

An adaptive statistical test to detect non Brownian diffusion from particle trajectories

Vincent BRIANE Myriam VIMOND Charles KERVRANN

INRIA Rennes Team Serpico & ENSAI

May 24, 2016



- 1 Introduction
- 2 The MSD method
- 3 Statistical procedure
- 4 Experimental results
- 5 Conclusion

Outline

- 1 Introduction
- 2 The MSD method
- 3 Statistical procedure
- 4 Experimental results
- 5 Conclusion

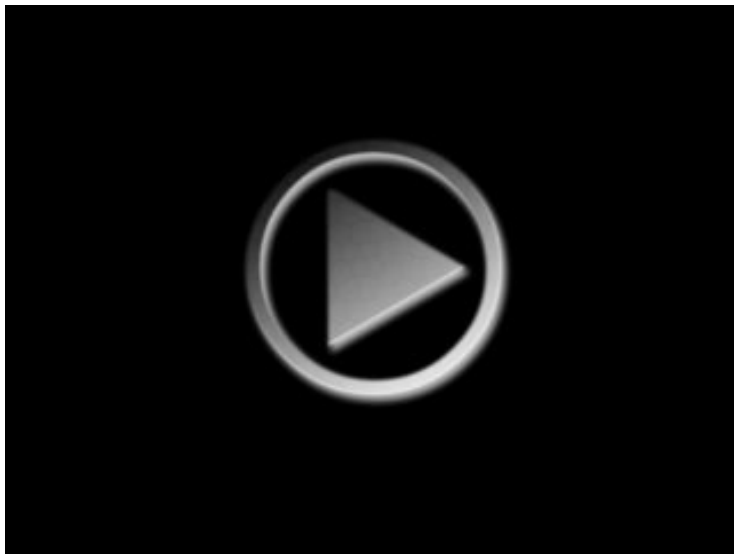


Figure: Simulation of a **directed**, **confined** and **free** diffusion.

Objective

We model the trajectories of intra-cellular object/molecules with **diffusions**:

- ① **Directed diffusion**: the particle is transported actively via molecular motors along the cytoskeleton [Bressloff and Newby, 2013].
- ② **Free diffusion** (or **Brownian** motion): the particle evolves freely in the cytosol [Bressloff and Newby, 2013].
- ③ **Confined diffusion**: the particle is confined in a domain or evolves in an open but crowded area [Berry and Chaté, 2014].

A commonly used method for classifying motions: Mean Square Displacement.
We propose a **statistical test** as an alternative.

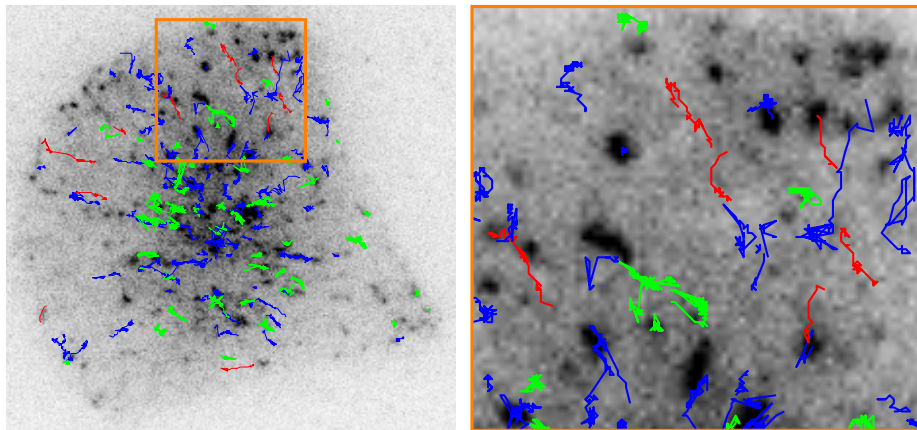


Figure: Labelling of the trajectories on the Rab11 protein sequence in a single cell (Courtesy of UMR 144 CNRS Institut Curie - PICT IBiSA).

