Statistical signature of interactions in heterogeneous cellular environments

Hippolyte Verdier, François Laurent, Maxime Duval, Christian L. Vestergaard, Jean-Baptiste Masson

hverdier@pasteur.fr, flaurent@pasteur.fr,mduval@pasteur.fr

clvestergaard@pasteur.fr, jbmasson@pasteur.fr

Decision and Bayesian Computation - Épiméthée (soon)
Computational Biology Department, CNRS USR 3756
Neuroscience Department, CNRS UMR 3571
INRIA (soon)
Institut Pasteur
Institut Prairie

October 28, 2021

Overview

- Problem statement
- The mapping hypothesis
- Method
  - Simulation-based inference
  - GRATIN : Graphs on Trajectories for Inference
- Results
  - On simulated data
  - Application to experimental data
- Conclusion & perspectives

Overview

- Problem statement
- The mapping hypothesis
- Method
  - Simulation-based inference
  - GRATIN : Graphs on Trajectories for Inference
- Results
  - On simulated data
  - Application to experimental data
- Conclusion & perspectives

Random walks at the membrane

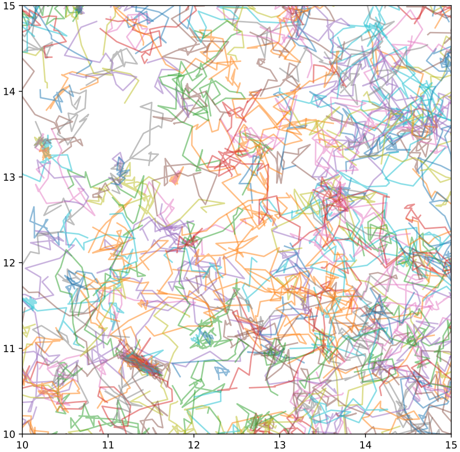


Figure 1: Trajectories of Gag, 2 minutes of observation[1]

[1] Data acquired by C. Favard — Floderer C. et al, . Sci Rep. 2018 Nov 2;8(1):16283

Overview

- Problem statement
- The mapping hypothesis
- Method
  - Simulation-based inference
  - GRATIN : Graphs on Trajectories for Inference
- Results
  - On simulated data
  - Application to experimental data
- Conclusion & perspectives

Problem: The Mapping Hypothesis

From a set $\mathcal{T} = \{\boldsymbol{r}_t^i | i \in [0, m_{max}], t \in [0, T_{max}]\}$ of localisations within a bounded domain $\mathcal{D} \in \mathbb{R}^l$ with $l \leq 3$, with an underlying model written as

$$d\boldsymbol{r}_t^i = \boldsymbol{a}_t\left(\boldsymbol{r}_t^i\right) dt + \boldsymbol{b}_t\left(\boldsymbol{r}_t^i\right) \circ dW(t)$$

we seek[2]

- A probabilistic assignment $\mathcal{S} = \{\sigma(t)_i^j \,|\forall (i,j) \in \mathcal{L}\}$ associated to $P_i^j\left(\boldsymbol{r}_t^i, \boldsymbol{r}_{t+\Delta t}^j | \boldsymbol{\theta}\right)$ between particle between time t and $t + \Delta t$ with $\boldsymbol{\theta}$ the set of parameters.
- A self organising mesh $\mathcal{M}$
- A set of maps $\mathcal{M}\{\dashv, \lfloor\} = \{(\boldsymbol{a}_t(\boldsymbol{r}), \boldsymbol{b}_t(\boldsymbol{r})) \,| \boldsymbol{r} \in \mathcal{D}\}$

Overdamped Langevin equation

$$\frac{d\boldsymbol{r}}{dt} = D_t(\boldsymbol{r})\left(\boldsymbol{f}_t(\boldsymbol{r}) + \lambda \frac{\nabla D_t(\boldsymbol{r})}{D_t(\boldsymbol{r})}\right) + \sqrt{2D_t(\boldsymbol{r})}\xi(t)$$

[2] **A. Serov et al., Phys. Rep., 2020 Mar 2;10(1):3783.** .

