mathbf{x}_1, \dots, \mathbf{x}_n) = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i$ ,
- $j$ -th point located on the trajectory:  $\mathbf{f}(\mathbf{x}_1, \dots, \mathbf{x}_n) = \mathbf{x}(j)$ , with  $j \in \{1, \dots, n\}$ .

As already mentioned, the trapped particles are modeled by confined free diffusion normally reflected at the boundaries of the trapping regions. In that case, we can see that different representative trajectory points yield very different spatial patterns. As illustrated in Figure 3.6, we represent the trajectory  $\mathbb{X}_n = (\mathbf{x}_{t_0}, \dots, \mathbf{x}_{t_n})$  by the mid-point  $\mathbf{x}_{\lfloor n/2 \rfloor}$  (related function  $\mathbf{f}(\mathbf{x}_1, \dots, \mathbf{x}_n) = \mathbf{x}(\lfloor n/2 \rfloor)$ ). It turns out that the spatial distribution of the representative points  $\mathbf{x}_{\lfloor n/2 \rfloor}^{(i)}$  of the trapped trajectories  $\mathbb{X}_n^{(i)}$  is uniform with

this choice, over the trapped region  $\mathcal{S}_1$  and  $\mathcal{S}_2$  (Figure 3.6(b)). This is due to proposition 1, as the trapped particles undergo a normally reflected confined free diffusion in  $\mathcal{S}_1$  and  $\mathcal{S}_2$ .

**Proposition 1 ([Pinsky 2003])**

Let  $(\mathbf{x}_t)$  be a normally reflected Brownian motion on a finite volume domain  $\mathcal{D}$ . Assume that the process  $(\mathbf{x}_t)$  has a stationary uniform distribution over  $\mathcal{D} : \mathcal{U}(\mathcal{D})$ . If the process  $(\mathbf{x}_t)$  has reached its stationary distribution, then for any  $t > 0$  and  $\mathcal{B} \subset \mathcal{D}$  it follows that

$$P(\mathbf{x}_t \in \mathcal{B}) = \frac{|\mathcal{B}|}{|\mathcal{D}|},$$

where  $|\mathcal{B}|$  denotes the area of domain  $\mathcal{B}$ .

Actually, the average position of the trapped trajectories are concentrated in the center of the trapped regions (Figure 3.6(a)). Therefore, in the purpose of accurately estimating the boundaries of  $\mathcal{S}_1$  and  $\mathcal{S}_2$ , it seems more appropriate to represent the trajectory  $i$  with  $\mathbf{x}_{\lfloor n/2 \rfloor}^{(i)}$  rather than by the average position, not necessary located on the trajectory.

### 3.5.3 A clustering algorithm: DBSCAN

In Step #3, we apply the popular clustering algorithm DBSCAN [Ester et al. 1996] to robustly estimate point clusters from noisy points due to presence of false detections. DBSCAN adds a point into a cluster if it is surrounded by at least  $\kappa$  neighboring points in a disk of radius  $\epsilon$  -this type of point is called a core point- or if it is within a disk of radius  $\epsilon$  from a core point -this type of point is called a boarder point-. On the contrary, the point does not belong to any cluster and is considered as noise. Other more sophisticated clustering algorithms could be applied. For instance, the so-called a contrario method allows a finer analysis of the clusters than DBSCAN through the concept of meaningful clusters ([Desolneux, Moisan, and Morel 2003], [F. Cao et al. 2007]) organized in a hierarchical structure. We focus on the DBSCAN algorithm since it is very popular spatial statistics. It is mainly controlled by two parameters  $(\epsilon, \kappa)$  as explained below.

#### Selection of the DBSCAN parameters

The set of parameters  $(\epsilon, \kappa)$  should be adapted to each cluster, as each cluster does not have the same density (or concentration) of points. However this information is available

once the clusters are estimated. In practice, a unique set of parameters  $(\epsilon, \kappa)$  is used for all clusters. A good choice is to set the parameters  $(\epsilon, \kappa)$  in order to detect the least dense cluster. Indeed, the clusters with more points are more easily detected than the least dense clusters. [Ester et al. 1996] argue that, in two-dimensional problems, we can set  $(\kappa = 4)$  as a rule of thumb. Once the parameter  $\kappa$  is set, we can choose  $\epsilon$  by considering two different strategies:

1. a data driven method based on the observed distribution of  $\kappa$ -nearest neighbors;
2. a parametric method to approximate the distribution of noisy points.

In the following, we present two data driven methods and two parametric methods for selecting  $\epsilon$ . Then, we compare the outcomes of the DBSCAN algorithm with different methods for selecting  $\epsilon$ , applied to FLUOSIM simulations.

**Data driven methods** The data driven methods enable to determine  $\epsilon$  from the sample  $d_1(\kappa), \dots, d_m(\kappa)$  where  $d_i(\kappa)$  denotes the distance of the  $\kappa$ -nearest neighbors of point  $\mathbf{x}_i \in \Omega$ . Let us denote  $d_1(\kappa), \dots, d_m(\kappa)$  the increasing-ordered sample. Note that if  $\epsilon = d_i(\kappa)$ , the  $i$  points corresponding to the values  $d_1(\kappa), \dots, d_i(\kappa)$  will be core points while the other points will be either boarder points or noisy points. The choice  $\epsilon < d_1(\kappa)$  corresponds to the case when all the points are noisy points (no core point), whereas the choice  $\epsilon \geq d_m(\kappa)$  corresponds to the case where all the points are core points. In the latter case, all the points belong to one single cluster. If the number of noisy points  $\aleph$  is known, a natural choice proposed by [Ester et al. 1996] is  $\epsilon = d_{(m-\aleph+1)}(\kappa)$ . An alternative graphical approach is also described in [Ester et al. 1996] to determine  $\epsilon$ . The authors plot the sequence  $d_m(\kappa), \dots, d_1(\kappa)$  and  $\epsilon$  is determined as the first  $d_{(i)}$  in the first valley of the sequence.

**Parametric methods** In [Daszykowski, Walczak, and Massart 2001], it is implicitly assumed that the noisy points are uniformly distributed. The parameter  $\epsilon$  is defined as the  $\alpha$  quantile of the distribution (e.g.  $\alpha = 0.05$ ) of the  $\kappa$ -nearest neighbors of  $|\Omega|$  points drawn from a uniform distribution over  $\mathcal{D}$ . If the observed noisy points are really drawn from a uniform distribution, none of them will be chosen as a core point with approximately  $1 - \alpha$  probability.

As the number  $\aleph$  of noisy points and the clusters  $C_i$  are unknown, it is not possible to propose a method which exactly controls the probability of detecting a noisy point as a

Type	Method	$\epsilon$
Data driven	Graphical method	5.64
Parametric	[Daszykowski, Walczak, and Massart 2001]	5.94
	[Friedman, Baskett, and Shustek 1975]	5.86

Table 3.2 – Values of the parameter  $\epsilon$  of DBSCAN obtained with different methods. The parameter  $\epsilon$  is estimated assuming  $\kappa = 4$  as recommended in [Ester et al. 1996]. The parameter  $\epsilon$  is expressed in pixel units. The image size is  $200 \times 200$  px.

core point. In [Daszykowski, Walczak, and Massart 2001],  $\epsilon$  is estimated with Monte-Carlo simulations, which is time consuming if  $|\Omega|$  is large.

It is worth noting that [Friedman, Baskett, and Shustek 1975] also studied the distribution of the  $\kappa$ -nearest neighbors of  $m$  points uniformly distributed over a finite domain  $\mathcal{D}$ . If  $|\Omega|$  is large enough to neglect boundary effects, the ratio of the volume of a two-dimensional sphere centered at a point containing  $\kappa$  neighbors and the volume of the whole space  $\mathcal{D}$  is governed by a Beta distribution  $f$  with parameters  $(\kappa, |\Omega| - \kappa)$ :

$$f(x) = \frac{m!}{(\kappa - 1)!(m - \kappa)!} x^{\kappa-1} x^{m-\kappa}, \quad 0 \leq x \leq 1.$$

Finally, we can define  $\epsilon$  as explained in [Daszykowski, Walczak, and Massart 2001]. If  $F^{-1}(\alpha)$  denotes the quantile of order  $\alpha$  of the distribution  $f$ , we get

$$\epsilon = \left( \frac{|\mathcal{D}| F^{-1}(\alpha)}{\pi} \right)^{1/d},$$

where  $|\mathcal{D}|$  is the volume (or area in the two-dimensional case) of  $\mathcal{D}$ .

### Comparison of strategies on FLUOSIM simulations

We apply the DBSCAN algorithm on the set  $\Omega$  obtained with the test procedure [Briane, Kervrann, and Vimond 2018] (see Figure 3.5(b)). The parameter  $\epsilon$  is estimated with the different methods. All the methods provided similar estimations of  $\epsilon$  in our experiments (e.g. see table 3.2). Consequently, we got the same set of clusters when we run DBSCAN as shown in Figure 3.7 by setting  $\kappa = 4$ . We found two clusters as expected, corresponding to the two trapping regions  $\mathcal{S}_1$  and  $\mathcal{S}_2$ . We got similar results in additional experiments, suggesting that selecting one method or another one would not affect the detection of trapping areas.

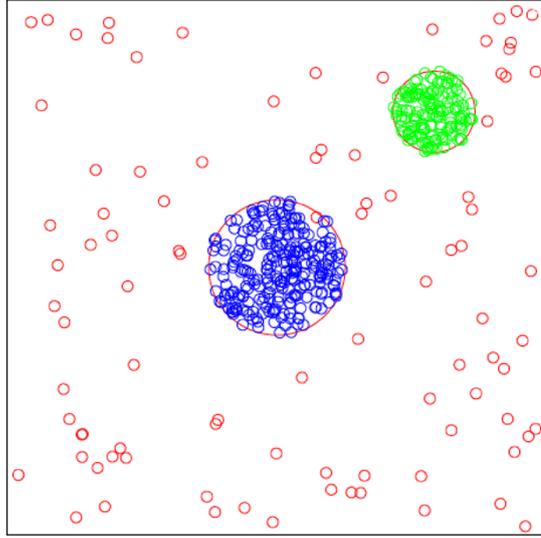


Figure 3.7 – Clusters detected by DBSCAN ( $\kappa = 4$ ) on the set of particles detected as subdiffusive with the test procedure [Briane, Kervrann, and Vimond 2018] (single test with a significance level of 5%). We choose  $\epsilon = 5.86$  derived with the Beta distribution assumption. The **noisy points** are in **red**, and points belonging to the same cluster are labelled with the same color (**blue** or **green**). The **red circles** represent the **boundaries** of the trapping regions  $\mathcal{S}_1$  (large circle) and  $\mathcal{S}_2$  (small circle). The black square delineates the borders of the whole domain  $\mathcal{D}$ .