Model

- The observed trajectory of the particle is the $(n + 1)$ vector in \mathbf{R}^2

$$X = (X_{t_0}, X_{t_1}, \dots, X_{t_n}) \quad (1)$$

vector of the successive positions of the particle.

- $t_i - t_{i-1} = \Delta t$ is the resolution time of the sensor.
- X is generated by the stochastic process $(X_t)_{t_0 \leq t \leq t_n}$ solution of the SDE.

Stochastic Differential Equation

- Stochastic Differential Equation (SDE):

$$dX_t = \mu(X_t, t)dt + \sigma dB_t.$$

- The **drift** $\mu(x, t)$ models the deterministic force.
- The **diffusion coefficient** σ models an isotropic random force.

Solution of a SDE

- $(X_t)_{t_0 \leq t \leq t_n}$ is a **Strong solution** of the SDE if it is solution of the SDE where the Brownian motion $(B_t)_{t_0 \leq t \leq t_n}$ is given.

- **Weak solution.** We can build :

$$(X_t, B_t)_{t_0 \leq t \leq t_n}$$

where $(X_t)_{t_0 \leq t \leq t_n}$ is a solution of the SDE driven by the Brownian $(B_t)_{t_0 \leq t \leq t_n}$.

- Condition on μ for having a weak solution, **the linear growth hypothesis** :

$$\exists K, \quad \forall t \in [t_0, t_n], \quad \forall x \in \mathbf{R}^2, \quad \|\mu(x, t)\|_2 \leq K(1 + \|x\|_2).$$

Outline

- 1 Introduction
- 2 The MSD method
- 3 Statistical procedure
- 4 Experimental results
- 5 Conclusion

Confined and directed diffusion.

- In the literature, confined and directed diffusion are defined from the **Mean Square Displacement** (MSD) :

$$\text{MSD}(t) = \mathbb{E}(\|X_t - X_0\|_2^2).$$

- For the Brownian motion σB_t we have :

$$\mathbb{E}(\|B_t - B_0\|_2^2) = \sigma t.$$

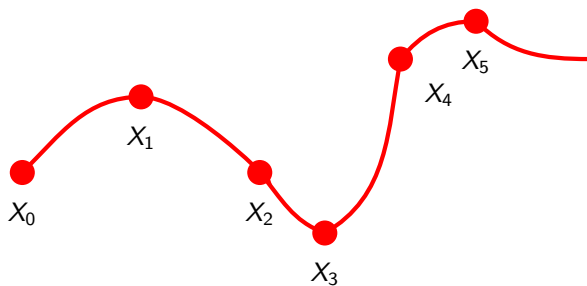
- If $t \mapsto \text{MSD}(t)$ **grows faster** than a linear function it is a **directed diffusion**, **slower** than the linear function it is a **confined diffusion**.

Mean Square Displacement Method

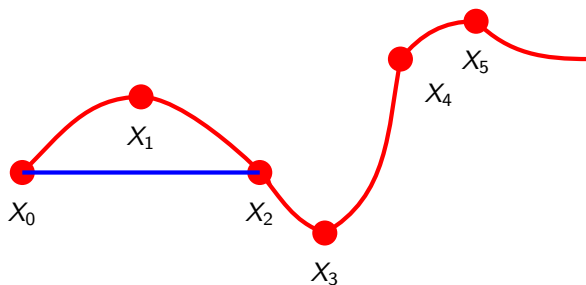
- We estimate $t \rightarrow \text{MSD}(t)$ from the trajectory X by :

$$\widehat{\text{MSD}}(r\Delta t) = \frac{1}{n-r} \sum_{i=1}^{n-r} \|X(t_{i+r}) - X(t_i)\|_2^2$$

