A machine learning approach to distinguish different subdiffusive dynamics in particle tracking

Giacomo Nardi, Matheus Santos Sano, Anne Brelot, Jean-Christophe Olivo-Marin, Thibault Lagache

▶ **To cite this version:**

Giacomo Nardi, Matheus Santos Sano, Anne Brelot, Jean-Christophe Olivo-Marin, Thibault Lagache. A machine learning approach to distinguish different subdiffusive dynamics in particle tracking. International Symposium on Biomedical Imaging (ISBI), IEEE, May 2024, Athens (Greece), Greece. pasteur-04626746v2

HAL Id: pasteur-04626746

https://pasteur.hal.science/pasteur-04626746v2

Submitted on 16 Aug 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.

MOTION CLASSIFICATION BASED ON GEOMETRICAL FEATURES OF TRAJECTORIES

M. Santos Sano⋆, A. Brelot†, J.-C. Olivo-Marin⋆, T. Lagache⋆, G. Nardi⋆

⋆ Institut Pasteur, Universite de Paris Cité, CNRS UMR 3691, BioImage Analysis Unit, Paris, France

† Institut Pasteur, Université de Paris Cité, CNRS UMR 3691, Dynamics of Host-Pathogen Interactions Unit, Paris, France

Corresponding author: giacomo.nardi@pasteur.fr

ABSTRACT

This paper proposes a novel approach for motion classification based on geometrical features computed on trajectories. The method follows a machine learning approach trained and validated on synthetic datasets simulating several stochastic models. The resulting model enables, in particular, the recognition of different subdiffusive behaviors, offering a finer classification than the standard method based on mean square displacement. The method is assessed on a biological dataset containing trajectories of CCR5 cell receptors.

***Index Terms*—** Motion classification, subdiffusive processes, machine learning, geometric features.

1. INTRODUCTION

Fluorescence microscopy coupled with tracking algorithms enables the dynamic observation of different cellular particles (like proteins or receptors). Motion classification represents an essential task to understand their dynamic and the influence of drugs on them. The classical approach for motion classification is based on the polynomial fitting of the function of mean square displacement. This function accounts for the variance (in time) of time-constant displacements along the trajectory. Based on its linearity for Brownian motion, a criterion has been established for classifying motion into three large categories: subdiffusive, Brownian, and superdiffusive [1]. An alternative criterion has been proposed in [2, 3] via a hypothesis test on the maximal distance from the starting point of the trajectory. This method is more suited to short trajectories but still performs a classification into the three classes cited above. However, a three-class classification approach can be very restrictive and unrealistic for many biological applications. In reality, at the subcellular scale, many phenomena reveal coexisting subdiffusive dynamics with intrinsically different behaviors corresponding to different environmental constraints. Then, a finer description of subdiffusive motions is a primary challenge for classification algorithms.

Some methods have already been proposed to characterize several subdiffusive behaviors. [4, 5] introduce mathematical quantifiers differentiating generalized diffusion (Fractional Brownian Motion, FBM) from motion with trapping (Continuous Time Random Walk, CTRW). [6] explains how to combine statistical tests to distinguish between confined (Ornstein-Uhlenbeck process, OU) or FBM, Brownian (BM), and directed (DIR) trajectories. Finally, [7] develops a machine learning method that distinguishes confined, anomalous, and diffusive motions. Although these works propose *ad-hoc* features for specific subdiffusive processes, no work establishes, to our knowledge, a unified method to distinguish all main types of motion widely studied in the literature.

This work proposes a novel unified approach for motion classification. We create a large dataset of trajectories simulating the principal processes (BM, OU, DIR, FBM, CTRW) and associate each with geometric features. Then, we train a machine learning supervised method to learn the original stochastic model based on the features defined on the trajectories. Features consider the angle between successive displacements along the trajectory or Ripley's indices on growing balls around the starting points of trajectories. These features discriminate confined or trapped motion from more diffusive paths via a five-class classification method. Compared to related methods, the proposed approach works with a more prominent family of movements based on fewer features, representing its main advantage regarding explicability and reliability. After validation of synthetic data, the method is assessed on a biological dataset of trajectories of cell receptors (CCR5) obtained from fluorescence microscopy videos via a tracking algorithm.