### 3.5.4 Detection of trapping areas

In Step #4, we simply use the convex hull to define the borders of the clusters. By definition of the convex hull, all the points of the cluster will be either inside the convex hull (or convex polygon) or be some vertexes of the polygon. We note that several extensions of the convex hull have been developed, including  $\alpha$ -shapes [Edelsbrunner, Kirkpatrick, and Seidel 1983]. In that case, an extra parameter  $\alpha > 0$  is used to control the smoothness of the estimated shape.

## 3.6 Experimental results

### 3.6.1 Evaluation on FLUOSIM simulated sequences

In this section, we use the FLUOSIM simulation scheme to illustrate the potential of our approach to detect trapping areas. We consider temporal series depicting 4500 trajectories made of  $n = 350$  points. The trajectories are analyzed once the equilibrium

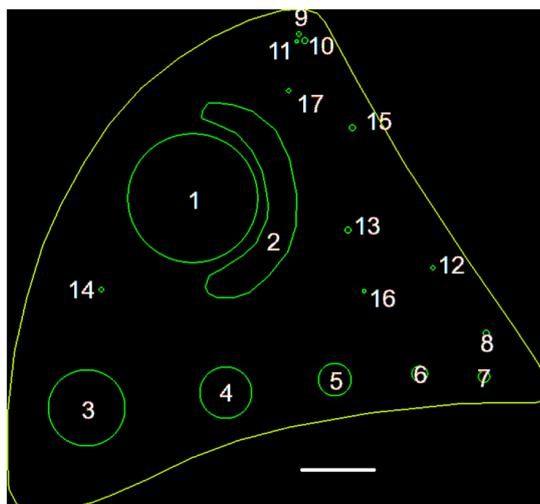


Figure 3.8 – Simulated crossbow-shaped cell with seventeen trapping regions of different sizes and shapes. The white line at the bottom accounts for  $16\ \mu\text{m}$ .

Type	Parameters	Value
Biophysics parameters	$\sigma_S^2$	$0.2\ \mu\text{m}^2/\text{s}$
	$\sigma_T^2$	$0.02\ \mu\text{m}^2/\text{s}$
	$k^+$	$10\ \text{s}^{-1}$
	$k^-$	$2\ \text{s}^{-1}$
Microscopy parameters	$\Delta t$	$0.02\ \text{s}$
	$\Delta x$	$0.160\ \mu\text{m}$

Table 3.3 – FLUOSIM parameters to simulate the first crossbow-shaped cell.

regime is reached. We recall that the trajectories can switch between two diffusion modes:

1. confined free diffusion in small microdomains  $\mathcal{S}_p$  with normal reflection on the boundaries  $\partial\mathcal{S}_p$ ;
2. Free diffusion in the whole domain  $\mathcal{D}$  with normal reflection on the boundaries  $\partial\mathcal{D}$ .

We designed a cell in with a crossbow shape and seventeen trapping regions depicted in Figure 3.8:

- one circular nucleus surrounded by one crescent- shaped zone,
- five circular zones with decreasing radius,
- seven tiny circular zones,
- three tiny circular zones really close to each other.

The parameters are given in table 3.3. In order to assess the efficiency of the method, we compute two criteria for each true (ground truth) trapping region. Firstly, we compute

	True regions					
Region index (see Figure 3.8)	1	2	3	4	5	6-17
Number of overlapping clusters	1	1	1	1	1	0
Maximum overlap (%)	90	36	60	29	15	0

Table 3.4 – Evaluation of the procedure on the crossbow-shaped cell simulation.

Cluster index	Proportion of subdiffusive trajectories (%)
1	3.3
2	5.6
3	18.5
4	8.3

Table 3.5 – Proportion of subdiffusive trajectories inside the detected regions from the crossbow-shaped cell simulation.

the number of detected regions overlapping with the real trapping regions. Secondly, we compute the proportion of overlap between the closest detected region to the trapping region (useful if multiple detected regions overlap with the true region). We define the proportion of overlap as the area of intersection of the true region and the detected region over the area of the true region. Then, ideally the two criteria should be equal to one, meaning that one detected region corresponds to one real trapping region and that the overlap is perfect. Also, for each detected cluster we compute the proportion of subdiffusive trajectories which should be compared to equation (3.7). In this case, the theoretical proportion is  $k^+/(k^+ + k^-) = 0.8333$ .

From table 3.4, we can see that the algorithm detects satisfyingly the regions #1 to #5 while it does not detect at all the regions #6 to #17. The best detected region is region #1 with 90% of overlap. The algorithm does not detect the small regions as there is either no detected subdiffusive trajectories or less than  $\kappa = 4$  detected subdiffusive trajectories (see Figure 3.9(a)). In the latter case, DBSCAN can not build a cluster as the minimum number of points in a cluster must be  $\kappa = 4$ . We also note that the proportion of subdiffusive trajectories in the five detected regions is much lower, as reported in table 3.5, than the expected value  $k^+/(k^+ + k^-) = 0.8333$ . Then, in the setting of this simulation, the algorithm under-detects subdiffusive trajectories. The parameters that influence the most the outcome of the classification step (Step #1) are the trajectory length, the microdomain size, and the diffusion coefficient of trapped particles. An optimal setting is a long trajectory with a high diffusion coefficient in a small domain. In this case, the particles bounce a lot against the edges of the trapping regions and its subdiffusive behavior

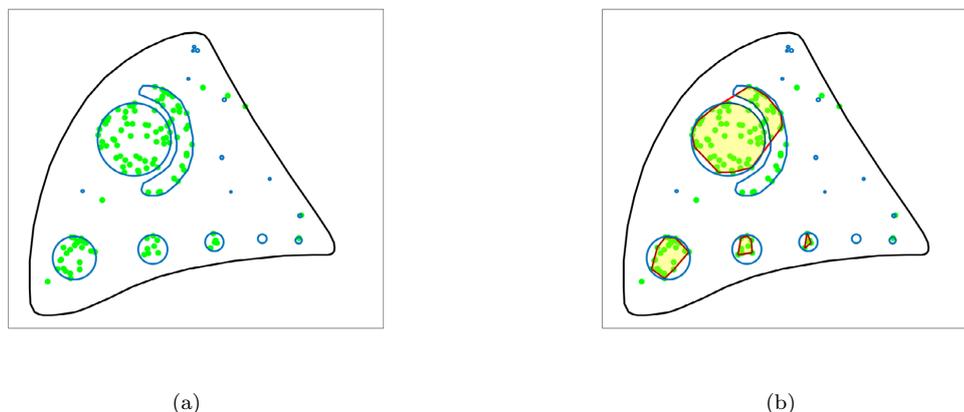


Figure 3.9 – Results of our procedure on the first crossbow-shaped cell simulation; (a) representative points of the trajectories detected as subdiffusive by the test procedure [Briane, Kervrann, and Vimond 2018] with a significance level of 5%; (b) trapping regions (yellow background, red boundaries) estimated with the convex hull from the clusters detected by DBSCAN. The true trapping regions are delineated with blue curves.

Type	Parameters	Value
Biophysics parameters	$\sigma_S^2$	$0.2 \mu\text{m}^2/\text{s}$
	$\sigma_T^2$	$0.1 \mu\text{m}^2/\text{s}$
	$k^+$	$10 \text{ s}^{-1}$
	$k^-$	$1 \text{ s}^{-1}$
Microscopy parameters	$\Delta t$	$0.02 \text{ s}$
	$\Delta x$	$0.160 \mu\text{m}$

Table 3.6 – FLUOSIM parameters to simulate the second crossbow-shaped cell.

is obvious. A more detailed simulation scheme would give the interval of parameters for which the classification step performs the best. Another issue is the proximity between two trapping regions. We can see in Figure 3.9(b) that a detected region overlaps both with region #1 (large disk) and region #2 (crescent shape). We note that the minimum distance to resolve two close regions will depend on the concentration of subdiffusive trajectories, that is on the binding parameters  $k^+$  and  $k^-$ . In fact, the outcome of DBSCAN depends on the density of points inside the clusters. We provide another simulation controlled by the parameters given in table 3.6. The results are shown in Figure 3.10 and the quantitative results are reported in tables 3.7 and 3.8. In conclusion, we did not choose the optimal simulation parameters here to assess our method. Still, we are able to detect some trapping regions even if the algorithm under-detects subdiffusive trajectories and

	True regions								
Region index (see Figure 3.8)	1	2	3	4	5	6	7-14	15	16-17
Number of overlapping clusters	4	3	1	1	1	1	0	1	0
Maximum overlap (%)	31	17	24	43	25	34	0	7	0

Table 3.7 – Evaluation of the procedure on the second simulated crossbow-shaped cell.

Cluster index	Proportion of subdiffusive trajectories (%)
1	42.0
2	19.0
3	5.0
4	15.0
5	8.5
6	11.0
7	11.0
8	8.0

Table 3.8 – Proportion of subdiffusive trajectories inside the detected regions from the crossbow-shaped cell simulation (2) to be compared ( $k^+/(k^+ + k^-) = 0.9091$ ).

merges close trapping regions. A more systematic simulation study should be performed to determine the range of parameters for which the method gives satisfying results.

