



HAL
open science

Enhancing Fluorescence Correlation Spectroscopy with machine learning to infer anomalous molecular motion

Nathan Quiblier, Jan-Michael Rye, Pierre Leclerc, Henri Truong, Abdelkrim Hannou, Laurent Héliot, Hugues Berry

► **To cite this version:**

Nathan Quiblier, Jan-Michael Rye, Pierre Leclerc, Henri Truong, Abdelkrim Hannou, et al.. Enhancing Fluorescence Correlation Spectroscopy with machine learning to infer anomalous molecular motion. 2025. hal-04650578v2

HAL Id: hal-04650578

<https://inria.hal.science/hal-04650578v2>

Preprint submitted on 8 Jan 2025

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.



Distributed under a Creative Commons Attribution 4.0 International License

Enhancing Fluorescence Correlation Spectroscopy with machine learning to infer anomalous molecular motion

Nathan Quiblier¹, Jan-Michael Rye¹, Pierre Leclerc², Henri Truong², Abdelkrim Hannou², Laurent Heliot², and Hugues Berry^{1,*}

¹AstroSight, Inria, Hospices Civils de Lyon, Universite Claude Bernard Lyon 1, F-69603 Villeurbanne, France

²Univ. Lille, CNRS, UMR 8523, PhLAM Laboratoire de Physique des Lasers, Atomes et Molécules, F-59658, Lille, France

*hugues.berry@inria.fr

January 8, 2025

1 Abstract

The random motion of molecules in living cells has consistently been reported to deviate from standard Brownian motion, a behavior coined as “anomalous diffusion”. To study this phenomenon in living cells, Fluorescence Correlation Spectroscopy (FCS) and Single-Particle Tracking (SPT) are the two main methods of reference. In opposition to SPT, FCS with its classical analysis methodology cannot consider models of motion for which no analytical expression of the auto-correlation function is known. This excludes for instance anomalous Continuous-Time Random Walks (CTRW) and Random Walk on fractal (RWf). Moreover, the whole acquisition sequence of the classical FCS methodology takes several tens of minutes. Here, we propose a new analysis approach that frees FCS of these limitations. Our approach associates each individual FCS recording with a vector of features based on an estimator of the auto-correlation function and uses machine learning to infer the underlying model of motion and to estimate the values of the motion parameters. Using simulated recordings, we show that this approach endows FCS with the capacity to distinguish between a range of standard and anomalous random motions, including CTRW and RWf. Our approach exhibits performances comparable to the best-in-class state-of-the-art algorithms for SPT and can be used with a range of FCS setup parameters. Since it can be applied on individual recordings of short duration, we show that with our method, FCS can be used to monitor rapid changes of the motion parameters. Finally, we apply our method on experimental FCS recordings of calibrated fluorescent beads in increasing concentrations of glycerol in water. Our results accurately predict that the beads follow Brownian motion with a diffusion coefficient and anomalous exponent which agree with classical predictions from Stokes-Einstein law even at large glycerol concentrations. Taken together, our approach significantly augments the analysis power of FCS to capacities that are similar to state-of-the-art SPT approaches.

2 Statement of significance

In this work, we propose to improve Fluorescence Correlation Spectroscopy (FCS), a key biophysical method used to study how molecules move inside cells. By using machine learning to analyze FCS measurements, we show that FCS can handle more complex and more varied types of molecule motion and analyze shorter data recordings. This can for instance be used to detect quick changes in molecule movement along time or to account for molecule motion models that were not accessible up to now in FCS. Using machine learning to analyze FCS measurements therefore makes FCS a more powerful and flexible tool for understanding how molecules move inside cells, a topics with wide applications in biology and medicine.