2. MOTION MODELS

This section presents the main stochastic processes commonly used to model particle dynamics. We also recall the standard criterion for motion classification; this enables the classification of motion into three main classes (subdiffusive, Brownian, and superdiffusive) based on the mean square displacement (MSD) function.

We define a trajectory as a set of successive positions over time $\mathbf{X} = (X_{t_1}, ..., X_{t_N})$, where $X_i \in \mathbb{R}^2$ with independent components and the time interval between successive positions is constant. Discretizing 2-dimensional contin-

uous stochastic processes can generate examples of trajectories with different dynamical behavior.

2.1. Brownian motion and superdiffusion

Brownian motion (BM) is a Gaussian stochastic process, denoted by B_t, with stationary and independent increments verifying $(B_t - B_s) \sim \mathcal{N}(0, (t-s)\mathbf{I}_2)$ for every $t > s$ where $\mathbf{I}_2$ denotes the 2-dimensional identity matrix. Brownian motion describes particles freely moving in the space whose displacement amplitude can be amplified by a factor σ, called the diffusion coefficient, considering the process $X_t = \sigma B_t$.

Brownian motion also helps describe objects driven by active motors. The so-called Directed Brownian (or directed) motion (DIR) verifies

$$dX_t = \mu dt + \sigma dB_t$$

where the drift component $\mu \in \mathbb{R}^2$ gives a constantly oriented input to the motion. Of course, depending on the ratio $\|\mu\|/\sigma$ the trajectory will have a more linear ($\|\mu\| >> \sigma$) or Brownian ($\|\mu\| << \sigma$) behavior.

2.2. Subdiffusive processes

This section introduces the principal stochastic processes modeling different subdiffusive behaviors. These types of motion reveal different environmental interactions (like trapping or viscoelastic constraints) and help understand the influence of exterior agents on particle mobility.

Ornstein-Uhlenbeck process (OU). This process suits particles with limited mobility due to an external force attracting them toward an equilibrium point. The continuous process verifies

$$dX_t = -\lambda(X_t - \theta)dt + \sigma dB_t \tag{1}$$

where θ is the equilibrium point and λ represent the drift term. In the following, we consider $\theta = (0,0)$. We note that if $\lambda << \sigma$, the motion will present a Brownian behavior, while, for larger λ, we obtain more confined trajectories.

Fractional Brownian Motion (FBM). It is a generalization of Brownian motion characterized by Gaussian increments that are stationary but not independent. This enables the modeling of particles moving in constrained or crowded environments. The correlation between increments is given by

$$E[(X_t - X_s)^2] = |t-s|^{2H}$$

where H is the Hurst parameter. For $H = 1/2$, this process reduces to Brownian motion, while for H close to 0 or 1, it describes more confined or directed trajectories, respectively.

Continuous-Time Random Walk (CTRW). This process helps describe objects often trapped in an obstacle and having trajectories alternating jumps and waiting times [8]. Jumps follow a Gaussian distribution $\mathcal{N}(0, \sigma^2)$, while the waiting times follow a power law distribution on $\psi(t) = t^{-\gamma-1}$ with $\gamma \in [0,1]$ and $t \geq 1$. After every jump, the trajectory $\mathbf{X}$ maintains the same position for a duration equal to the related waiting time. This makes CTRW a Gaussian process with independent but non-stationary increments. Moreover, for large γ, large waiting times are less likely to occur, and CTRW is more similar to Brownian motion.