### 3.6.2 Experiments on real TIRF images: dynamics of Rab11 involved in exocytosis

In this section, we demonstrate the potential of our approach on Rab11-mCherry proteins involved in exocytosis mechanisms and observed in 2D TIRF (Total Internal Reflection Fluorescence) microscopy (Nikon TE2000 inverted microscope equipped with a  $100\times$  TIRF objective: numerical aperture = 1.49, exposure time  $\Delta_t = 100$  ms,  $dx = dy = 160$  nm). Figure 3.11 shows three  $256 \times 353$  images at different time points, depicting fluorescent vesicles close to the plasma membrane. The image sequences are typically composed of 600 images, and show the late steps of exocytosis driven by Rab11 at the plasma membrane, namely the tethering-docking process [Schafer et al. 2014].

Given several hundreds of Rab11 trajectories estimated with the ICY software [Chenouard, Bloch, and Olivo-Marin 2013] [Chenouard, Smal, et al. 2014], we first applied our test procedure [Briane, Kervrann, and Vimond 2018] to classify the trajectories. In Figure 3.12(a), we displayed all trajectory points onto a 2D three-color map: **free diffusion** in **blue**, **subdiffusion** in **green**, and **superdiffusion** in **red**. In the particular case shown in Figure 3.11,

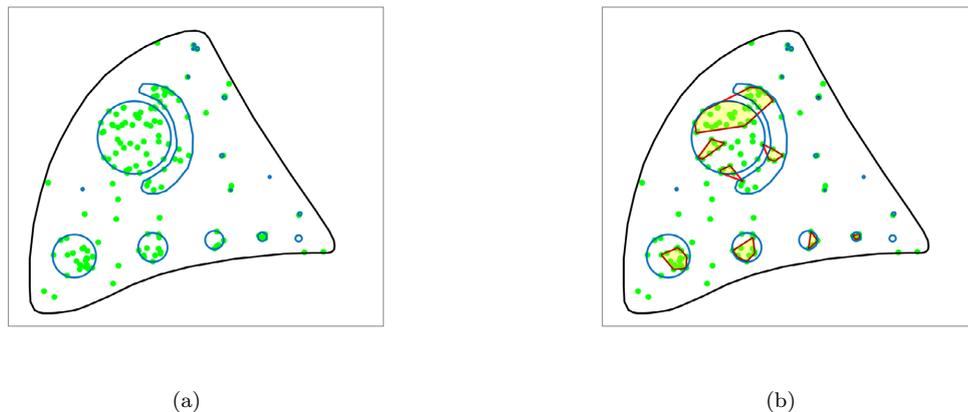


Figure 3.10 – Results of our procedure on the second crossbow-shaped cell simulation; (a) representative points of the trajectories detected as subdiffusive by the test procedure [Briane, Kervrann, and Vimond 2018] with a significance level of 5%; (b) trapping regions (yellow background, red boundaries) estimated with the convex hull from the clusters detected by DBSCAN. The true trapping regions are delineated with blue curves.

the M10 cells were treated with Latrunculin A, which inhibits actin polymerization and then reduces active transport of vesicles [Pécot et al. 2018]. This effect is confirmed on Figure 3.12(a) since no track was labeled a superdiffusive trajectory.

Second, we applied our computational approach on subdiffusive trajectories (green pixels in Figure 3.12(a)) represented by points in Figure 3.12(b). In the sequel, we assume that the tethering-docking process under study is associated to subdiffusion because of multiple interactions with actin filaments and interactions with the plasma membrane. All points reported in Figure 3.12(b) correspond to vesicles that fuse at the surface of the plasma membrane. First, we observe an approximately uniform spatial distribution of points (black dots). Nevertheless, we detected a few additional clusters corresponding to higher densities of fusion events in a local spatial neighborhood. These clusters can be interpreted as local domains related to more intensive interactions with the plasma membrane and actin filaments. Several clusters are actually localized in regions with a high concentration of spots (e.g. see top right of the cell). Trajectory analysis is then helpful to focus the attention on specific areas which cannot be easily detected by examining fluctuations of fluorescence intensities shown in Figure 3.11.

Finally, we illustrate in Figure 3.12(c) the potential of the [Hozé, Nair, et al. 2012] method on the same sequence. In particular, we reported the spatially varying diffusion coefficients on a 2D map since the drift vectors are very noisy and not interpretable. In

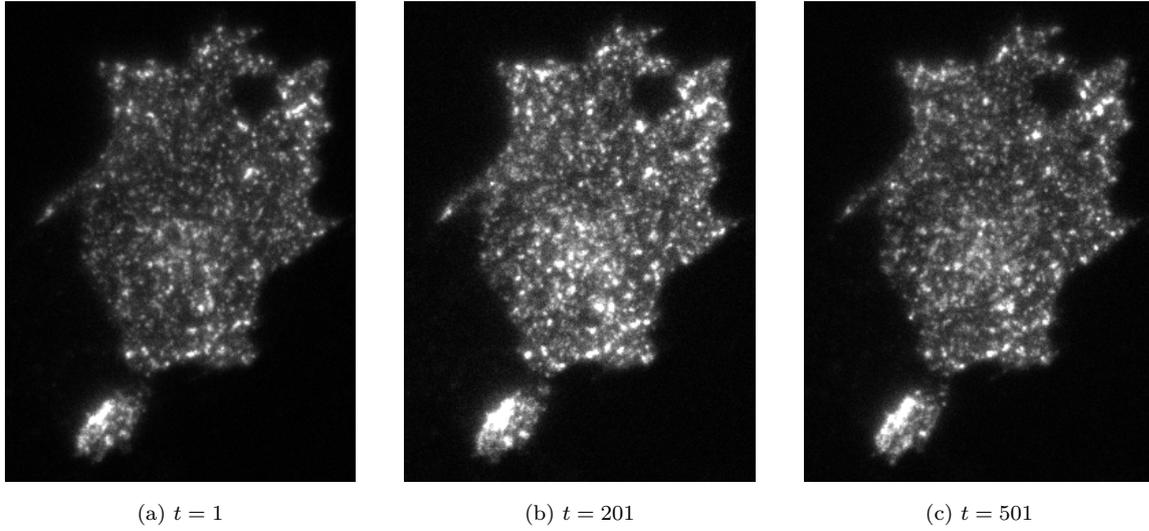


Figure 3.11 – Rab11-mCherry proteins within a M10 cell treated with Latrunculin and observed at several time points. The images are extracted from a 2D TIRFM sequence composed of 600 images (exposure time  $\Delta t = 100$  ms,  $dx = dy = 160$  nm).

our study, the diffusion map is actually not really helpful to draw conclusions if the drift values are noisy. The number of estimated trajectories (about one thousand) is too small to safely apply the methods described in [Hozé, Nair, et al. 2012] the second chapter of this thesis.

### 3.7 Conclusion and discussion

In this chapter, we presented a method to detect the trapping regions where the particles are confined and thereby undergo subdiffusion. We combined the clustering algorithm DBSCAN [Ester et al. 1996] and a test procedure [Briane, Kervrann, and Vimond 2018] in order to identify the areas with a high concentration of subdiffusive particles, assumed to correspond to trapping regions. With this approach, we were also able to estimate the proportion of particles inside the trapping regions which were indeed confined in the domain. Our approach is more flexible than the Eulerian approach of [Hozé, Nair, et al. 2012], which is not able to handle mixtures of motions, or to capture the motion of particles not confined in the trapping regions but still going through these regions.

We illustrated the potential of our procedure on simulations generated with the FLUOSIM software, and on real TIRF images. However, the simulation results provided here

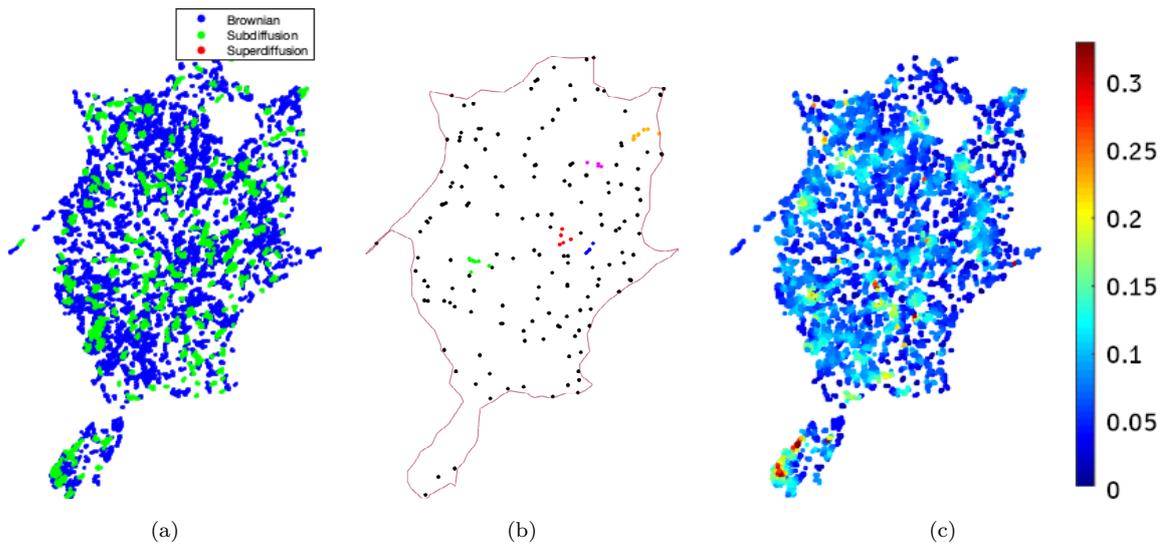


Figure 3.12 – Analysis and clustering of trajectories; (a) trajectory classification into three diffusion groups [Briane, Kervrann, and Vimond 2018]; (b) selection/clustering of subdiffusive trajectories: the point clusters are labeled with colors while individual subdiffusive trajectory are represented with black dots; (c) estimated diffusion coefficient map generated with our method presented in Chapter 2.

do not intend to assess systematically the procedure for a large range of parameters values ( $k^+$ ,  $k^-$ ,  $\sigma_S$ ,  $\sigma_T$ , the shapes of the regions). For future works, it would typically involve:

1. determining jointly the region sizes and trajectory lengths for which the classification procedure can detect subdiffusive trajectories (Step #1) and further defining a criterion to select the region size for different shapes;
2. determining the minimum distance between two regions for the clustering algorithm to be able to detect two distinct regions and not merge them. This distance is expected to depend on the concentration of subdiffusive trajectories in the trapping regions.