The mapping hypothesis

Bayesian Inference

$$P(U|T) = \frac{\overbrace{P(T|U)}^{\text{Likelihood}}\,\overbrace{\pi(U)}^{\textit{Prior}}}{\underbrace{P(T)}_{\text{Evidence}}}$$

Likelihood: solving the Fokker-Planck equation

$$\frac{\partial \overbrace{P(\boldsymbol{r}, t|\boldsymbol{r}_0, t_0)}^{\text{Likelihood}}}{\partial t} = -\nabla\left[\left(D(\boldsymbol{r})\,\boldsymbol{f}(\boldsymbol{r}) + \lambda \nabla D(\boldsymbol{r})\right) P(\boldsymbol{r}, t|\boldsymbol{r}_0, t_0)\right] + \nabla\left[D(\boldsymbol{r})\,\nabla P(\boldsymbol{r}, t|\boldsymbol{r}_0, t_0)\right]$$

Prior: Physics of the environment

$$\pi(D) \propto \exp\left(-\int d^2\boldsymbol{r}\mu_{\boldsymbol{r}}|\nabla D(\boldsymbol{r},t)|^2 + \mu_t\left(\dot{D}(\boldsymbol{r},t)\right)^2\right)$$

$$\pi(\mathbf{f}) \propto \exp\left(-\int d^2\boldsymbol{r}\lambda_{\boldsymbol{r}}|\nabla \mathbf{f}(\mathbf{r},\mathbf{t})|^{\mathbf{2}} + \lambda_{\mathbf{t}}\left(\dot{\mathbf{f}}(\mathbf{r},\mathbf{t})\right)^{\mathbf{2}}\right)$$

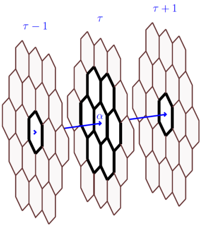


Figure 2: Space time inference

The mapping hypothesis (cont.)

Physics-informed stochastic optimization[3]

- Perform in parallel
- Randomly select a local domain (α, τ) with $d(\alpha, \alpha') > d_s(\mu_r, \lambda_r)$ and $d(\tau, \tau') > d_\tau(\mu_\tau, \lambda_\tau)$
- Sample a minibatch $\Delta \boldsymbol{r}_{\mathcal{B}_{\alpha,\tau}} = \cup_{(\alpha',\tau') \in \mathcal{B}_{\alpha',\tau'}}$ in the neighbourhood of (α, τ)
- update $\theta_{\alpha,\tau}^{(k)} = \theta_{\alpha,\tau}^{(k-1)} + \Delta\theta\left(\Delta \boldsymbol{r}_{\mathcal{B}_{\alpha,\tau}}, \theta_{\mathcal{S}_{\alpha,\tau}}^{(k-1)}\right)$
- approximate local posterior:

$$f_{\alpha,\tau}\left(\theta_{\mathcal{B}_{\alpha,\tau}}\right) = -\log p\left(\Delta \boldsymbol{r}_{\alpha,\tau} | \theta_{\mathcal{R}_{\alpha,\tau}}\right) + \mu_{\boldsymbol{r}} q_\alpha\left(D_{\mathcal{R}_{\alpha,\tau}}\right) + \mu_t q_{\mathcal{T}}\left(D_{\alpha,\mathcal{T}_\tau}\right) + \ldots$$

The mapping hypothesis (cont.)