3 Introduction

Deviation of random motion from standard Brownian motion (BM) has received considerable attention in the literature to describe diverse physical situations [1, 2, 3]. For instance, anomalous diffusion, where the mean-squared displacement scales non-linearly with time, $\langle r^2(t) \rangle = Dt^\alpha$, has been reported to describe the motion of several proteins or particles in living cells [4, 5, 6, 7, 8, 9, 10]. In this case, the exponent α is usually referred to as the anomalous exponent, and D is the diffusion coefficient. All anomalous subdiffusion motion models exhibit $\alpha < 1$, whereas $\alpha = 1$ for standard BM. However, anomalous subdiffusion is a characteristic shared by several unrelated types of motion. For instance, continuous-time random walk (CTRW), fractional Brownian motion (fBM) or random walk on a fractal support (RWf), heterogeneous diffusion processes (HDP) or scaled Brownian Motion (sBM) all exhibit anomalous subdiffusion while the physical processes they describe are very different [9]: heavy-tailed residence time distribution for CTRW, correlation between successive jumps for fBM, changes of diffusion coefficient for HDP and sBM or the fractal geometry of the object on which RWf takes place [11, 12]. Therefore, the complete characterization of the motion of a biomolecule in a live cell requires the completion of two tasks: (i) a classification or selection task to decide what model is the best at explaining the observations (e.g., BM, fBM, RWf or CTRW) and (ii) an inference or calibration task, to estimate the parameter values of the selected model given an experimental observation.

In recent years, the advent of single-particle tracking supra-resolution microscopy [13, 14] has generalized the use of individual trajectories to quantify the motion of biomolecules or particles in living cells. A range of methods have been proposed for the classification and inference tasks based on individual trajectories [15], from simple (non-)linear regression [5, 16], statistical tests [17, 18] or Bayesian inference [19, 20], to machine- [21, 22] and deep-learning [23, 24]. On the other hand, Fluorescence Correlation Spectroscopy (FCS) is the main methodological alternative to single-particle-based techniques for motion characterization of biomolecules in living cells [25, 26]. In FCS, the biomolecules of interest are labelled with a fluorophore, and one monitors the fluctuations of the fluorescence signal due to their interaction with the light beam illuminating the sample. Although alternative approaches have been proposed [27], data analysis in FCS is usually based on the auto-correlation of the fluorescence signal, $G(\tau)$. In the case of BM and fBM, theoretical considerations yield explicit non-linear functions for the expression of $G(\tau)$ as a function of the correlation delay τ , the parameters of the optical setup and the parameters of the model of motion [25, 11]. Fitting this expression to the measured auto-correlation can be used for both model classification and selection with information criteria as well as for parameter inference [28, 16].

Each approach, whether FCS or SPT, comes with its own specificity [12]. FCS can yield good results with a few individual molecules in the illumination volume, but is not a single-molecule approach, as opposed to SPT. The time scales they address are usually different: typically between 1 μ s to 1 ms for FCS *vs* 100 ms to 1 s for SPT. In SPT, one usually has to reconstruct the trajectories from the measured individual localizations. Tracking errors during these reconstructions can induce significant measurement errors [29]. In FCS, the signal-to-noise ratio of the auto-correlation function is usually low, so one has to continuously monitor the signal over long durations (more than 1 second) and average large numbers of consecutive measurements (often more than 100). Because of this, FCS is usually not able to track changes of the motion parameters if they occur over a time scale shorter than several minutes. Finally, analytical expressions for the auto-correlation function $G(\tau)$ are available for BM and fBM, but they are still lacking for other anomalous models, e.g. RWf or CTRW [11]¹. Therefore, FCS is usually considered not to be applicable to the characterization of RWf or CTRW. Recent studies have started to propose alternative FCS measurement and analysis methods to this circumvent this issue [30].