In this chapter, we focused on the modeling and estimation framework and demonstrated its potential with preliminary experiments. Then, we emphasize the fact that it is possible to use different procedures at each step of the proposed pipeline to adapt to different scenarios. For instance, in the case of trajectories switching over time, we can use sub-trajectories estimated with the method described in [Briane, Vimond, Valades-Cruz, et al. 2020], instead of the classification procedure [Briane, Kervrann, and Vimond 2018] used in Step #1 of our algorithm. The same way, another choice of representative

point and other strategies of clustering or shape estimation can also be used, respectively, in Steps #2, #3 and #4. We proposed a default choice of procedures for each step of the pipeline but the user is free to select the appropriate procedures depending on the application.

# CONCLUSION

---

## Contributions of the thesis

In this thesis, we developed algorithmic tools to analyze fluorescence microscopy data. In particular, we built new methods reliably producing cellular maps and detecting intracellular features from tracking data or microscopy image sequences, while outperforming existing methods.

In the first chapter, we gave an overview of the mathematical notions and tools necessary to understand our research. After introducing stochastic processes, Brownian motion, and the notion of diffusion, we presented the theory behind particle motion modeling, and in particular the approach of Langevin [Langevin 1908]. We then reviewed several types of fluorescence microscopy techniques, whether classic or groundbreaking, 2D or 3D, diffraction-limited or super-resolved. On the algorithmic side, we presented different biomolecule tracking methods, and made an extensive presentation of Vincent Briane’s work on automated trajectory motion classification and motion class switching among trajectories [Briane, Kervrann, and Vimond 2018] [Briane, Vimond, Valades-Cruz, et al. 2020].

In the second chapter, we proposed a new mapping method to robustly estimate diffusion and drift in the cell from particle tracks. After a review of two techniques dedicated to this purpose, we presented in details the theoretical elements of our statistical method, which uses space-time kernel estimators guided by particle motion classification on particle positions. Our computational approach was tested on an extensive set of experiments, and gave satisfactory results on both simulated, real, 2D and 3D data, providing high-resolution estimated maps, even with a low number of trajectories, and proving to be more robust than existing methods. This implies that our method can be used on a more diverse set of fluorescence microscopy techniques, and can detect subtle biological events in image sequences.

In the third chapter, we studied the detection of confinement areas [Briane, Salomon, et al. 2020]. We introduced a method which detects the cellular microdomains where particles get trapped. After a presentation of the mathematical and biophysical framework

---

of this problem, we introduced its related work. Then, we presented our method, which uses trajectory classification as well as a combination of clustering procedures. Its main idea is to detect high concentrations of subdiffusive particles, by reducing subdiffusive trajectories to a set of representative points, and then running said set through a clustering algorithm such as DBSCAN. Our method was evaluated on simulated sequences obtained with FLUOSIM, and on real TIRF image sequences. It could prove its potential and flexibility, even if there is still room for improvement.

This corpus of research and algorithms should provide biologists with new analytical tools that will further exploit the potential of fluorescence microscopy. Notably, our work could have an impact on fundamental biological research, with potential applications in fields such as cancer research (e.g. breast cancer), the study of viruses (e.g. HIV), and neurology. For instance, [Baloul et al. 2023] uses our work to study the behavior of CSL transcription complexes in the Notch signaling pathway, whose dysfunction is linked to cancer development [Nowell and Radtke 2017][Ntziachristos et al. 2014].

## Future work and extensions

While our work is worthy of interest, there is still room for improvement and extensions, and we would like to suggest some avenues for development here.

First of all, there are two features for our diffusion and drift mapping method that could unfortunately not be developed in time. The first one would have been an automated and adaptive kernel size selector. For the moment being, the size of the estimation kernel of our method is manually selected by the user, through one single constant parameter called  $m$ . It could be interesting to let the method automatically choose the value of this parameter in function of the available data density. In the first instance, this could be performed globally by calculating one single value for the whole data set. It would basically replace the current user choice and relieve them from this sometimes drastic choice. In a second step, a more sophisticated adaptive kernel could be implemented, that could locally change its size depending on the local space-time data density and quality, allowing the method to be more robust and precise. Implementing these two features would at the same time simplify the user experience and make the method more accurate. Another feature that was roughly developed but could not be tested is the artificial data quantity augmentation by bootstrapping. This consists of processing an image sequence several times through the same tracking algorithm, while incrementally removing the first

---

frames. Each iteration generates stochastic variations in the detected particle coordinates at time  $t = t_i$ , and results can be concatenated into a denser set of particle positions. This feature would allow to potentially go even further in the ability to analyze lightweight data sets.

Another development path concerns the implementation of our method. Our algorithm was developed in MATLAB, which is a slow, interpreted language and software, and processing times can be concerningly high when working on large data sets. Another problem with this tool is the display of 3D results, which is often sub-optimal in terms of reading and difficult to navigate. Also, MATLAB is a proprietary software on which our method depends. A solution to these problems could be to rewrite our code in a compiled language such as C++. This would allow to generate a fast, standalone version of our algorithm, featuring a clear and accessible Graphic User Interface (GUI), as well as a more user-friendly display for 3D results. This could also facilitate the technology transfer work that has already been initiated with our MATLAB code, in particular with the biotechnology start-up Myriade which develops innovative imaging techniques for nanometric particles based on interferometry.

Finally, as a challenge to our work, possibilities offered by machine learning could also be explored. As a cutting-edge and widely applicable technology, machine learning based methods often provide impressive results and are yet to be tested on every aspect of microscopy data analytics. Previous experiments made in our research team showed that these methods do not outperform more classic approaches on particle motion classification. Still, it could be an interesting path of research to test this technology on a more global approach, covering both particle motion classification, diffusion and drift mapping, and cluster detection.

As an final note, we hope to pursue at least some of the above-mentioned developments in the future, and we would warmly welcome anyone interested in taking our work as a starting point for their research.



# APPENDIX

---

## A.1 Derivation of the global binding and unbinding rates

We demonstrate that the equalities (3.3) and (3.4) hold. Let us consider the equality (3.3) involving  $\gamma_1$ :

$$\begin{aligned} P(\phi(X_{t+h}) = 1 \mid \phi(X_t) = 0) &= P(\phi(X_{t+h}) = 1, X_{t+h} \in \mathcal{S} \mid \phi(X_t) = 0) + \\ &\quad P(\phi(X_{t+h}) = 1, X_{t+h} \in \overline{\mathcal{S}} \mid \phi(X_t) = 0) \\ &= P(\phi(X_{t+h}) = 1 \mid \phi(X_t) = 0, X_{t+h} \in \mathcal{S}) \times \\ &\quad P(X_{t+h} \in \mathcal{S} \mid \phi(X_t) = 0) + 0. \end{aligned}$$

The particle can not be trapped outside  $\mathcal{S}$ . Consequently, the second probability of the sum is zero. As the probability to be in region  $\mathcal{S}$  at  $t + h$  is independent from the event  $\{\phi(X_t) = 0\}$  to be a free at  $t$ , we finally get:

$$P(\phi(X_{t+h}) = 1 \mid \phi(X_t) = 0) = (k^+h + o(h)) P(X_{t+h} \in \mathcal{S}).$$

Normally reflected Brownian motion in a domain  $\mathcal{D}$  with finite volume has a stationary distribution. This distribution is the uniform distribution over  $\mathcal{D}$  [Pinsky 2003]. Then, under the uniform distribution assumption, we get:

$$\begin{aligned} P(\phi(X_{t+h}) = 1 \mid \phi(X_t) = 0) &= (k^+h + o(h)) \frac{|\mathcal{S}|}{|\mathcal{D}|} \\ &= k^+ \frac{|\mathcal{S}|}{|\mathcal{D}|} h + o(h). \end{aligned}$$

It follows that  $\gamma_1 = k^+ |\mathcal{S}| / |\mathcal{D}| = k^+ p_{\mathcal{S}}$ .

---

We now prove the equality (3.4) involving  $\gamma_2$ :

$$\begin{aligned}
P(\phi(X_{t+h}) = 0 \mid \phi(X_t) = 1) &= P(\phi(X_{t+h}) = 0, X_{t+h} \in \mathcal{S} \mid \phi(X_t) = 1) + \\
&P(\phi(X_{t+h}) = 0, X_{t+h} \in \bar{\mathcal{S}} \mid \phi(X_t) = 1) \\
&= P(\phi(X_{t+h}) = 0 \mid \phi(X_t) = 1, X_{t+h} \in \mathcal{S}) \times \\
&P(X_{t+h} \in \mathcal{S} \mid \phi(X_t) = 1) + 0.
\end{aligned}$$

As  $\{\phi(X_t) = 1\} \subset \{X_{t+h} \in \mathcal{S}\}$ , we get the result:

$$P(\phi(X_{t+h}) = 0 \mid \phi(X_t) = 1) = (k^- h + o(h)) \times 1,$$

which implies  $\gamma_2 = k^-$ .

## A.2 Additional results: langerin in RPE-1 cells

We used here an image sequence acquired at Institut Curie with a 3D multi-angle TIRF microscope showing YFP-tagged langerin in a RPE-1 crossbow-shaped cell. The image sequence was 2D-projected and tracking was performed with U-TRACK. The data has notably been used in [Boulanger et al. 2014] and [Balsollier et al. 2023].

Our method produced the outputs shown in Figure A.1 for the tracking data analysis and in Figure A.2 for the diffusion maps. Additionally, Figure A.2 also illustrates the drift map.

Figure A.2(a) and Figure A.2(b) show the diffusion point cloud output and the interpolated diffusion bitmap inferred from the point cloud, respectively. Figure A.2(c) displays the diffusion histogram corresponding to the point cloud. Figure A.2(d) show the drift vector field.