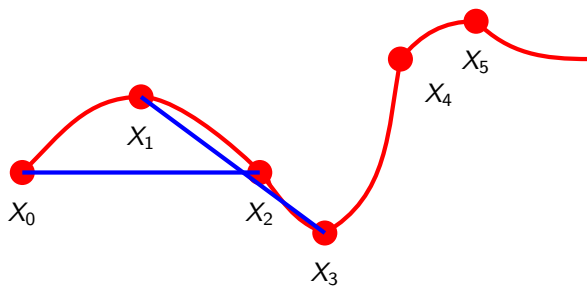
- This estimator is valid if the increments of the process are stationary.
- Not valid if μ depends on t .
- The terms of the average are not independent leading to non efficient estimation.
- Estimator at different lags non independent,
- The variance increases with increasing lag.



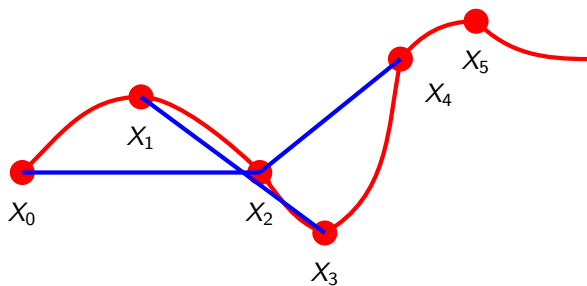
Estimation of $\text{MSD}(2\Delta t)$



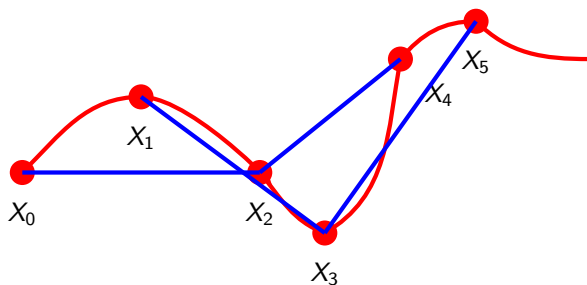
Estimation of $\text{MSD}(2\Delta t)$



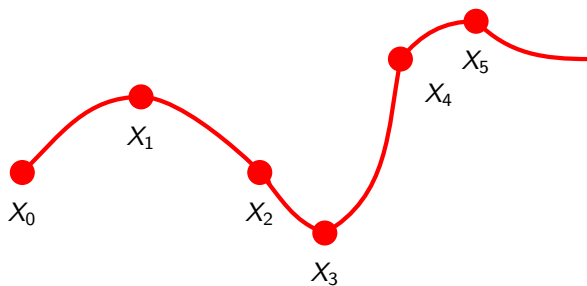
Estimation of $\text{MSD}(2\Delta t)$



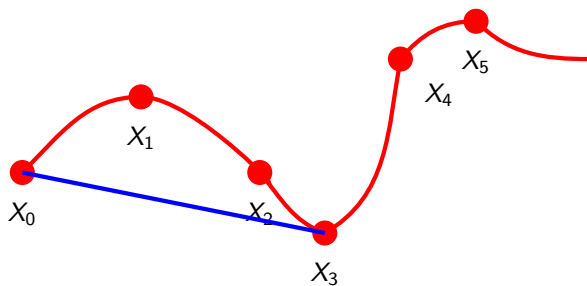
Estimation of $\text{MSD}(2\Delta t)$



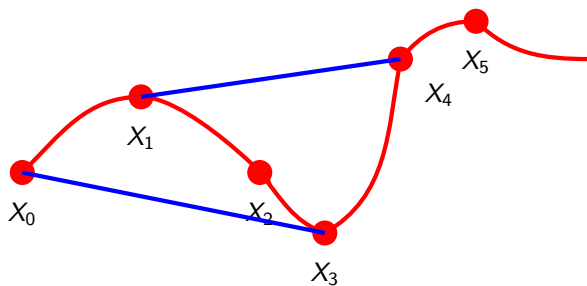
Estimation of $\text{MSD}(2\Delta t)$



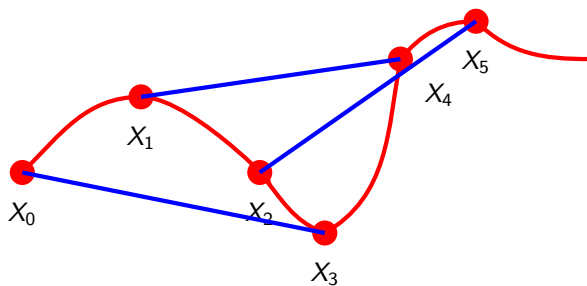
Estimation of $\text{MSD}(3\Delta t)$



Estimation of $\text{MSD}(3\Delta t)$



Estimation of $\text{MSD}(3\Delta t)$



Estimation of $\text{MSD}(3\Delta t)$

Classification with the MSD

Fit $t \rightarrow \widehat{\text{MSD}}(t)$ to $t \rightarrow Ct^\beta$.

Classify according the value of $\beta > 0$ [Feder et al., 1996]:

- ① $\beta < 0.1$: motionless.
- ② $0.1 < \beta < 0.9$: confined diffusion.
- ③ $0.9 < \beta < 1.1$: free diffusion.
- ④ $\beta > 1.1$: directed diffusion.

Empirical procedure: no underlying model.