Here we show that most of the above shortcomings of FCS for the classification and characterization of biomolecule motions can be overcome. Instead of fitting the auto-correlation function by a theoretical expression, we use machine learning based on the auto-correlation function to perform the classification and inference tasks. With synthetic FCS data, we show that this approach renders FCS a powerful tool to distinguish between a range of standard and anomalous motions (BM, fBM, CTRW and RWf). The performance of our approach for the classification task and for parameter inference is found to be similar to the best-in-class state-of-the-art SPT algorithms on long trajectories. Our approach

¹Actually, an analytical expression can be obtained for motions defined by stationary processes with anomalous diffusion at all times and Gaussian distribution of the spatial displacements [11]. In practice, this usually restricts to fBM.

accommodates a wide range of FCS experimental setup parameters (beam width and illumination intensity) and uses recordings that are both unique (one recording per estimation) and short (≥ 100 -200 ms). We show that it can be used to accurately track changes of the parameter motions even with 1 Hz parameter-change frequency. Finally we apply the method on experimental data using calibrated beads in water with an increasing concentration of glycerol. Our predictions regarding the model of motion and physical parameters follow the Stokes-Einstein law and serves as a validation of our method.

4 Results

This study focuses on models for anomalous diffusion, *i.e.*, random motions for which the mean squared displacement $\langle r^2(t) \rangle$ scales non-linearly with time:

$$\langle r^2(t) \rangle = 2dDt^\alpha \quad (1)$$

where $r(t)$ is the position of a random walker at time t , $\langle \cdot \rangle$ denotes ensemble averaging (averaging over a population of walkers at time t), $\alpha \in (0, 1]$ is the anomalous coefficient and d the dimension of the space (here $d = 3$). In the equation, $D \in \mathbb{R}_+$ is not a proper diffusion coefficient, since *e.g.*, its unit depends on α , but a “generalized” diffusion coefficient. However, for readability, we will keep referring to it as a diffusion coefficient. The literature refers to motions with $\alpha < 1$ as “subdiffusive” *vs* “superdiffusive” for $\alpha > 1$ ($\alpha = 1$ being standard BM) [3, 11, 12].

We selected three models of motion that give rise to anomalous diffusion:

- Fractional Brownian motion (fBM) that introduces correlations in the lengths of consecutive steps of the walk [2, 31]
- Continuous Time Random Walk (CTRW) where the time spent between two successive steps is a random variable with heavy-tailed distribution [32]
- Random Walk on a fractal (RWf) where the walkers are constrained to take their successive positions on a fractal geometry object [33].

These three models all exhibit the power-law scaling of the mean-squared displacement of eq.(1) but due to totally different reasons. The goal of the present study is to develop an analysis approach able to distinguish between these three models and between standard BM based on standard FCS measurements. More information about numerical simulation of these models can be found in supplementary information section [SI.1.1](#).

4.1 Motion classification and parameter inference on synthetic data

We generated a learning set of more than 3.4 millions simulated FCS experiments, corresponding to 945 values of motion parameters α and D sampled uniformly in $(0, 1)$ and $(0, 10]$, respectively (see sec. [SI.1](#) for details on the generation of synthetic FCS data). For each pair of sampled parameters, 4 sets of trajectories were simulated with the following models: Brownian motion (BM, for which α was set to 1), fractional Brownian motion (fBM), continuous-time random walk (CTRW) and random walk on a fractal (RWf) (more details on the models in sec. [SI.1.1](#)). One constant concern in this study was to develop a method that is robust enough to accommodate a wide range of experimental setup parameters, as encountered in the FCS laboratories worldwide. To this end, for each of the 945×4 trajectories generated with the sampled parameters, we generated 900 FCS recordings by covering a wide range of experimental setup parameters: illumination beam waists $\omega_{xy} \in \{200, 225, 250, 275, 300\}$ nm, and $\omega_z \in \{400, 500, 600\}$ nm and recording durations $T_{\text{obs}} \in \{0.1, 0.25, 0.5, 0.75, 1, 1.25, 1.5, 2\}$ s. Figure 1 provides illustrations of the types of trajectories generated (Fig. 1a1,b1,c1,d1), as well as the corresponding estimators of the auto-correlation $\hat{G}(\tau)$ (defined by eq. 5) for simulated FCS recordings of 0.15 or 1.5 seconds (Fig. 1a2,b2,c2,d2). Due to the FCS signal-to-noise ratio (SNR) of the auto-correlation $\hat{G}(\tau)$, in fitting FCS methodologies, one typically accumulates and averages a large number of measurements (several hundreds), in order to, precisely, compensate for the lower SNR of individual measurements. Here, our objective was to test whether machine learning could exploit the information