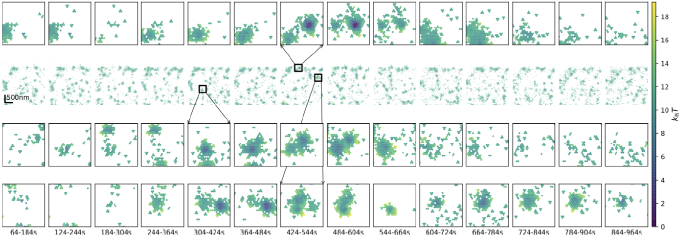
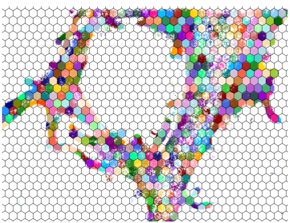


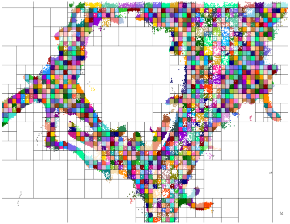
Figure 3: Space time inference of transient Virion assembly. 100k parameters

[3] **F. Laurent et al.**, **Phys Biol**, **17**, **015 003**, **2019**.

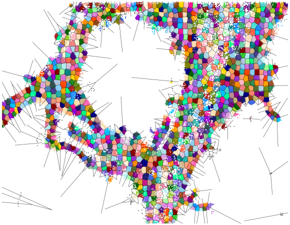
Example segmentations



(a) Hexagonal



(b) Quad-tree



(c) k-Means

Example segmentations (cont.)

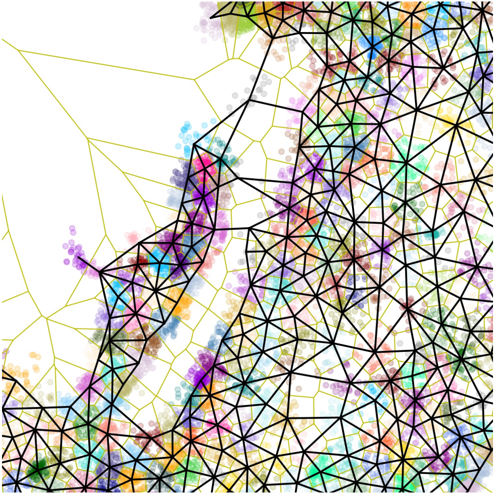
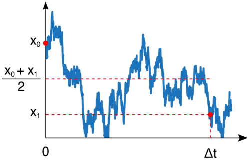


Figure 5: *Growing-When-Required*-based tessellation

Stochastic integrals

$$dX_t = \underbrace{a(X_t)}_{\text{drift}} dt + \underbrace{b(X_t)}_{\text{diffusion}} \circ dW(t)$$



We integrate from any $x \in [x_0; x_1]$ with $0 \leq \lambda \leq 1$

$$x = (1-\lambda)\, x_0 + \lambda x_1$$

$$\mathbb{E}(\Delta x) = a(x_0)\,\Delta t + \underbrace{\lambda b(x_0)\, b'(x_0)}_{\text{diffusion gradient}}\,\Delta t + \mathrm{O}\left(\Delta t^2\right)$$

Bayesian evidence analysis

- H_0 : Heterogeneous diffusion environment
- H_1 : Heterogeneous diffusion environment with active forces.

$$B_{1,0} \equiv \frac{P\left(x|M_1\right)}{P\left(x|M_0\right)}$$

$$B_{1,0} = \eta^d \frac{\int_0^1 d\lambda \left[\nu + \eta^2 \left(\xi_t - \lambda\xi_{sp}\right)^2\right]^{-p}}{\int_0^1 d\lambda \left[\nu + \left(\xi_t - \lambda\xi_{sp}\right)^2\right]^{-p}}$$

$\xi_{sp} \equiv \frac{\nabla \boldsymbol{b} \Delta t}{\sqrt{V}}$: SNR spurious force

$\xi_t \equiv \frac{\overline{\Delta \boldsymbol{r}}}{\sqrt{V}}$: SNR for total force