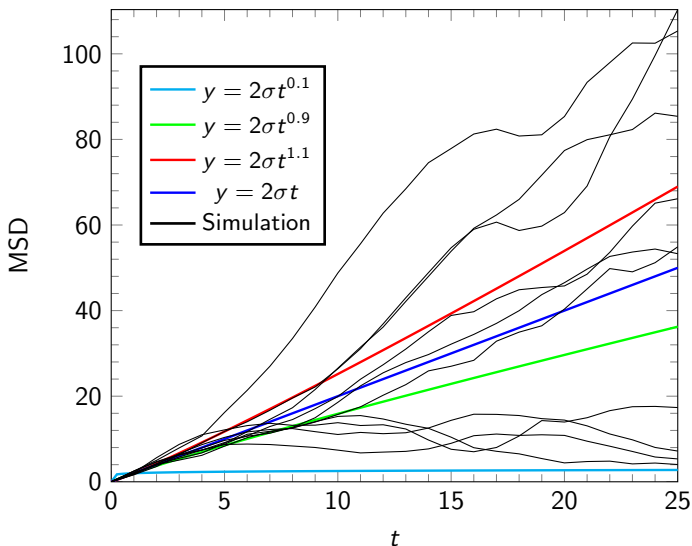


Figure: Classification of estimated MSD curves of 10 simulated ($\sigma = 1$) 2D Brownian trajectories.

Outline

- 1 Introduction
- 2 The MSD method
- 3 Statistical procedure**
- 4 Experimental results
- 5 Conclusion

Statistical test

- Our procedure can be seen as the **statistical test**:

H_0 "the observed trajectory X comes from a Brownian motion, $X_t = \sigma B_t$ "
versus

H_1 "the observed trajectory X comes from a confined or directed diffusion."

- Unlike a test which is binary, we split H_1 in 2 distinct outcomes.

Intuitive test statistic

- An intuitive statistic (or measure) to distinguish a **confined**/**directed** diffusion from **Brownian** motion:

$$S_n = \max_{i=0,\dots,n} \|X(t_i) - X(t_0)\|_2.$$

- This statistic allows to answer the question :
"How far from its initial position did the process go during the period of length $t_n - t_0$?"
- S_n low : the process stays close to X_{t_0} during $[t_0, t_n]$.
- S_n high : the process goes far from X_{t_0} during $[t_0, t_n]$.

Scaling of the test statistic

- Under H_0 the distribution of S_n depends on unknown parameter σ .
- We scale S_n as follows:

$$T_n = \frac{S_n}{\hat{\sigma} \sqrt{t_n - t_0}}$$

where $\hat{\sigma}$ is a consistent estimator of σ .