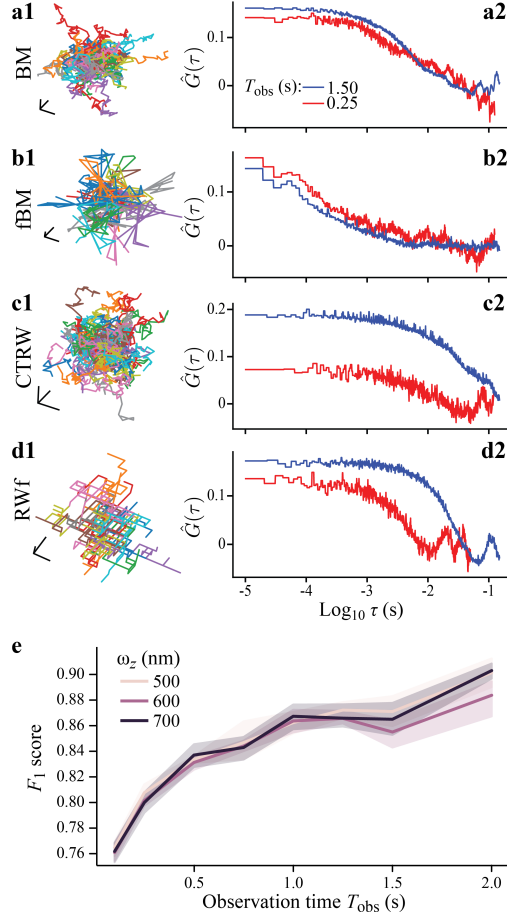


Figure 1: Classification of motion models by FCS on synthetic data. (**a-d**) Illustrations of trajectories and corresponding features $\hat{G}(\tau)$ for Brownian motion (BM, *a*), fractional Brownian motion (fBM, *b*), continuous-time random walks (CTRW, *c*) and random walk on a fractal (RWf, *d*). A classifier was trained to predict the model of motion for such synthetic data. Its performance as a function of the duration of the FCS recordings T_{obs} is shown in (**e**), where full curves show the F_1 scores averaged over the whole test set, grouped by beam waist values ω_z as indicated in the legend. Shaded areas locate ± 1 standard-deviation. For the illustrations of (*a1*, *b1*, *c1*, *d1*), 50 trajectories of 50 steps were selected at random in the learning set, and their initial location set to $(0, 0, 0)$ for readability. Parameter values were $(D, \alpha) = (4.2, 1)$ (*a*), $(2.9, 0.27)$ (*b*), $(2.9, 0.84)$ (*c*) and $(4.1, 0.67)$ (*d*). In each panel the black lines represent (in each dimension): 50 (*a1*), 500 (*b1*), 20 (*c1*) nm or 1 μm (*d1*). In (*a2*, *b2*, *c2*, *d2*), the features $\hat{G}(\tau)$ are shown for recordings of duration 0.25 (*red*) or 1.50 (*blue*) seconds.