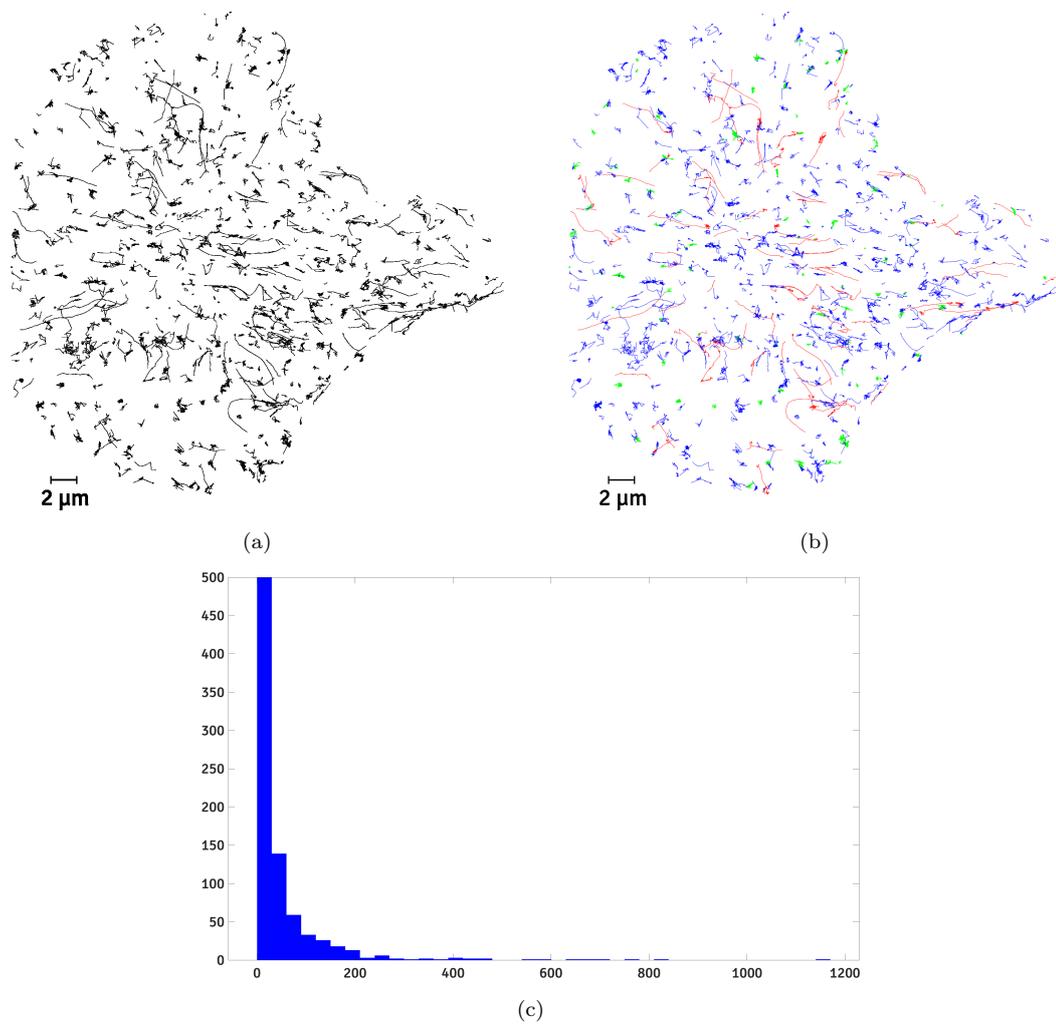


Figure A.1 – The tracking data obtained from the langerin in crossbow-shaped cells data set is analyzed. **(a)** The set of all simulated particle trajectories is shown. **(b)** Particle trajectories are classified by motion type. Free diffusion is represented in blue, subdiffusion is represented in green and superdiffusion is represented in red. **(c)** The histogram of the input trajectory lengths is shown.

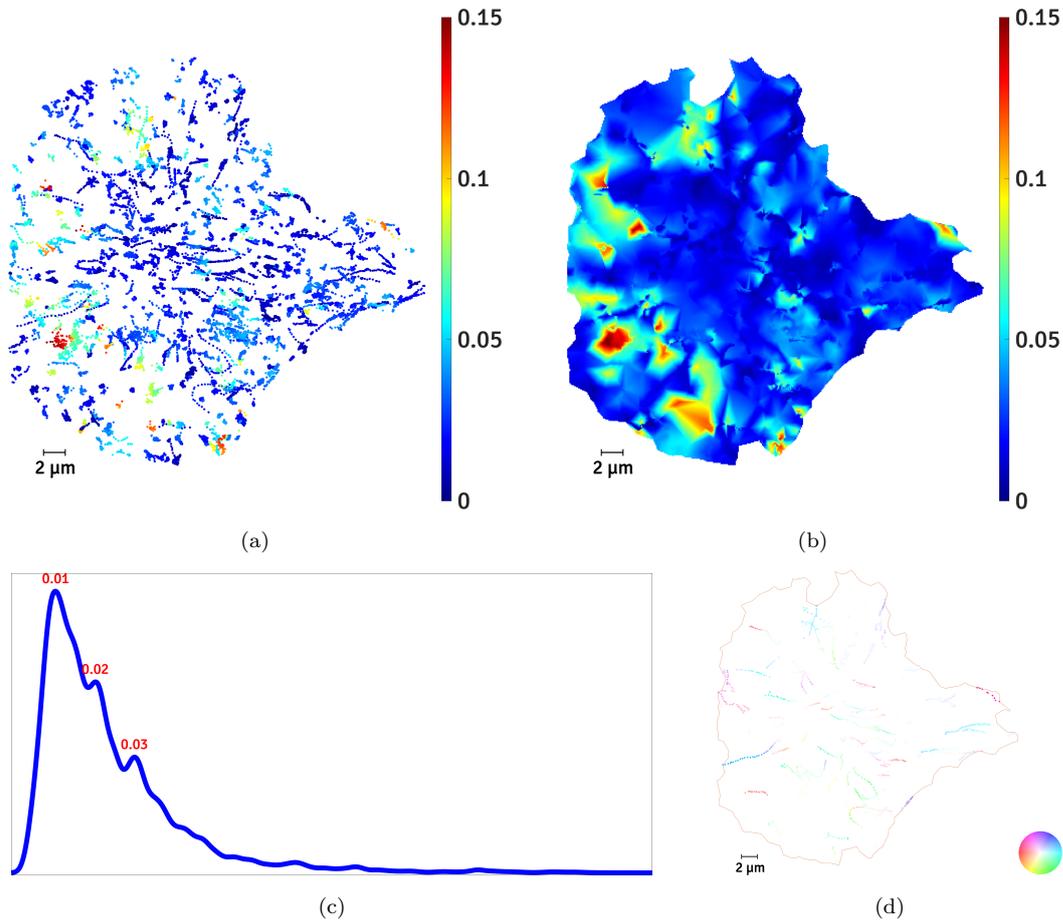


Figure A.2 – The results obtained with our method on real data (langerin in crossbow-shaped cells), are shown. **(a)** The point cloud output represents the spatial distribution of the diffusion coefficient inside the cell, in  $\mu\text{m}^2/\text{s}$ . **(b)** The point cloud is derived into a map by interpolation, with the same scale units. **(c)** The point cloud is derived into a diffusion histogram. The values displayed on top of the modes are also in  $\mu\text{m}^2/\text{s}$  **(d)** This second point cloud represents the spatial distribution of the drift inside the cell. Each point represents a drift vector. The wheel indicates both the direction and the amplitude of the vectors.

# BIBLIOGRAPHY

---

- Abrahamsson, Sara et al. (Jan. 2013), « Fast multicolor 3D imaging using aberration-corrected multifocus microscopy », *in: Nature Methods* 10.1, pp. 60–63.
- Airy, George Biddell (1835), « On the diffraction of an object glass with circular aperture », *in: Transactions of the Cambridge Philosophical Society* 5.3, pp. 283–291.
- Allersma, Miriam et al. (Oct. 2004), « Visualization of regulated exocytosis with a granule-membrane probe using Total Internal Reflection Microscopy », *in: Molecular Biology of the Cell* 15.10, pp. 4658–4668.
- Axelrod, Daniel (2003), « Total internal reflection fluorescence microscopy in cell biology », *in: Methods in Enzymology* 361, pp. 1–33.
- Axelrod, Daniel, Thomas Burghardt, and Nancy Thompson (June 1984), « Total Internal Reflection Fluorescence », *in: Annual Review of Biophysics and Bioengineering* 13.1, pp. 247–268.
- Axelrod, Daniel and Michael W. Davidson (n.d.), *Specialized microscopy techniques - Total Internal Reflection Fluorescence microscopy - introduction and theoretical aspects*, URL: <https://www.olympus-lifescience.com/fr/microscope-resource/primer/techniques/fluorescence/tirf/tirfintro/>.
- Baddeley, Adrian, Imre Bárány, and Rolf Schneider (2007), « Spatial point processes and their applications », *in: Stochastic Geometry, Lectures given at the C.I.M.E. Summer School held in Martina Franca, Italy, September 13–18, 2004*, vol. 1892, Lecture Notes in Mathematics, Springer Berlin Heidelberg, pp. 1–75.
- Baloul, Sarah et al. (Aug. 2023), « Changes in searching behaviour of CSL transcription complexes in Notch active conditions », *in: bioRxiv* 2023.08.10.552734.
- Balsollier, Lisa et al. (Mar. 2023), « A generative model to synthesize spatio-temporal dynamics of biomolecules in cells », *in: arXiv* 2303.06951.
- Basset, Antoine et al. (July 2017), « An extended model of vesicle fusion at the plasma membrane to estimate protein lateral diffusion from TIRF microscopy images », *in: BMC Bioinformatics* 18.1, p. 352.