2.3. Mean square displacement criterion

The mean squared displacement function is defined as :

$$\text{MSD}(\Delta t) = \frac{1}{N - \Delta t + 1} \sum_{k=1}^{N-\Delta t} \|X_{t_k+\Delta t} - X_{t_k}\|^2, \tag{2}$$

and it holds that $MSD(\Delta t) \propto \Delta t$ if $\mathbf{X}$ is a Brownian motion, and $\text{MSD}(\Delta t) \propto \Delta t^\alpha$ with $\alpha > 1$ if $\mathbf{X}$ is a directed motion. These properties define a widely used classification criterion for trajectories based on the fitting of the MSD function with a polynomial function Δt^α. The case $\alpha < 1$ corresponds to subdiffusive motion, including several stochastic models. The MSD criterion has two many drawbacks. First, it is very sensitive to the trajectory length, which influences the precision of the fitting. Secondly, in the case $\alpha < 1$, the MSD criterion enables their classification as local and restricted evolutions but does not give any information about their intrinsic behavior. This paper addresses the second issue and proposes a novel method to distinguish the family of subdiffusive processes presented in the previous section.

3. GEOMETRICAL FEATURES

This section defines the main features used to train the machine learning approach. They account for the directionality and amplitude of movement and summarize the intrinsic properties of previous processes.

The directionality of a movement can be studied by analyzing the turning angle of a displacement compared to the previous one [1]. For every three consecutive points of the trajectory $(X_{t_i}, X_{t_{i+1}}, X_{t_{i+2}})$, we consider the angle

$$\theta_{\mathbf{X}}(t_i) = \pi - \angle X_{t_i} X_{t_{i+1}} X_{t_{i+2}}.$$

To consider the motion direction, we define counterclockwise (with respect to $X_{t_{i+1}} - X_{t_i}$) angles as positive. In the case of CTRW, most angles are null since the particle often remains stationary at the same point. Therefore, non-null angles are computed between triplets of not necessarily consecutive displacements. The histogram of angles $\{\theta_{\mathbf{X}}\}$ defines the related probability density $p_{\theta_{\mathbf{X}}}$. Numerical simulations in Fig. 1 show that the analytical shape of the $p_{\theta_{\mathbf{X}}}$ depends on the process. This indicates a uniform-like distribution for Brownian motion, while directed motion exhibits a Gaussian distribution centered at zero. On the other hand, confined dynamics reveal a predominance of the largest angles.

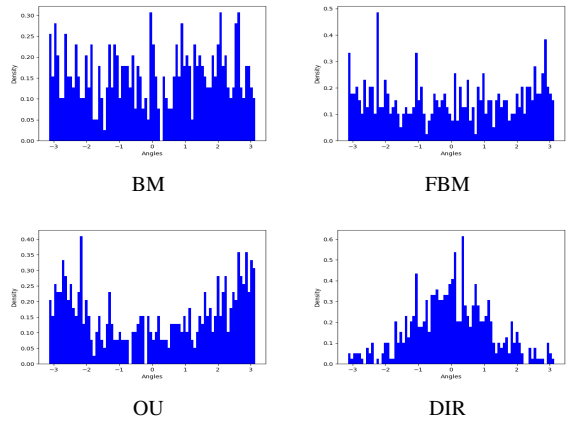

Fig. 1. Histograms for angles $\theta_{\mathbf{X}}$ for different models: BM ($\sigma = 1$), FBM ($H = 0.35$), OU ($\sigma = 1$ and $\lambda = 1$), DIR ($\sigma = 1$ and $\mu = (1, 1)$).

To estimate the shape of the angles distribution, the following fitting is performed:

$$p_{\theta_{\mathbf{X}}}(x) \propto ax^2 \,. \tag{3}$$

The value of a defines a geometrical feature of the trajectory linked to its directional variability.

Moreover, similarly to [1], we consider the index of directionality $P_d(\mathbf{X})$ defined as

$$P_d(\mathbf{X}) = \mathbb{P}(|\theta_{\mathbf{X}}| < \pi/2) - \mathbb{P}(|\theta_{\mathbf{X}}| \geq \pi/2) \,. \tag{4}$$

This feature highlights the existence of a preferred motion direction along the trajectory. In particular, a null P_d corresponds to random motion.