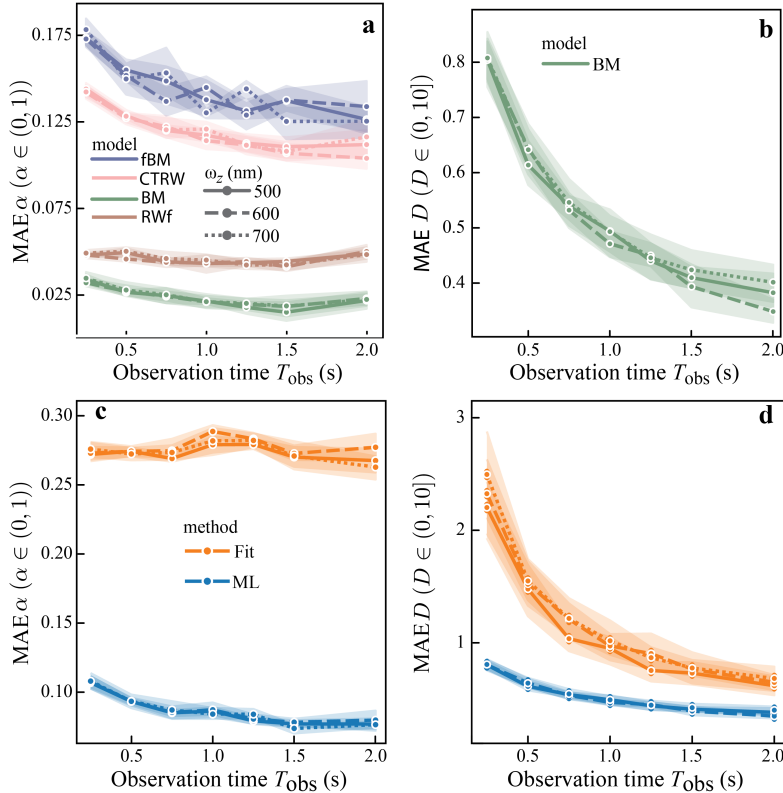


Figure 2: Parameter estimation by FCS on synthetic data. A regressor was trained to estimate the motion parameters of the trajectories: the anomalous exponent α (a,c) and the diffusion coefficient D (b,d). Its accuracy was computed as the mean absolute error (MAE), as a function of the duration of the FCS recordings T_{obs} . Full curves show averages over the test set, grouped by beam waist values ω_z as indicated in the legend. Shaded areas locate ± 1 standard-deviation. Comparison of the accuracy of the classical FCS method (non-linear fit, *orange*) with our machine learning algorithm (ML, *blue*) on synthetic data is shown in (c) and (d). In (a), the MAE is computed for all the data of the test set, using a colorcode that distinguishes the trajectories generated with fBM (indigo), CTRW (pink), BM (green) or RWf (brown), independently of their classification, whereas panel (c) shows averages over all motion types (BM,fBM, CTRW or RWf) and their parameters. Only motions classified as BM are considered in (b) and (d).

contained in individual auto-correlation measurements, despite their low SNR, in the absence of any averaging or accumulation procedure.

Figure 1d shows the performance of our machine learning strategy for the model classification task. Our strategy, described in Online Methods, sec. 6.1.2 is based on histogram gradient boosting and exclusively uses individual auto-correlation measurements as illustrated in Fig. 1a2,b2,c2,d2. Despite the low SNR of individual FCS recordings, our method exhibits very good classification accuracy, as measured by the F_1 -score ($F_1 = TP/[TP + 0.5(FN + FP)]$, with $TP = \#$ true positives, $FN = \#$ false negatives, $FP = \#$ false positives). With observation times larger than 1.0 s, the average F_1 -scores reach large values, in the range [0.88 – 0.90]. As expected, performance decreases with the FCS measurement time, but even with the smallest value used, $T_{\text{obs}} = 0.1$ s, the F_1 -scores remain large, with values close to 0.76. Our algorithm manages to exhibit very similar values for all the beam waists ω_z tested. The shadings of these curves show the standard deviations of the score computed for different values of the motion parameters but also for the different values of ω_{xy} in the test set. The small amplitudes of the shadings again shows the robustness of our algorithm towards the value of the motion parameters and the beam waists. Therefore, the large values of the F_1 -scores exhibited by our algorithm reveal its capacity to perform a robust classification of the motion types, even with individual (non-averaged) and short FCS measurements, and even when CTRW or RWf are part of the possible motions.