- 
- Beheiry, Mohamed El, Maxime Dahan, and Jean-Baptiste Masson (July 2015), « InferenceMAP: mapping of single-molecule dynamics with Bayesian inference », *in: Nature Methods* 12.7, pp. 594–595.
- Beheiry, Mohamed El, Silvan Türkcan, et al. (Mar. 2016), « A primer on the Bayesian approach to high-density single-molecule trajectories analysis », *in: Biophysical Journal* 110.6, pp. 1209–1215.
- Berry, Hugues and Hugues Chaté (Feb. 2014), « Anomalous diffusion due to hindering by mobile obstacles undergoing Brownian motion or Orstein-Uhlenbeck processes », *in: Physical Review E* 89.2, p. 022708.
- Betzig, Eric et al. (Sept. 2006), « Imaging intracellular fluorescent proteins at nanometer resolution », *in: Science* 313.5793, pp. 1642–1645.
- Blackman, Samuel S. (Jan. 2004), « Multiple hypothesis tracking for multiple target tracking », *in: IEEE Aerospace and Electronic Systems Magazine* 19.1, pp. 5–18.
- Boulanger, Jérôme et al. (Nov. 2014), « Fast high-resolution 3D total internal reflection fluorescence microscopy by incidence angle scanning and azimuthal averaging », *in: Proceedings of the National Academy of Sciences of the United States of America* 111.48, pp. 17164–17169.
- Brémaud, Pierre (2020), *Markov chains, Gibbs fields, Monte Carlo simulation and queues*, vol. 31, Texts in Applied Mathematics, Springer Cham.
- Bressloff, Paul and Jay Newby (Jan. 2013), « Stochastic models of intracellular transport », *in: Reviews of Modern Physics* 85.1, pp. 135–196.
- Briane, Vincent, Charles Kervrann, and Myriam Vimond (June 2018), « Statistical analysis of particle trajectories in living cells », *in: Physical Review E* 97.6, p. 062121.
- Briane, Vincent, Antoine Salomon, et al. (Feb. 2020), « A computational approach for detecting micro-domains and confinement domains in cells: a simulation study », *in: Physical Biology* 17.2, p. 025002.
- Briane, Vincent, Myriam Vimond, and Charles Kervrann (July 2020), « An overview of diffusion models for intracellular dynamics analysis », *in: Briefings in Bioinformatics* 21.4, pp. 1136–1150.
- Briane, Vincent, Myriam Vimond, Cesar Augusto Valades-Cruz, et al. (Jan. 2020), « A sequential algorithm to detect diffusion switching along intracellular particle trajectories », *in: Bioinformatics* 36.1, pp. 317–329.

- 
- Cao, Frédéric et al. (Feb. 2007), « A unified framework for detecting groups and application to shape recognition », *in: Journal of Mathematical Imaging and Vision* 27.2, pp. 91–119.
- Cao, Hongyuan and Wei Wu (Dec. 2015), « Changepoint estimation: another look at multiple testing problems », *in: Biometrika* 102.4, pp. 974–980.
- Chandler, David (1987), *Introduction to modern statistical mechanics*, Oxford University Press.
- Chaumont, Fabrice et al. (July 2012), « Icy: an open bioimage informatics platform for extended reproducible research », *in: Nature Methods* 9.7, pp. 690–696.
- Chen, Bi-Chang et al. (Oct. 2014), « Lattice light-sheet microscopy: imaging molecules to embryos at high spatiotemporal resolution », *in: Science* 346.6208, p. 1257998.
- Chenouard, Nicolas, Isabelle Bloch, and Jean-Christophe Olivo-Marin (Nov. 2013), « Multiple Hypothesis Tracking for cluttered biological image sequences », *in: IEEE Transactions on Pattern Analysis and Machine Intelligence* 35.11, pp. 2736–3750.
- Chenouard, Nicolas, Ihor Smal, et al. (Mar. 2014), « Objective comparison of particle tracking methods », *in: Nature Methods* 11.3, pp. 281–289.
- Chmyrov, Andriy et al. (Aug. 2013), « Nanoscopy with more than 100,000 'doughnuts' », *in: Nature Methods* 10.8, pp. 737–740.
- Cole, Richard, Tushare Jinadasa, and Claire Brown (Dec. 2011), « Measuring and interpreting point spread functions to determine confocal microscope resolution and ensure quality control », *in: Nature Protocols* 6.12, pp. 1929–1941.
- Coutin, Laure and Zhongmin Qian (Jan. 2002), « Stochastic analysis, rough path analysis and fractional Brownian motions », *in: Probability Theory and Related Fields* 122.1, pp. 108–140.
- Daszykowski, Michal, Beata Walczak, and D.L Massart (May 2001), « Looking for natural patterns in data: Part 1. Density-based approach », *in: Chemometrics and Intelligent Laboratory Systems* 56.2, pp. 83–92.
- Dertinger, Thomas et al. (Dec. 2009), « Fast, background-free, 3D super-resolution optical fluctuation imaging (SOFI) », *in: Proceedings of the National Academy of Sciences of the United States of America* 106.52, pp. 22287–22292.
- Desolneux, Agnès, Lionel Moisan, and Jean-Michel Morel (Dec. 2003), « Maximal Meaningful Events and applications to image analysis », *in: The Annals of Statistics* 31.6, pp. 1822–1851.

- 
- Edelsbrunner, Herbert, David Kirkpatrick, and Raimund Seidel (July 1983), « On the shape of a set of points in the plane », *in: IEEE Transactions on Information Theory* 29.4, pp. 551–559.
- Einstein, Albert (Jan. 1905), « On the motion of small particles suspended in liquids at rest required by the molecular-kinetic theory of heat », *in: Annalen der Physik* 322.8, pp. 549–560.
- Endesfelder, Ulrike and Mike Heilemann (2015), « Direct Stochastic Optical Reconstruction Microscopy (dSTORM) », *in: Advanced Fluorescence Microscopy, Methods and Protocols*, vol. 1251, Methods in Molecular Biology, Springer New York, pp. 263–276.
- Ester, Martin et al. (Aug. 1996), « A density-based algorithm for discovering clusters in large spatial databases with noise », *in: Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, Portland, Oregon, USA: AAAI Press, pp. 226–231.
- Finkel, Raphael and Jon Louis Bentley (Mar. 1974), « Quad trees: a data structure for retrieval on composite keys », *in: Acta Informatica* 4.1, pp. 1–9.
- Floderer, Charlotte et al. (Nov. 2018), « Single molecule localisation microscopy reveals how HIV-1 Gag proteins sense membrane virus assembly sites in living host CD4 T cells », *in: Scientific Reports* 8.1, p. 16283.
- Friedman, Jerome H., Forest Baskett, and Len J. Shustek (Oct. 1975), « An algorithm for finding nearest neighbors », *in: IEEE Transactions on Computers* C-24.10, pp. 1000–1006.
- Friedrich, Rudolf and Joachim Peinke (Feb. 1997), « Description of a turbulent cascade by a Fokker-Planck equation », *in: Physical Review Letters* 78.5, pp. 863–866.
- Fusco, Dahlene et al. (Jan. 2003), « Single mRNA molecules demonstrate probabilistic movement in living mammalian cells », *in: Current Biology* 13.2, pp. 161–167.
- Gidon, Alexandre et al. (June 2012), « A Rab11A/myosin Vb/Rab11-FIP2 complex frames two late recycling steps of langerin from the ERC to the plasma membrane », *in: Traffic* 13.6, pp. 815–833.
- Gomez Lamarca, Maria et al. (Mar. 2018), « Activation of the notch signaling pathway in vivo elicits changes in CSL nuclear dynamics », *in: Developmental Cell* 44.5, pp. 611–623.
- Granik, Naor et al. (July 2019), « Single-particle diffusion characterization by deep learning », *in: Biophysical Journal* 117.2, pp. 185–192.

- 
- Greenfield, Derek et al. (June 2009), « Self-organization of the Escherichia coli chemotaxis network imaged with super-resolution light microscopy », *in: PLoS Biology* 7.6, e1000137.
- Gustafsson, Mats G.L. (May 2000), « Surpassing the lateral resolution limit by a factor of two using structured illumination microscopy », *in: Journal of Microscopy* 198.2, pp. 82–87.
- Gustafsson, Mats G.L. et al. (June 2008), « Three-dimensional resolution doubling in wide-field fluorescence microscopy by structured illumination », *in: Biophysical Journal* 94.12, pp. 4957–4970.
- Gustafsson, Nils et al. (Aug. 2016), « Fast live-cell conventional fluorophore nanoscopy with ImageJ through super-resolution radial fluctuations », *in: Nature Communications* 7.1, p. 12471.
- Hajj, Bassam, Mohamed El Beheiry, and Maxime Dahan (Mar. 2016), « PSF engineering in multifocus microscopy for increased depth volumetric imaging », *in: Biomedical Optics Express* 7.3, pp. 726–731.
- Hajj, Bassam, Mohamed El Beheiry, Ignacio Izeddin, et al. (Aug. 2014), « Accessing the third dimension in localization-based super-resolution microscopy », *in: Physical Chemistry Chemical Physics* 16.31, pp. 16340–16348.
- Hajj, Bassam, Jan Wisniewski, et al. (Nov. 2014), « Whole-cell, multicolor superresolution imaging using volumetric multifocus microscopy », *in: Proceedings of the National Academy of Sciences of the United States of America* 111.49, pp. 17480–17485.
- Hell, Stefan and Jan Wichmann (June 1994), « Breaking the diffraction resolution limit by stimulated emission: stimulated-emission-depletion fluorescence microscopy », *in: Optics Letters* 19.11, pp. 780–782.
- Holcman, David, Pierre Parutto, et al. (Oct. 2018), « Single particle trajectories reveal active Endoplasmic Reticulum luminal flow », *in: Nature Cell Biology* 20.10, pp. 1118–1125.
- Holcman, David and Zeev Schuss (Mar. 2017), « 100 years after Smoluchowski: stochastic processes in cell biology », *in: Journal of Physics A: Mathematical and Theoretical* 50.9, p. 093002.
- Hozé, Nathanaël and David Holcman (Dec. 2014), « Residence times of receptors in dendritic spines analyzed by stochastic simulations in empirical domains », *in: Biophysical Journal* 107.12, pp. 3008–3017.