Finally, to detect trapped motion, the following feature is also taken into account :

$$P_{stop}(\mathbf{X}) = \mathbb{P}(\theta_{\mathbf{X}} = 0) \,. \tag{5}$$

The last feature considers how the particle unfurls in space. To this end, we consider the Ripley's index K_r in a ball $B(X_{t_1}, r)$ of radius r centered at the starting point:

$$K_r = |\{X_{t_i} \in \mathbf{X} \,|\, X_{t_i} \in B(X_{t_1}, r)\}| \,/\, N \,.$$

This allows the definition of a feature describing the behavior of the trajectory through balls of increasing radius. Using the reference radius $R = \frac{1}{N}\sum_{i=1}^{N-1} \|X_{t_{i+1}} - X_{t_i}\|$ we compute the vector $K_{\mathbf{X}} = (K_R, ..., K_{N*R})$. The vector $K_{\mathbf{X}}$ can be seen as a sampling of a function depending on r, and using the fitting

$$K_{\mathbf{X}} \propto 1 - e^{-br} \tag{6}$$

the corresponding feature b is defined. Fig. 2 shows the values of this feature for different models. These values indicate that the $K_{\mathbf{X}}$ graph will be more linear (b small) for diffusive trajectories. In contrast, larger values hold for more confined trajectories due to plateaus appearing in the $K_{\mathbf{X}}$ graph since small values of r.

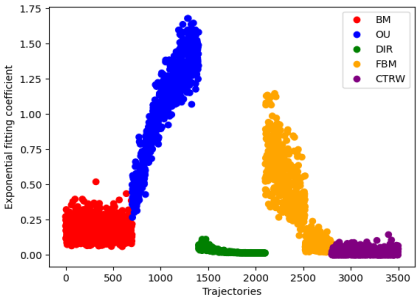

Fig. 2. Fitting coefficient for the vector $K_{\mathbf{X}}$ computed on the dataset used for synthetic simulation in Section 4.

4. MACHINE LEARNING METHOD AND RESULTS

Synthetic dataset. For each model presented in Section 2, we simulated 1000 trajectories of the same length but using different parameters (BM with $\sigma = 1$; OU with $\sigma = 1$ and λ uniformly distributed between 0 and 2; DIR with $\sigma = 1$ and $\mu = u*(1, 1)$ with u uniformly sampled on $[0.3, 2]$; FBM with H between 0.1 and 0.8 but different of 0.5; CTRW with γ uniformly sampled between 0 and 1). Parameters are chosen to avoid generating the same dynamical behavior by different processes (for example, Brownian trajectories, either by BM or by FBM with $H = 0.5$).

For each trajectory, the features dataset collects the parameters defined in (3), (4), (5), and (6). The dataset is split into training and test sets following the ratio of 70%-30% in a balanced way to models and parameters. A Random Forest model (10 trees) is trained using the model's name as labels and validated on the test set. In Fig. 3, we present the results of the method for different trajectories lengths. Most misclassifications concern short FBM trajectories (classified as OU or BM) due to the variability of their dynamic behaviors depending on H. This error decreases significantly with increasing length, proving that FBM describes an intrinsically different dynamic fully characterized by these features. On the other hand, CTRW behavior is easily learned, also for short trajectories, due to successive waiting times and the limit configuration of $p_{\theta_{\mathbf{X}}}$ and $K_{\mathbf{X}}$. Finally, Fig. 4 shows the features'

Length	Acc.	BM	OU	DIR	FBM	CTRW
50	87	87.3	89.3	89	70	100
100	95.3	95	97	95.7	88.7	100
300	98.8	99.3	100	98.7	96	100

Fig. 3. Accuracy and Sensitivity per class obtained via the Random Forest model depending on the trajectory length.

importance for each class based on the mean decrease accuracy method. For each feature, a random permutation of its values is performed, and the trained model is applied to the

new dataset. For each class, the difference between the original and the new sensitivity estimates how that feature discriminates the class (because of randomness, we average the results of 10 independent permutations). The computation is made on the training set for the model with a length of 100.

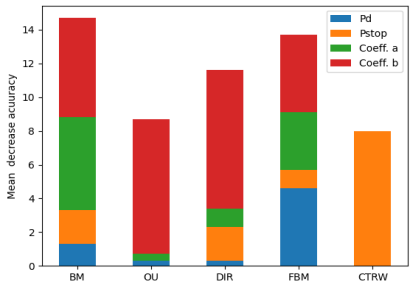

Fig. 4. Feature-specific mean decrease accuracy per class.