Regarding now the regression task, the accuracy of our machine learning algorithm is shown on Figure 2, with separate inference of the anomalous exponent α (Fig. 2a) and the diffusion coefficient

(Fig. 2b). The estimation of α exhibits very good accuracy with MAE (mean absolute error) values around 0.12 for the largest observation times, both for fBM and CTRW. The inference error is smaller with RWf, and even smaller for BM. The smaller error with RWf is due to the fact that the simulation method for RWf generates trajectories that can take only 12 discrete values of α between 0.49 and 0.97 (see supplementary information section [SI.1.2](#)), whereas in data generated as CTRW or fBM, the value of α takes any real-number value in $(0, 1]$ uniformly at random. The inference of α for RWf is thus easier than CRW or fBM. Now the MAE for BM data (*green*) is minimal because for all data that are correctly classified as BM by our algorithm, we set the inferred value of α to exactly 1.0, yielding a zero error. The residual MAE is due to BM data that are misclassified as CTRW, fBM or RWf, for which the inferred value of α is close to the real value 1.0 but not exactly 1.0. As shown by the very good F_1 -score of fig. 1e, the amount of these misclassifications is limited, so the contribution of the misclassified BM data to the MAE remains small. In all cases, the estimation of α of course deteriorates with decreasing recording times, but the loss of accuracy down to $T_{\text{obs}} = 250$ ms remains limited (not larger than 0.15). The accuracy for the estimation of the diffusion coefficient of BM motions is even better. The MAE values are around 0.70 for long T_{obs} , a very good performance given that the real value is sampled uniformly at random in $(0, 10]$. Here again the accuracy decreases with smaller observation times, but even with the smaller value used here, $T_{\text{obs}} = 250$ ms, the error is less than twice the error with $T_{\text{obs}} = 2$ s.

We compared the accuracy of our method with the standard methodology of FCS, that is based on non-linear fitting of the auto-correlation function. Indeed, for BM and fBM, theoretical expressions can be derived for the decay of the auto-correlation function [11]:

$$\bar{G}_{\text{BM}}(\tau) = \frac{1}{N} \left(1 + \frac{4D\tau}{\omega_{xy}^2} \right)^{-1} \left(1 + \frac{4D\tau}{\omega_z^2} \right)^{-1/2} \quad (2)$$

and

$$\bar{G}_{\text{fBM}}(\tau) = \frac{1}{N} \left(1 + \left(\frac{4D\tau}{\omega_{xy}^2} \right)^\alpha \right)^{-1} \left(1 + \left(\frac{4D\tau}{\omega_z^2} \right)^\alpha \right)^{-1/2} \quad (3)$$

Fitting the expression corresponding to the model of motion of the measured auto-correlation function allows one to estimate the value of the free parameters D and α . To our knowledge, such an expression is not available for CTRW nor RWf, so this method cannot be used for parameter estimation in CTRW and RWf. We show in figure 2c a comparison of the accuracy obtained using the above non-linear fits with the one obtained with our machine learning method. Both methods were applied to individual (non-averaged) auto-correlation functions like those shown in fig. 1a2,b2,c2,d2. Given the level of noise present in these recordings, it is not surprising that the estimation of α by standard non-linear fitting is not very good, with accuracies that are 3- to 4-times lower than our machine learning approach (Fig. 2c). Note that to be fair, part of this limited accuracy of the fit method stems from the fact that, for CTRW and RWf, we used the expression for $\bar{G}_{\text{fBM}}(\tau)$ above to fit the autocorrelation function, whereas this expression is not expected to be valid for CTRW and RWf. For the estimation of D , the accuracy of the non-linear fits is markedly better (Fig. 2d). Our ML approach is still approx. 1.8-times more accurate than the standard non-linear fit method at very small T_{obs} , but the accuracy values of both methods converge at long T_{obs} . Therefore, the machine-learning approach proposed in the current study demonstrates better accuracy on individual (non-averaged) synthetic FCS recordings than the standard non-linear fit methods.