- 
- Hozé, Nathanaël and David Holcman (Mar. 2017), « Statistical methods for large ensembles of super-resolution stochastic single particle trajectories in cell biology », *in: Annual Review of Statistics and Its Application* 4.1, pp. 189–223.
- Hozé, Nathanaël, Deepak Nair, et al. (Oct. 2012), « Heterogeneity of AMPA receptor trafficking and molecular interactions revealed by superresolution analysis of live cell imaging », *in: Proceedings of the National Academy of Sciences of the United States of America* 109.42, pp. 17052–17057.
- Huang, Bo et al. (Feb. 2008), « Three-dimensional super-resolution imaging by stochastic optical reconstruction microscopy », *in: Science* 319.5864, pp. 810–813.
- Hummer, Gerhard (Jan. 2005), « Position-dependent diffusion coefficients and free energies from Bayesian analysis of equilibrium and replica molecular dynamics simulations », *in: New Journal of Physics* 7.1, p. 34.
- Hurst, Harold Edwin (Jan. 1951), « Long-term storage capacity of reservoirs », *in: Transactions of the American Society of Civil Engineers* 116.1, pp. 770–799.
- Jaqaman, Khuloud et al. (Aug. 2008), « Robust single-particle tracking in live-cell time-lapse sequences », *in: Nature Methods* 5.8, pp. 695–702.
- Jeon, Jae-Hyung et al. (Jan. 2011), « In vivo anomalous diffusion and weak ergodicity breaking of lipid granules », *in: Physical Review Letters* 106.4, p. 048103.
- Johannes, Ludger and Cesar Valades (Aug. 2021), « The final twist in endocytic membrane scission », *in: Nature Cell Biology* 23.8, pp. 812–813.
- Karlin, Samuel and Howard M. Taylor (1981), *A second course in stochastic processes*, Academic Press.
- Klar, Thomas and Stefan Hell (July 1999), « Subdiffraction resolution in far-field fluorescence microscopy », *in: Optics Letters* 24.14, pp. 954–956.
- Klebaner, Fima C. (2012), *Introduction to stochastic calculus with applications*, Imperial College Press.
- Kolmogorov, Andrey Nikolaevich (Jan. 1941), « The local structure of turbulence in incompressible viscous fluid for very large Reynolds numbers », *in: Proceedings of the USSR Academy of Sciences* 30, pp. 301–305.
- Kou, Shingchang (June 2008), « Stochastic modeling in nanoscale biophysics: subdiffusion within proteins », *in: The Annals of Applied Statistics* 2.2, pp. 501–535.
- Kramers, Hendrik Anthony (Apr. 1940), « Brownian motion in a field of force and the diffusion model of chemical reactions », *in: Physica* 7.4, pp. 284–304.

- 
- Lachuer, Hugo et al. (Feb. 2023), « Spatial organization of lysosomal exocytosis relies on membrane tension gradients », *in: Proceedings of the National Academy of Sciences of the United States of America* 120.8, e2207425120.
- Lagardère, Matthieu et al. (Nov. 2020), « FluoSim: simulator of single molecule dynamics for fluorescence live-cell and super-resolution imaging of membrane proteins », *in: Scientific Reports* 10.1, p. 19954.
- Langevin, Paul (Mar. 1908), « Sur la théorie du mouvement brownien », *in: Comptes Rendus Hebdomadaires des Séances de l'Académie des Sciences* 146, pp. 530–533.
- Lemaigre, Camille et al. (Apr. 2023), « N-BAR and F-BAR proteins - Endophilin-A3 and PSTPIP1 - control clathrin-independent endocytosis of L1CAM », *in: Traffic* 24.4, pp. 190–212.
- Liu, Zhe, Luke Lavis, and Eric Betzig (May 2015), « Imaging live-cell dynamics and structure at the single-molecule level », *in: Molecular Cell* 58.4, pp. 644–659.
- Loerke, Dinah, Walter Stühmer, and Martin Oheim (Sept. 2002), « Quantifying axial secretory-granule motion with variable-angle evanescent-field excitation », *in: Journal of Neuroscience Methods* 119.1, pp. 65–73.
- Mandelbrot, Benoit B. and John W. Van Ness (Oct. 1968), « Fractional Brownian motions, fractional noises and applications », *in: SIAM Review* 10.4, pp. 422–437.
- Martin-Fernandez, Marisa, Christopher Tynan, and Stephen Webb (Oct. 2013), « A ‘pocket guide’ to total internal reflection fluorescence », *in: Journal of Microscopy* 252.1, pp. 16–22.
- Masson, Jean-Baptiste, Didier Casanova, et al. (Jan. 2009), « Inferring maps of forces inside cell membrane microdomains », *in: Physical Review Letters* 102.4, p. 048103.
- Masson, Jean-Baptiste, Patrice Dionne, et al. (Jan. 2014), « Mapping the energy and diffusion landscapes of membrane proteins at the cell surface using high-density single-molecule imaging and Bayesian inference: application to the multiscale dynamics of glycine receptors in the neuronal membrane », *in: Biophysical Journal* 106.1, pp. 74–83.
- Meroz, Yasmine and Igor Sokolov (Apr. 2015), « A toolbox for determining subdiffusive mechanisms », *in: Physics Reports* 573, pp. 1–29.
- Metzler, Ralf and Joseph Klafter (Dec. 2000), « The random walk’s guide to anomalous diffusion: a fractional dynamics approach », *in: Physics Reports* 339.1, pp. 1–77.

- 
- Momboisse, Fanny et al. (July 2022), « Tracking receptor motions at the plasma membrane reveals distinct effects of ligands on CCR5 dynamics depending on its dimerization status », *in: eLife* 11, e76281.
- Monnier, Nilah, Zachary Barry, et al. (Sept. 2015), « Inferring transient particle transport dynamics in live cells », *in: Nature Methods* 12.9, pp. 838–840.
- Monnier, Nilah, Syuan-Ming Guo, et al. (Aug. 2012), « Bayesian approach to MSD-based analysis of particle motion in live cells », *in: Biophysical Journal* 103.3, pp. 616–626.
- Nowell, Craig and Freddy Radtke (Mar. 2017), « Notch as a tumour suppressor », *in: Nature Reviews Cancer* 17.3, pp. 145–159.
- Ntziachristos, Panagiotis et al. (Mar. 2014), « From fly wings to targeted cancer therapies: a centennial for notch signaling », *in: Cancer Cell* 25.3, pp. 318–334.
- Oheim, Martin et al. (Sept. 2019), « Calibrating evanescent-wave penetration depths for biological TIRF microscopy », *in: Biophysical Journal* 117.5, pp. 795–809.
- Okhonin, Victor (July 1991), « A method of examination of sample microstructure », USSR Invention Patent SU1374922-A1.
- Parutto, Pierre et al. (Aug. 2022), « High-throughput super-resolution single-particle trajectory analysis reconstructs organelle dynamics and membrane reorganization », *in: Cell Reports Methods* 2.8, p. 100277.
- Pécot, Thierry et al. (Aug. 2018), « A quantitative approach for analyzing the spatio-temporal distribution of 3D intracellular events in fluorescence microscopy », *in: eLife* 7, e32311.
- Pinsky, Ross (May 2003), « Asymptotics of the principal eigenvalue and expected hitting time for positive recurrent elliptic operators in a domain with a small puncture », *in: Journal of Functional Analysis* 200.1, pp. 177–197.
- Qian, Hong, Michael Patrick Sheetz, and Elliot L. Elson (Oct. 1991), « Single particle tracking. Analysis of diffusion and flow in two-dimensional systems », *in: Biophysical Journal* 60.4, pp. 910–921.
- Racine, Victor, Ariane Hertzog, et al. (Apr. 2006), « Multiple-target tracking of 3D fluorescent objects based on simulated annealing », *in: 3rd IEEE International Symposium on Biomedical Imaging: Nano to Macro*, Arlington, Virginia, USA: IEEE, pp. 1020–1023.
- Racine, Victor, Martin Sachse, et al. (Mar. 2007), « Visualization and quantification of vesicle trafficking on a three-dimensional cytoskeleton network in living cells », *in: Journal of Microscopy* 225.3, pp. 214–228.

- 
- Rayleigh, John W. (Oct. 1879), « Investigations in optics, with special reference to the spectroscope », *in: The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 8.49, pp. 261–274.
- Rego, Elizabeth et al. (Dec. 2011), « Nonlinear structured-illumination microscopy with a photoswitchable protein reveals cellular structures at 50-nm resolution », *in: Proceedings of the National Academy of Sciences of the United States of America* 109.3, E135–E143.
- Reid, Donald B. (Dec. 1979), « An algorithm for tracking multiple targets », *in: IEEE Transactions on Automatic Control* 24.6, pp. 843–854.
- Renard, Henri-François et al. (Jan. 2015), « Endophilin-A2 functions in membrane scission in clathrin-independent endocytosis », *in: Nature* 517.7535, pp. 493–496.
- Römer, Winfried et al. (Nov. 2007), « Shiga toxin induces tubular membrane invaginations for its uptake into cells », *in: Nature* 450.7170, pp. 670–675.
- Roudot, Philippe, Liya Ding, et al. (Nov. 2017), « Piecewise-stationary motion modeling and iterative smoothing to track heterogeneous particle motions in dense environments », *in: IEEE Transactions on Image Processing* 26.11, pp. 5395–5410.
- Roudot, Philippe, Wesley Legant, et al. (July 2022), « U-track 3D: measuring and interrogating intracellular dynamics in three dimensions », *in: bioRxiv 2020.11.30.404814*.
- Rust, Michael, Mark Bates, and Xiaowei Zhuang (Oct. 2006), « Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM) », *in: Nature Methods* 3.10, pp. 793–796.
- Saffarian, Saveez and Tomas Kirchhausen (Mar. 2008), « Differential evanescence nanometry: live-cell fluorescence measurements with 10-nm axial resolution on the plasma membrane », *in: Biophysical Journal* 94.6, pp. 2333–2342.
- Saxton, Michael J. (Feb. 1994), « Anomalous diffusion due to obstacles: a Monte Carlo study », *in: Biophysical Journal* 66.2, pp. 394–401.
- Saxton, Michael J. and Ken Jacobson (June 1997), « Single-particle tracking: applications to membrane dynamics », *in: Annual Review of Biophysics and Biomolecular Structure* 26.1, pp. 373–399.
- Sbalzarini, Ivo and Petros Koumoutsakos (Aug. 2005), « Feature point tracking and trajectory analysis for video imaging in cell biology », *in: Journal of Structural Biology* 151.2, pp. 182–195.
- Schafer, Jenny C. et al. (Mar. 2014), « Rab11-FIP2 interaction with MYO5B regulates movement of Rab11a-containing recycling vesicles », *in: Traffic* 15.3, pp. 292–308.