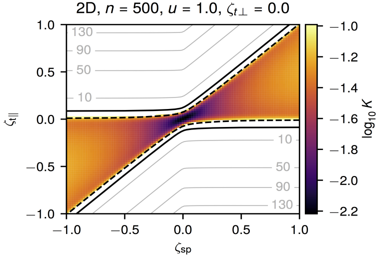
$V = \overline{\left(\Delta \boldsymbol{r} - \overline{\Delta \boldsymbol{r}}\right)^2}$: one-jump variance

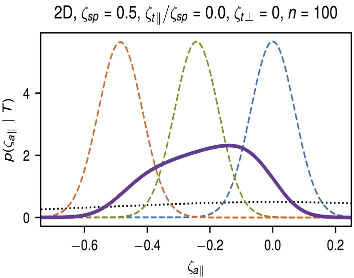
$\nu = 1 - \frac{n_\pi V_\pi}{nV}$: ratio of jump variances

$\eta = \sqrt{\frac{n_\pi}{n+n_\pi}}$: normalized # points

$p\left(d\right) = \frac{d(n+n_\pi-1)}{2} - 1$: exponent

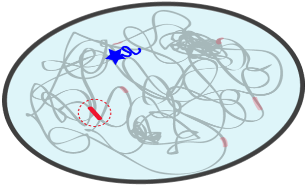
Bayesian Inference

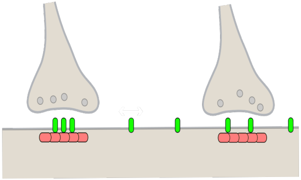




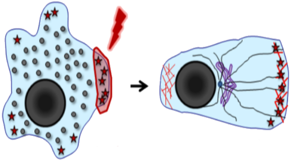
Versatile Applications

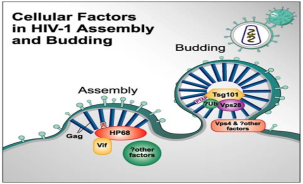
CRISPR-Cas9 dynamics[4]





Cell Motility[5]





Receptor-Scaffold Interactions[6]

Virion Assembly[7]

[4] “S .C. Knight et al., Science, 350, 823-826, 2015. T. Blanc et al., Nat Methods, 17, 1100-1102, 2020.

[5] A. Remorino et al., Cell Rep, 21, 1922-1935, 2017. M. El Beheiry, M. Dahan & J. B. Masson, Nat Methods, 12, 594-595, 2015.

[6] J. B. Masson et al., Biophys J, 106, 74-83, 2014. S. Turkcan & J.-B. Masson, PLOS ONE, 8, e82799, 2013.

[7] C. Floderer et al., Sci Rep, 8, 17 426, 2018. A.S. Serov et al., Sci Rep, 10, 3783, 2020.

TRamWAy: parallel Python software for random walk analysis

- Based on inferenceMAP[8]
- Non-tracking with Belief Propagation[9]
- Inference performed on multiple meshes in space and time[10]
- Mapping of biophysical properties on cell[11]
- Graph Neural Network approach to models of random walks[12]
- $\sim$ 60 000 lines

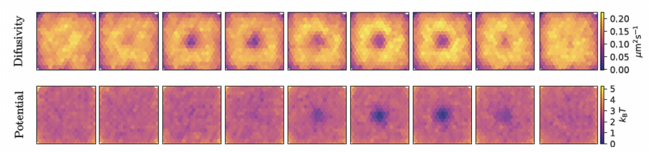


Figure 6

[8] M. El Beheiry, M. Dahan & J. B. Masson, Nat Methods, 12, 594-595, 2015.
[9] C. Vestergaard et al, In preparation.
[10] F. Laurent et al., Phys Biol, 17, 015 003, 2019.
[11] C. Floderer et al., Sci Rep, 8, 17 426, 2018. A.S. Serov et al., Sci Rep, 10, 3783, 2020.
[12] H. Verdier et al, 2021 arXiv:2103.11738.