4.2 Sensitivity to the brightness

Another parameter that is very likely to vary between experimental setups and even from one individual experiment to the next is total illumination, *i.e.* the mean rate of photon emission of the experiment. To test the sensitivity of our algorithm to variation of total illumination, we validated it on additional synthetic data where we only changed the maximal intensity of the excitation light source (parameter Φ_0 in section [SI.1.3](#)) compared to the learning set. For each of these simulations we set the value of Φ_0 uniformly at random in the range $[3.6, 8.4] \times 10^4$ (for reference, we used $\Phi_0 = 6 \times 10^4$ before). We then evaluated the performance of our algorithm on the classification and inference tasks and averaged the results by groups of T_{obs} and total brightness (fig 3). As expected, for a given value of the brightness, the performance of our model increases with observation times, as already reported above. Moreover, for a given observation time, the best results are obtained for illuminations that are

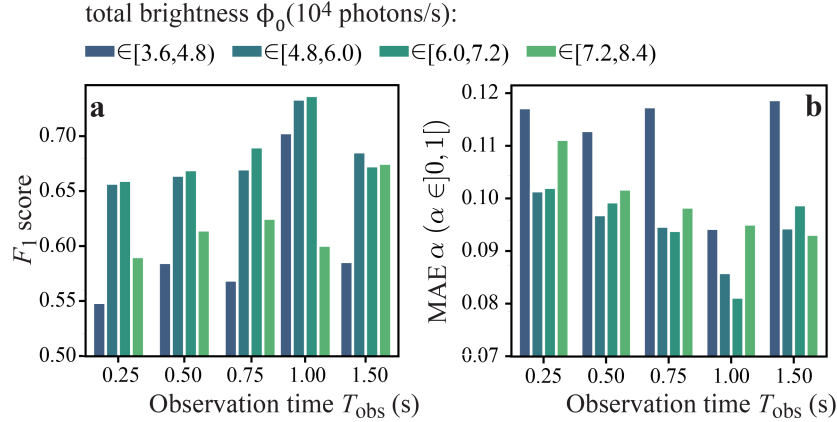


Figure 3: Impact of the total brightness on the performance of the model classification task (a) and the accuracy of the parameter estimation task (b). The MAE for the estimation of α and the F_1 -score are shown as a function of the duration of the FCS recordings T_{obs} on synthetic data. Each bar color refers to a different range of illumination parameters Φ_0 , with the color code given in the legend.

close to the value we used for the training set ($\Phi_0 = 6 \times 10^4$). Performance especially deteriorates for short simulations (less than 1s) with weak illumination (less than 4.8×10^4), thus illustrating the need for a minimum total number of emitted photons. Nevertheless the performance seems robust as long as the illumination remain in $[4.8, 7.2] \times 10^4$. We conclude from this figure that the quality of the estimations made by our machine learning algorithm is robust to restricted perturbations of the total brightness around the value used for training. To recover performance as large as above or for larger perturbations, re-training the algorithm with an illumination matching that of the experimental data would be needed.

4.3 Monitoring fast variations of the motion parameters

We then explored whether our method could be used to monitor rapid changes of the parameters of motion. To this end we used the simulation methodology presented in section SI.1 to generate synthetic FCS recordings of 10 s duration, where we changed the parameter of motion every second. For CTRW motion, we resampled the value of the anomalous exponent α every second according to a uniform distribution in $(0, 1)$. For BM, we resampled the coefficient of diffusion D with the same frequency, using a uniform distribution in $(0, 10]$. Figure 4a1 and b1 show examples of the resulting constant-by-part evolution of the real values of both parameters (*red*). We then applied our algorithm as a sliding window of length 500 ms with a shift of 100 ms after every prediction. Figure 4a1 shows the corresponding estimations of the anomalous exponent for the CTRW case (*gray* trace). The estimation follows the changes of the true value well, with occasional delays and over estimations especially for large values of real α (>0.85), where our algorithm tends to classify the trajectory as BM, thus setting α to exactly 1. On average, however, the estimation error is large only for the first 500 ms after the parameter change, where the sliding window of the segment overlaps two true values (fig. 4a2). Outside of these 500 ms period of overlap, the MAE converges back to the value exhibited with constant α , i.e. around 0.13 for $T_{\text{obs}} = 0.5$ s (compare with fig. 2a). The estimation appears slightly better for the estimation of D , that follows the changes of the true value quite closely (fig. 4b1). Like for α , the mean error on D drastically increases for the first 500 ms after the change of the true value and then returns to low values (fig. 4b2), reaching MAE values similar to those obtained with constant values of the true D (fig. 2b).