Lemma

Let $\hat{\sigma}$ a consistent estimator of σ such that the distribution of $\hat{\sigma}/\sigma$ does not depend on σ . Then, under H_0 , the distribution of T_n does not depend on σ .

- Then $q_n(\alpha)$ the quantile of order α of T_n does not depend on σ .

- The critical region for testing Brownian versus confined diffusion at level α :

$$\{T_n < q_n(\alpha)\}$$

- The critical region for testing Brownian versus directed diffusion at level α :

$$\{T_n > q_n(1 - \alpha)\}$$

- We aggregate these 2 tests and define the critical region :

$$\{T_n < q_n(\alpha/2)\} \cup \{T_n > q_n(1 - \alpha/2)\}$$

Consistency of the procedure

Theorem

Under H_0 , the test statistic T_n converges in distribution to $S_0 = \sup_{0 \leq s \leq 1} \|B_s\|_2$ as $n \rightarrow \infty$. Here $(B_t)_{t \geq 0}$ is the standard 2D Brownian motion that is the Brownian motion of variance \mathbf{I}_2 and initialization $B_0 = (0, 0)^T$.

The **cumulative distribution** of S_0 is :

$$P(S_0 \leq x) = \sum_{k=1}^{\infty} \frac{2e^{-j_{0,k}^2/2x^2}}{j_{0,k}J_1(j_{0,k})}$$

where $x \geq 0$, J_ν the Bessel function of order ν and $0 < j_{\nu,1} < j_{\nu,2} < \dots$ the positive zeros of J_ν .

Then $q_n(\alpha) \rightarrow q(\alpha)$ the quantile of order α of S_0 , as $n \rightarrow \infty$.

Decision rule

- For small n we have to use $q_n(\alpha)$ and not $q(\alpha)$.
- We estimate $q_n(\alpha)$ with Monte Carlo-simulations.
- We classify the motions according to the **decision rule**:
 - ① T_n outside in the critical region $\Rightarrow (X_t)_{t_0 \leq t \leq t_n}$ is **Brownian**.
 - ② $T_n < q_n(\alpha/2) \Rightarrow (X_t)_{t_0 \leq t \leq t_n}$ is a **confined diffusion**.
 - ③ $T_n > q_n(1 - \alpha/2) \Rightarrow (X_t)_{t_0 \leq t \leq t_n}$ is a **directed diffusion**.

Input: n, α, L

// the length n of the trajectory

// the probability $\alpha \in (0, 1)$

// the number N of Monte Carlo experiments

Result: $q_n^{(N)}(\alpha)$.

for $i=1$ **to** N **do**

 // Simulation of a Brownian trajectory of size n , of variance
 $\sigma = 1$ and with resolution time $\Delta t = 1$.

 initialization $Y_0^{(i)} = (0, 0)^\top$;

for $j=1$ **to** n **do**

 Draw $\epsilon \sim \mathcal{N}(0, \mathbf{I}_2)$;

$Y_j^{(i)} = Y_{j-1}^{(i)} + \epsilon$;

end

 // Computation of the test statistic

 Compute the ratio $T_n^{(i)} = S_n^{(i)} / \hat{\sigma}_n^{(i)}$ from $(Y_0^{(i)}, \dots, Y_n^{(i)})$;

end

Algorithm 1: Simulation of a N -sample $(T_n^{(1)}, \dots, T_n^{(N)})$ of the distribution of the statistic T_n under H_0 .