Anomalous diffusion

Mean Square Displacement

$$\langle (r_t - r_0)^2 \rangle \propto t^{\alpha} \ , \ 0 \leq \alpha \leq 2, \ \alpha \neq 1$$

Mean Square Displacement

$$\langle (r_t - r_0)^2 \rangle \propto t^{\alpha}\ ,\ 0 \leq \alpha \leq 2,\ \alpha \neq 1$$

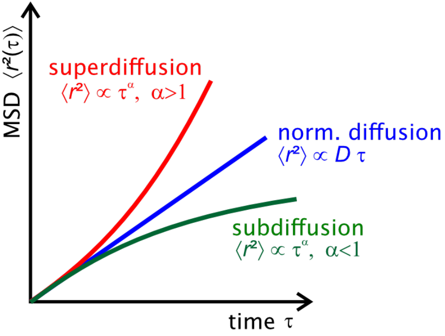
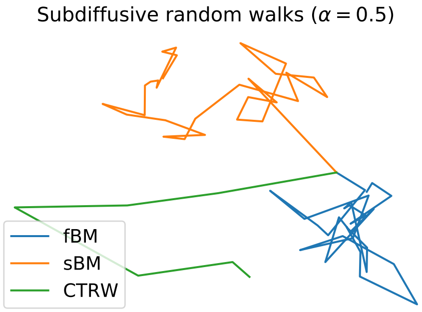


Figure 7: MSD scaling (Wikipedia)

Anomalous diffusion

$\alpha < 1$: Subdiffusion



Anomalous diffusion (cont.)

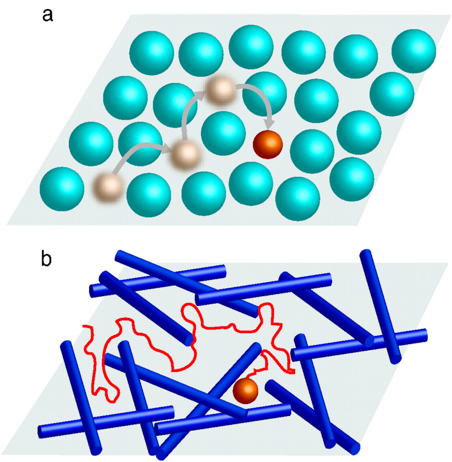
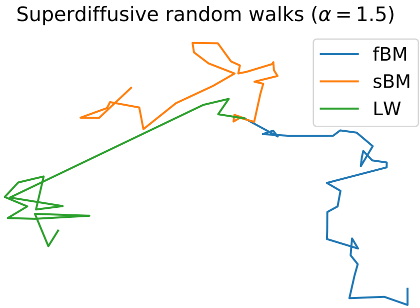


Figure 8: Origins of subdiffusion
Condamin, Tejedor, Voituriez, *et al.* [1]

Anomalous diffusion (cont.)

$\alpha > 1$: Superdiffusion



Inverse problem

From a trajectory $\mathbf{r}_{1:T}$, infer relevant parameters :

- **Motion identity** m, through probabilities of belonging to each class $\hat{p} = (\hat{p}(m = 1), \ldots \hat{p}(m = k))$
- **Anomalous exponent** α
- Intensity of **drift**
- Intensity of **confinement**

Challenges

- No analytic likelihood in general
- Highly stochastic processes
- Ability to generalize

Overview

- Problem statement
- The mapping hypothesis
- Method
  - Simulation-based inference
  - GRATIN : Graphs on Trajectories for Inference
- Results
  - On simulated data
  - Application to experimental data
- Conclusion & perspectives

Simulation-based inference

- Numerous physical models explored with simulations
- These simulations are poorly suited for inference
- ABC: historical approach
- New initiatives stemming from statistical learning[13]
- Amortised likelihood approach

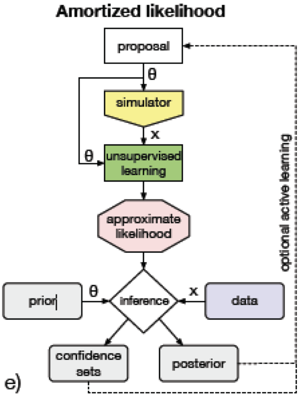


Figure 9: derived from [1]