Biological dataset. The previous method is used to study the dynamic of the cell membrane receptors CCR5, involved in HIV infection. Receptors can be images using TIRF microscopy and acquired videos (33 frames per second) are analyzed via the *Spot tracking* plugin of the Icy software [9]. We get trajectories of approximately 100-time points and consider the receptor immobile at t if $\|X_{t+1}-X_t\|$ is smaller than a fixed threshold. The model previously trained for length 100 points out various types of motions within the family of receptors. The results in Fig. 5 indicate several subpopulations whose dynamic properties reflect environmental constraints and receptor fate. Thanks to motion analysis, it is possible to understand better the behavior of these groups of receptors and their response to drugs.

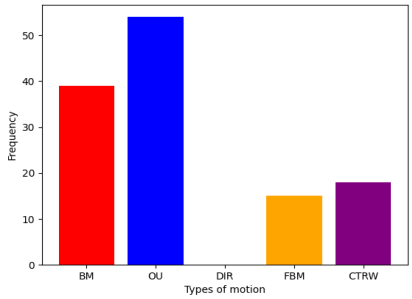

Fig. 5. Motion classification for the trajectories of cell receptors CCR5.

5. CONCLUSION

This paper introduces a novel approach to motion classification based on geometric features describing how the trajectory occurs in space. The proposed method, trained on synthetic data, enables the correct classification of trajectories simulated from the principal stochastic processes used for motion analysis. The obtained scores prove that these features describe these processes' intrinsic properties, supporting the method's explicability. In particular, this method is reliable in distinguishing different subdiffusive behaviors, making it suitable for biological applications.

6. REFERENCES

[1] N. Gal, D. Lechtman-Goldstein, and D. Weihs, "Particle tracking in living cells: a review of the mean square displacement method and beyond," *Rheologica Acta*, vol. 52, pp. 425–443, 2013.

[2] V. Briane, C. Kervrann, and M. Vimond, "Statistical analysis of particle trajectories in living cells," *Physical Review E*, vol. 97, no. 6, p. 062121, 2018.

[3] F. Momboisse, G. Nardi, P. Colin, M. Hery, N. Cordeiro, S. Blachier, O. Schwartz, F. Arenzana-Seisdedos, N. Sauvonnet, J.-C. Olivo-Marin, *et al.*, "Tracking receptor motions at the plasma membrane reveals distinct effects of ligands on ccr5 dynamics depending on its dimerization status," *Elife*, vol. 11, p. e76281, 2022.

[4] Y. Meroz and I. M. Sokolov, "A toolbox for determining subdiffusive mechanisms," *Physics Reports*, vol. 573, pp. 1–29, 2015.

[5] M. Magdziarz, A. Weron, K. Burnecki, and J. Klafter, "Fractional brownian motion versus the continuous-time random walk: A simple test for subdiffusive dynamics," *Physical review letters*, vol. 103, no. 18, p. 180602, 2009.

[6] A. Weron, J. Janczura, E. Boryczka, T. Sungkaworn, and D. Calebiro, "Statistical testing approach for fractional anomalous diffusion classification," *Physical Review E*, vol. 99, no. 4, p. 042149, 2019.

[7] T. Wagner, A. Kroll, C. R. Haramagatti, H.-G. Lipinski, and M. Wiemann, "Classification and segmentation of nanoparticle diffusion trajectories in cellular micro environments," *PloS one*, vol. 12, no. 1, p. e0170165, 2017.

[8] J. Klafter and I. M. Sokolov, *First steps in random walks: from tools to applications*. OUP Oxford, 2011.

[9] F. De Chaumont, S. Dallongeville, N. Chenouard, N. Hervé, S. Pop, T. Provoost, V. Meas-Yedid, P. Pankajakshan, T. Lecomte, Y. Le Montagner, *et al.*, "Icy: an open bioimage informatics platform for extended reproducible research," *Nature methods*, vol. 9, no. 7, pp. 690–696, 2012.