Estimation of the diffusion coefficient

- A first estimator of σ^2 :

$$\hat{\sigma}_1^2 = \frac{1}{2n\Delta t} \sum_{j=1}^n \|X_{j\Delta t} - X_{(j-1)\Delta t}\|_2^2.$$

- Consistent under H_0 but can be biased under H_1 .
- A non-parametric estimator as an alternative for constant drift over time :

$$\hat{\sigma}_{np}^2 = \frac{1}{2n\Delta t} \sum_{j=1}^n \|X_{j\Delta t} - X_{(j-1)\Delta t} - \hat{\mu}(X_{(j-1)\Delta t})\|_2^2.$$

$$\text{where } \hat{\mu}(x) = \frac{\sum_{j=1}^n K\left(\frac{X_{j\Delta t} - x}{h}\right)(X_{(j+1)\Delta t} - X_{j\Delta t})}{\sum_{j=1}^n K\left(\frac{X_{j\Delta t} - x}{h}\right)}, \text{ } K \text{ a kernel density, } h \text{ the bandwidth.}$$

Outline

- 1 Introduction
- 2 The MSD method
- 3 Statistical procedure
- 4 Experimental results**
- 5 Conclusion

Alternative parametric hypothesis

- **Evaluation** of our test procedure with two parametric diffusion processes.
- For illustrating directed diffusion, we use the **directed Brownian** :

$$dX_t^i = v_i dt + \sigma dB_t^i \quad i = 1, 2,$$

- For illustrating confined diffusion, we use the **Ornstein-Uhlenbeck** process:

$$dX_t^i = -\lambda(X_t - \theta_i)dt + \sigma dB_t^i \quad i = 1, 2,$$

- Power of the test $P(\text{accept } H_1 \mid H_1)$
- We can show that the power depends on:
 - 1 $\|v\| \sqrt{\Delta t} / \sigma$ for **directed Brownian**
 - 2 $\lambda \Delta t$ for **Ornstein-Uhlenbeck**

Power curves

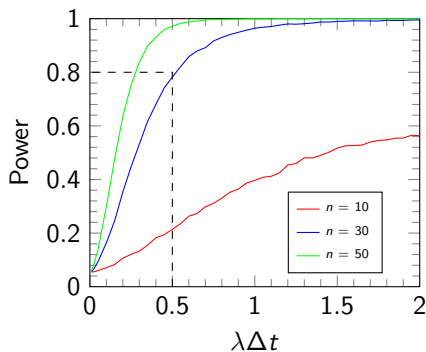
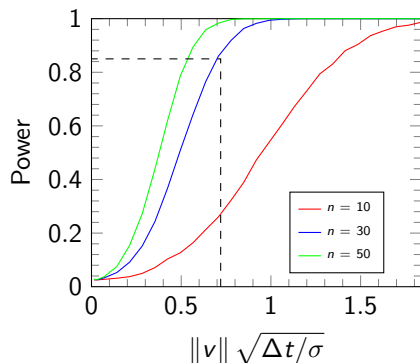


Figure: Power of the test in the **directed Brownian** case (left) and in the **Ornstein-Uhlenbeck** case (right).

Simulations

- We simulate trajectories of size $n = 30$.
- We take $\|v\| \sqrt{\Delta t} / \sigma = 0.72$. For instance $\Delta t = 0.1s$ $\sigma^2 = 4\mu m^2.s^{-1}$ and $\|v\| = 4.6\mu m.s^{-1}$.
- We take $\lambda \Delta t = 0.5$. For instance $\Delta t = 0.1s$ $\lambda = 5s^{-1}$.