[1] K. Cranmer et al., PNAS 117, 30055–30062 (2020)

Simulation-based inference

1. Simulate a diversity of trajectories

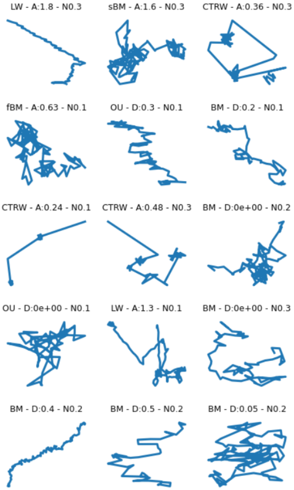


Figure 10: Simulated trajectories for training

Simulation-based inference

1. Simulate a diversity of trajectories
2. Train a model to infer parameters from these trajectories

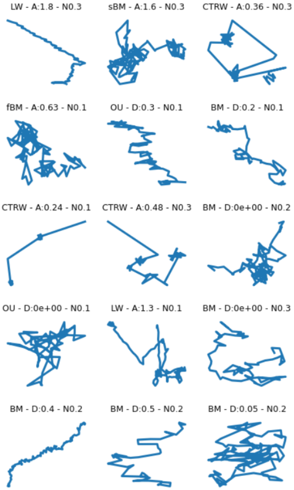


Figure 10: Simulated trajectories for training

Simulation-based inference

1. Simulate a diversity of trajectories
2. Train a model to infer parameters from these trajectories
3. Perform inference on experimental observations !

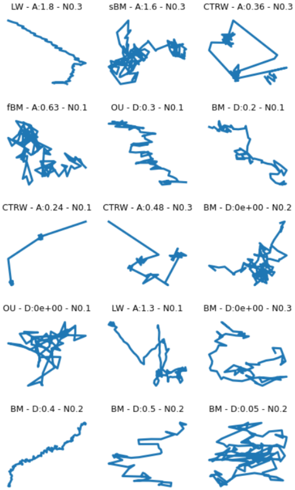
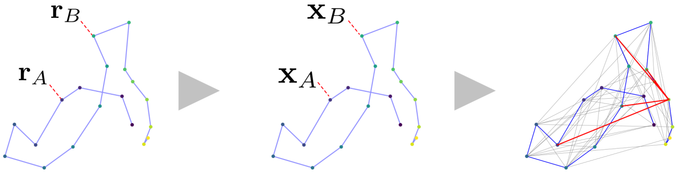


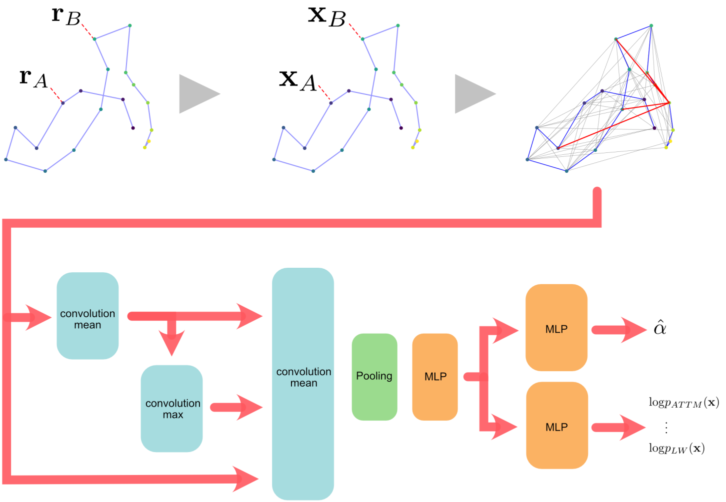
Figure 10: Simulated trajectories for training