- 
- Scher, Harvey and Elliott W. Montroll (Sept. 1975), « Anomalous transit-time dispersion in amorphous solids », *in: Physical Review B* 12.6, pp. 2455–2477.
- Schermelleh, Lothar et al. (Jan. 2019), « Super-resolution microscopy demystified », *in: Nature Cell Biology* 21.1, pp. 72–84.
- Schön, Ingmar et al. (Sept. 2011), « Binding-Activated Localization Microscopy of DNA structures », *in: Nano Letters* 11.9, pp. 4008–4011.
- Schuss, Zeev (2010), *Theory and applications of stochastic processes, An analytical approach*, vol. 170, Applied Mathematical Sciences, Springer New York.
- Sharonov, Alexey and Robin M. Hochstrasser (Dec. 2006), « Wide-field subdiffraction imaging by accumulated binding of diffusing probes », *in: Proceedings of the National Academy of Sciences of the United States of America* 103.50, pp. 18911–18916.
- Shroff, Hari, Helen White, and Eric Betzig (Mar. 2013), « Photoactivated Localization Microscopy (PALM) of adhesion complexes », *in: Current Protocols in Cell Biology* 58.1, pp. 4.21.1–4.21.28.
- Sibarita, Jean-Baptiste (2005), « Deconvolution microscopy », *in: Microscopy Techniques*, vol. 95, Advances in Biochemical Engineering/Biotechnology, Springer Berlin Heidelberg, pp. 201–243.
- (June 2014), « High-density single-particle tracking: quantifying molecule organization and dynamics at the nanoscale », *in: Histochemistry and Cell Biology* 141.6, pp. 587–595.
- Siegert, Silke, Rudolf Friedrich, and Joachim Peinke (July 1998), « Analysis of data sets of stochastic systems », *in: Physical Letters A* 243.5–6, pp. 275–280.
- Sims, Ruth et al. (Sept. 2020), « Single Molecule Light Field Microscopy », *in: Optica* 7.9, pp. 1065–1072.
- Singer, Robert, Ronald Sea, and Kim Housewright (July 1974), « Derivation and evaluation of improved tracking filter for use in dense multitarget environments », *in: IEEE Transactions on Information Theory* 20.4, pp. 423–432.
- Smal, Ihor et al. (Dec. 2008), « Multiple object tracking in molecular bioimaging by Rao-Blackwellized marginal particle filtering », *in: Medical Image Analysis* 12.6, pp. 764–777.
- Smith, Stephen, Claudia Cianci, and Ramon Grima (June 2017), « Macromolecular crowding directs the motion of small molecules inside cells », *in: Journal of the Royal Society Interface* 14.131, p. 20170047.

- 
- Soubies, Emmanuel et al. (Apr. 2019), « Improving 3D MA-TIRF reconstruction with deconvolution and background estimation », *in: 2019 IEEE 16th International Symposium on Biomedical Imaging*, Venice, Italy: IEEE, pp. 316–320.
- Spilger, Roman et al. (Jan. 2020), « A recurrent neural network for particle tracking in microscopy images using future information, track hypotheses, and multiple detections », *in: IEEE Transactions on Image Processing* 29, pp. 3681–3694.
- Stabley, Daniel et al. (Sept. 2015), « Real-time fluorescence imaging with 20 nm axial resolution », *in: Nature Communications* 6.1, p. 8307.
- Stelzer, Ernst et al. (Nov. 2021), « Light Sheet Fluorescence Microscopy », *in: Nature Reviews Methods Primers* 1.1, p. 73.
- Uhlenbeck, George Eugene and Leonard Salomon Ornstein (Sept. 1930), « On the theory of the Brownian motion », *in: Physical Review* 36.5, pp. 823–841.
- Van Kampen, Nicolaas Godfried (2007), *Stochastic processes in physics and chemistry*, Elsevier.
- Voronoi, Georges (1908a), « Nouvelles applications des paramètres continus à la théorie des formes quadratiques. Premier mémoire. Sur quelques propriétés des formes quadratiques positives parfaites », *in: Journal für die Reine und Angewandte Mathematik* 133, pp. 97–178.
- (1908b), « Nouvelles applications des paramètres continus à la théorie des formes quadratiques. Deuxième mémoire. Recherches sur les paralléloèdres primitifs. », *in: Journal für die Reine und Angewandte Mathematik* 134, pp. 198–287.
- Waterston, John James, Francis Beaufort, and John William Strutt (Dec. 1892), « On the physics of media that are composed of free and perfectly elastic molecules in a state of motion », *in: Philosophical Transactions of the Royal Society A* 183, pp. 1–79.
- Zhang, Bo, Josiane Zerubia, and Jean-Christophe Olivo-Marin (Apr. 2007), « Gaussian approximations of fluorescence microscope Point-Spread Function models », *in: Applied Optics* 46.10, pp. 1819–1829.
- Zwanzig, Robert (2001), *Nonequilibrium statistical mechanics*, Oxford University Press.



# ABBREVIATIONS

---

**2D** Two-Dimensional

**3D** Three-Dimensional

**BALM** Binding-Activated Localization Microscopy

**CTRW** Continuous Time Random Walk

**D3M2** Dense Diffusion and Drift Mapping Method

**DIC** Differential Interference Contrast

**DiNa** Differential Nanometry

**fBm** Fractional Brownian Motion

**FCS** Fluorescence Correlation Spectroscopy

**FM** Fluorescence Microscopy

**FRAP** Fluorescence Recovery After Photobleaching

**FWHM** Full Width at Half-Maximum

**GFP** Green Fluorescent Protein

**GLE** Generalized Langevin Equation

**HILO** Highly Inclined and Laminated Optical sheet

---

**LFM** Light Field Microscopy

**LLSM** Lattice Light-sheet Microscopy

**LSFM** Light Sheet Fluorescence Microscopy

**MAP** Maximum A Posteriori

**MHT** Multiple Hypothesis Tracking

**MSD** Mean Square Displacement

**NA** Numerical Aperture

**NL** Non-linear

**PAF** PhotoActivation of Fluorescence

**PAINT** Point Accumulation for Imaging in Nanoscale Topography

**PALM** Photoactivated Localization Microscopy

**PSF** Point-Spread Function

**mRNA** Messenger Ribonucleic Acid

**ROI** Region Of interest

**SDE** Stochastic Differential Equation

**SIM** Structured Illumination Microscopy

**SMLM** Single Molecule Localization Microscopy

---

**SPT** Single-Particle Tracking

**SRM** Super Resolution Microscopy

**STAR** Simultaneous Two-wavelength Axial Ratiometry

**STED** Stimulated Emission Depletion

**STICS** Spatio-Temporal Image Correlation Microscopy

**STORM** Stochastic Optical Reconstruction Microscopy

**dSTORM** Direct Stochastic Optical Reconstruction Microscopy

**TIRF** Total Internal Reflection Fluorescent

**WFFM** Wide-Field Fluorescence Microscopy

**Titre :** Modélisations statistiques pour l'analyse de la diffusion des molécules et du trafic intracellulaire en microscopie de fluorescence

**Mot clés :** Microscopie de fluorescence, dynamique des particules, suivi de particules, diffusion, dérive, cartographie, classification de mouvements, région de confinement

**Résumé :** En permettant de visualiser de manière sélective les dynamiques des particules dans les cellules, la microscopie de fluorescence est un outil de premier plan pour la recherche biomédicale. Il existe par conséquent une forte demande d'outils algorithmiques capables d'analyser automatiquement des données microscopiques brutes. Après une présentation de la théorie et des techniques entourant les dynamiques des particules, la microscopie de fluorescence, les méthodes de suivi et la classification de mouvements, nous présentons dans cette thèse une nouvelle méthode de cartographie basée sur des estimateurs à noyaux spatio-temporels permettant d'estimer de manière robuste la diffu-

sion et la dérive intracellulaires à partir des données de suivi. Nous l'évaluons dans un ensemble étendu d'expériences utilisant des données simulées, réelles, 2D et 3D et montrons que notre méthode fournit des cartes de diffusion et de dérive précises et exactes tout en étant plus performante que les méthodes existantes. En tant que telle, elle permet les biologistes d'étudier les dynamiques intracellulaires de particules spécifiquement marquées avec une plus large gamme d'acquisitions et de techniques de microscopie de fluorescence. En outre, nous présentons une méthode de détection de domaines de confinement utilisant également le suivi de particules et la classification de mouvements.

**Title:** Statistical modelling for the analysis of molecular diffusion and intracellular trafficking in fluorescence microscopy

**Keywords:** Fluorescence microscopy, particle dynamics, single-particle tracking, diffusion, drift, mapping, motion classification, confinement region

**Abstract:** Fluorescence microscopy is a tool of primary interest in biomedical research as it allows to selectively visualize particle dynamics within the cell. Hence, there is a high demand for algorithmic tools capable of automatically analyzing raw microscopic data. After a presentation of the theory and techniques surrounding particle dynamics, fluorescence microscopy, tracking methods and motion classification, we present in this thesis a new mapping method based on spatiotemporal kernel estimators that robustly estimate intracellular

diffusion and drift from tracking data. We evaluate it in an extensive set of experiments using simulated, real, 2D and 3D data and show that our method provides precise and accurate diffusion and drift maps while outperforming existing methods. As such, it allows biologists to study intracellular dynamics of specifically tagged particles with a wider range of acquisitions and fluorescence microscopy techniques. In addition, we present a confinement domain detection method, using particle tracks and motion classification as well.