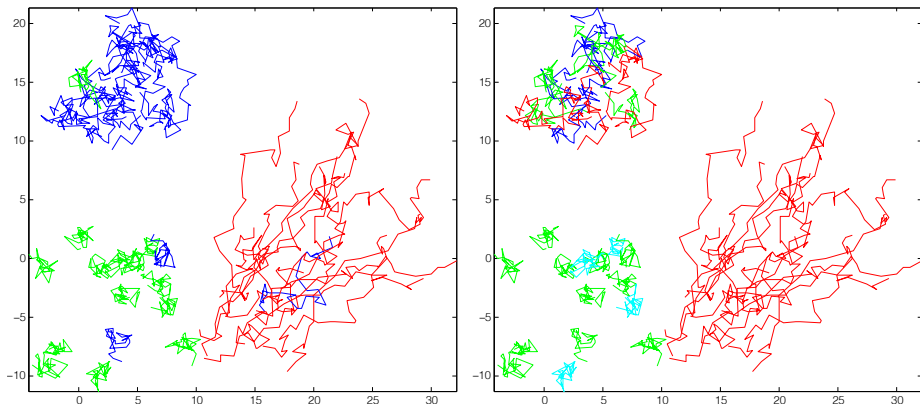


Figure: Simulated trajectories of Brownian, directed Brownian, Ornstein-Uhlenbeck.

Left panel: classification with our test at level $\alpha = 5\%$.

Right panel classification with the MSD method.

In cyan trajectories labelled as Motionless.

Test label	BR	DBR	OU	BR	DBR	OU
Ground truth	without noise			with noise		
BR	94.6	3	2.7	94.2	1.3	4.5
DBR	12.7	87.3	0	19.7	80.3	0
OU	26.6	0	73.4	19.8	0	80.2

Table: Confusion matrices of our test. $N = 10000$ simulated trajectories of each processes were used with parameters defined above. BR stands for **Brownian**, OU for **Ornstein-Uhlenbeck** and DBR for **directed Brownian**. Results are written in %. We read 0.1% of the simulated **Brownian** trajectories with noise are labelled as **directed Brownian** by our test. For the noisy case we set $\sigma_{err} = 0.2$ which gives $SNR=1$.

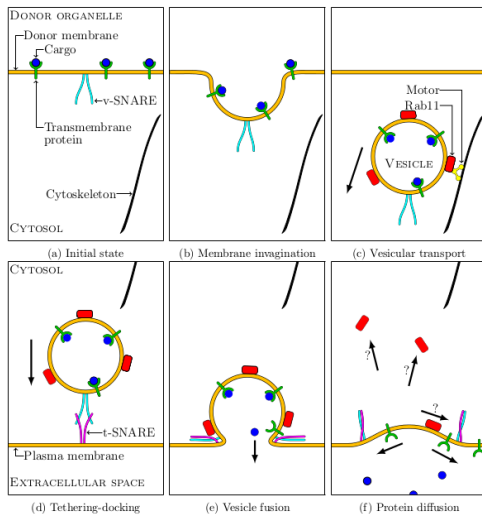


Figure: Main steps of exocytosis

Our test captures the **dominant motion** of Rab11.

Real sequences

- Sequences of fluorescent images, **TIRF microscopy**, depicting the traffic of **Rab11 protein**.
- 600 frames of size 256×240 (1 pixel=160nm) acquired with $\Delta t = 0.1s$
- Trajectories computed with the ICY tracker[De Chaumont et al., 2012].
- **Pre-processing of the data**: we keep only the particles moving enough.

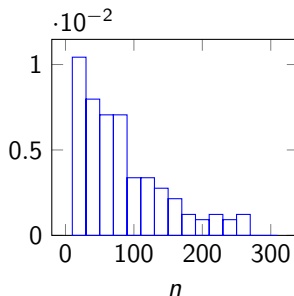


Figure: Histogram of the length of Rab11 trajectories

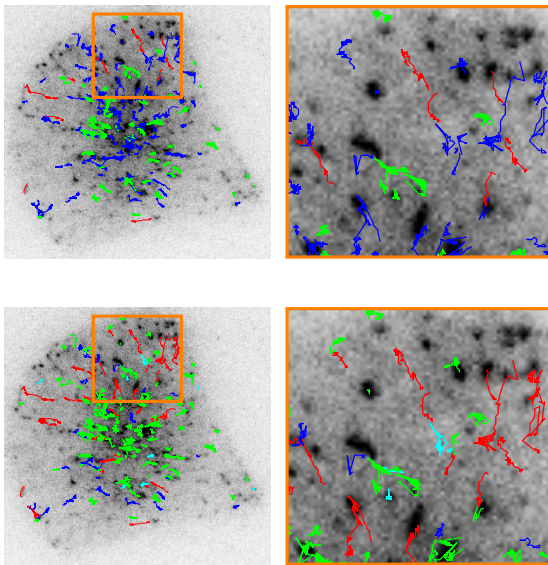


Figure: Labelling of the trajectories on the Rab11 protein sequence. Top panel with our test, bottom panel with the MSD method.