$\mathbf{r}_B$
$\mathbf{r}_A$
$\mathbf{x}_B$
$\mathbf{x}_A$

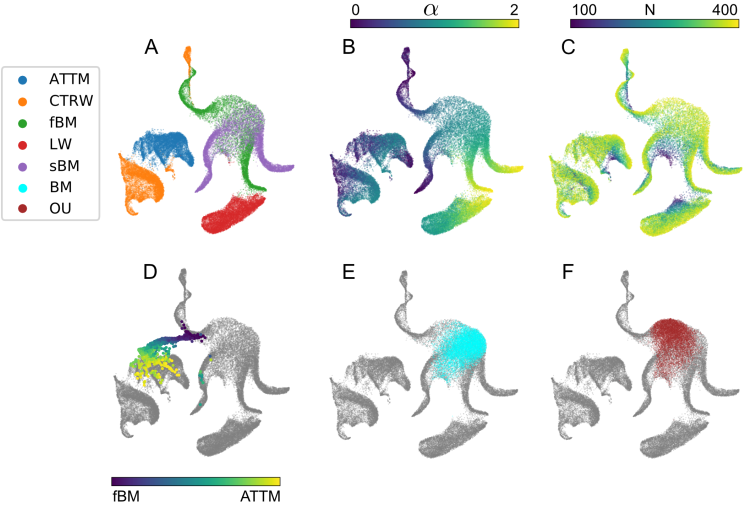
GRATIN[14]

Graphs on Trajectories for Inference



[14] H. Verdier et al, J. Phys. A. (2021)

Latent representation of random walks[15]



[15] **H. Verdier et al, J Phys A: Math. Theor. 2021**.

A new vision of synapses

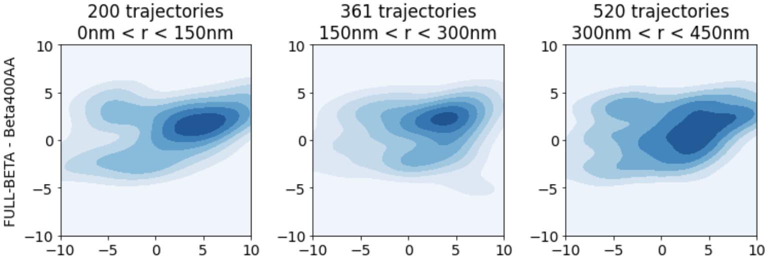


Figure 11: Evolution of distribution of position in the latent space as a function of radius within the synapses

Overview

- Problem statement
- The mapping hypothesis
- Method
  - Simulation-based inference
  - GRATIN : Graphs on Trajectories for Inference
- Results
  - On simulated data
  - Application to experimental data
- Conclusion & perspectives

Results on simulated data

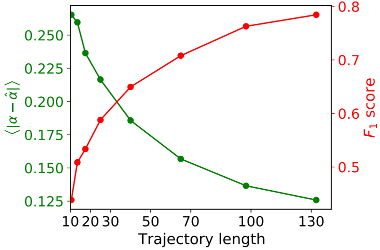


Figure 12: Performance : Regression of α and classification
Noise level equivalent to PALM conditions

Results on simulated data

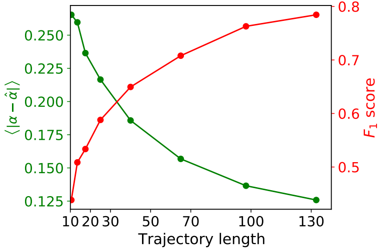


Figure 12: Performance : Regression of α and classification
Noise level equivalent to PALM conditions

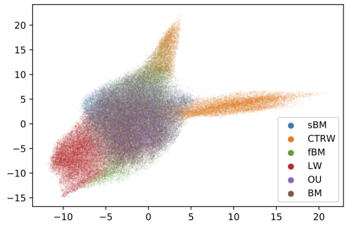


Figure 13: Latent space, colored by motion type
One point = one trajectory
$10 \leq L \leq 30$