Method	Test	MSD
Label		
Brownian	61	14
Confined	32	59
Directed	7	20
Motionless	0	4 - 3

Table: Results (in %) of the classification of trajectories dynamics on the sequences of Rab11 protein with our test and the MSD method.

Outline

- 1 Introduction
- 2 The MSD method
- 3 Statistical procedure
- 4 Experimental results
- 5 Conclusion**

Conclusion

- Our test procedure allows to classify particles motions within cell.
- Our test is statistically **consistent**.
- It is an alternative to the MSD method showing more reliable results from Monte Carlo simulations.
- Future work: **detection of change of motion dynamic** over time.

Thank you for your attention!

Estimation of $q_n(x)$

Estimated quantiles	Trajectory size			
quantile order	10	30	100	asyp
2.5%	0.725	0.754	0.785	0.834
97.5%	2.626	2.794	2.873	2.940

Table: Estimation of $q_n(\alpha/2)$ and $q_n(1 - \alpha/2)$ ($\alpha = 5\%$) for different size trajectory n , using Monte Carlo simulations with $N = 1\,000\,001$ replications. Estimations are accurate at ± 0.001 .

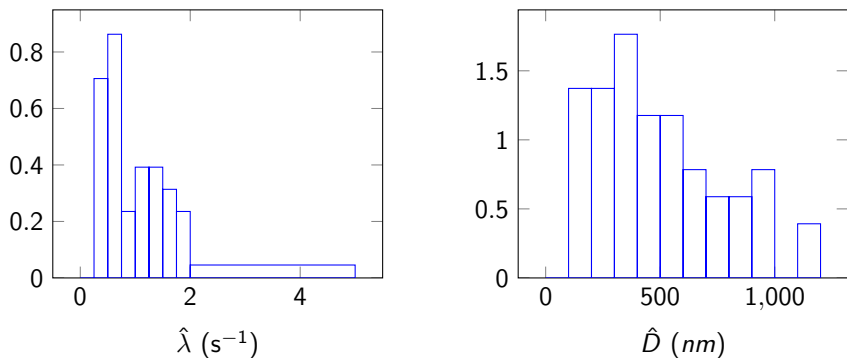


Figure: Histograms of $\hat{\lambda}$ (Method of Moment estimator) left, and \hat{D} for the trajectories labelled as **Ornstein-Uhlenbeck** by our test at level 5%.

Reading note: D is the radius of the confinement area defined as $P(\|X_t - \theta\| > D) = 0.05$.

For **directed Brownian**, $\text{Mean}(\|\hat{v}\|) = 468 \text{ nm}$ on 11 trajectories.

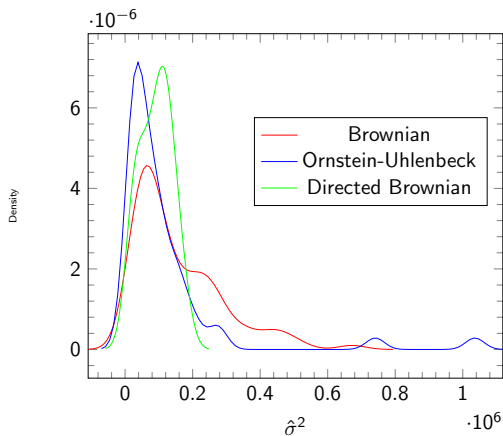


Figure: Estimated distribution of $\hat{\sigma}^2$ (MLE) of for the different processes on the Rab11 sequence.