Illustration on experimental data

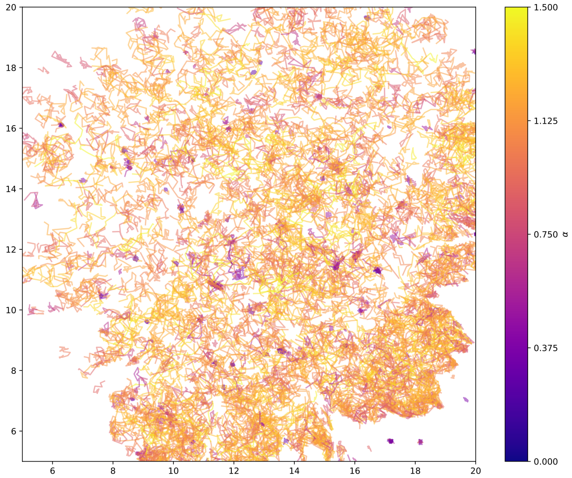


Figure 14: Trajectories of CAAX Gag [16]

[16] Data acquired by C. Favard — Floderer C. et al, . Sci Rep. 2018 Nov 2;8(1):16283

Gag interacts more than GT46 and Tetherin [17]

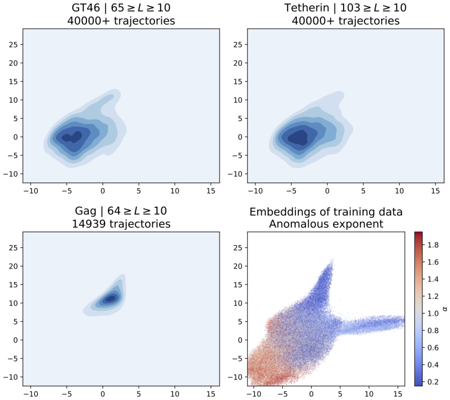


Figure 15: Latent representations

[17] Analysis made on data acquired by P. Sengupta

Gag interacts more than GT46 and Tetherin [17]

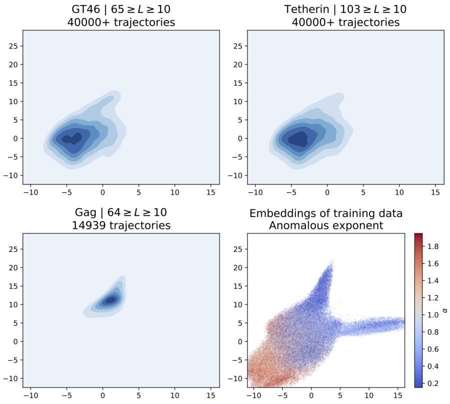


Figure 15: Latent representations

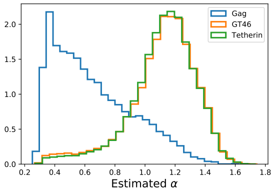


Figure 16: Estimations of α

[17] Analysis made on data acquired by P. Sengupta

Overview

- Problem statement
- The mapping hypothesis
- Method
  - Simulation-based inference
  - GRATIN : Graphs on Trajectories for Inference
- Results
  - On simulated data
  - Application to experimental data
- Conclusion & perspectives

Conclusion & perspectives

Conclusions :

- Spatio-temporal mapping
- individual RW model inference

Perspectives :

- Unsupervised learning
- statistical testing in latent space

Recent Funding & Miscellaneous

Funding

- ANR: TramWAy, SiNCoBe
- European: Marie-Curie, EMBL-EU.
- Other: Ile de France, DIM-Elicit, Inception, SESAME PIA 2018, PSL, Janelia, ACIP
- Sponsorship : CRPCEN, Gilead Science, EDF, Institut Curie, PSL
- Companies : Avatar Medical, Sanofi, NVIDIA

Visiting positions

- Janelia Research Campus (JBM)
- Cambridge MRC (JBM, FL)
- EMBL (JBM)
- CPT Marseille (CLV)