4.4 Application to the analysis of experimental data

The previous series of results show that our approach provides a robust and accurate solution to motion classification and inference tasks using synthetic FCS recordings. Interpreting these results as a first validation of our method, we applied it on real experimental data. To this aim, we carried out experimental FCS measurements of calibrated 40 nm fluorescent beads in water with an increased

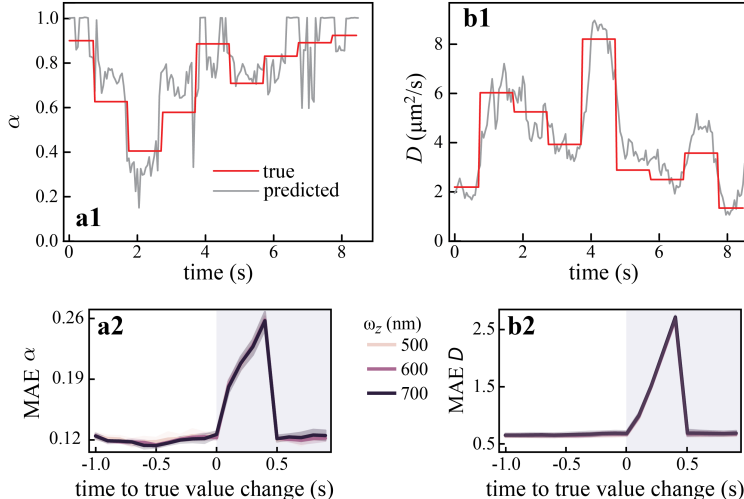


Figure 4: Tracking changes of the motion parameters. 10 seconds FCS recording of CTRW (a) or BM (b) trajectories were simulated as described in sec. SI.1.1, except for the value of parameters α or D , that were resampled from a uniform distribution every second (constant-by-parts *red* curves in *a1,b1*). The resulting 10-s simulated FCS recording was cut into consecutive overlapping segments of 0.5 s duration, with a 0.1 s shift and our machine-learning algorithm was applied to estimate the corresponding values of the parameters (*gray* curves in *a1,b1*). Estimation of the accuracy in each chunk of 1 second between two successive parameter changes is shown via the MAE of α (*a2*) and D (*b2*), grouped by beam waist values ω_z as indicated in the legend. Shaded areas locate ± 1 standard-deviation.

concentration of glycerol (see section 6.2). We applied our algorithm on these 1 second measurement as sliding window of length 500 ms with a shift of 100 ms after every prediction.

Figure 5a shows the results of the classification task with an increasing concentration of glycerol. Note that for these experimental data, the algorithm sometimes predict anomalous subdiffusion but with an exponent α close to 1.0. As a rule of thumb, we therefore opted to re-classify as BM all measurements that were initially classified as anomalous with $\alpha \geq 0.9$. With a small concentration of added glycerol (6%), our algorithm classifies almost all the motion segments as BM (95%), while a only a few of them is classified as RWf (5%). The corresponding estimation of α evidences a mostly bi-modal distribution for 6% glycerol (Figure 5b, blue), with BM motion at $\alpha = 1$ and RWf around $\alpha = 0.9$, thus confirming the overwhelming prevalence of BM. The inferred diffusion coefficient D (Fig. 5c, blue) exhibits an unimodal distribution centred around $9 \mu\text{m}^2/\text{s}$, a value that underestimates the theoretical value of $10.4 \mu\text{m}^2/\text{s}$ for this glycerol concentration (red-grey circles). Note that we have trained our algorithm with values of $D \in (0, 10] \mu\text{m}^2/\text{s}$, so the theoretical value of the diffusion coefficient of the beads in 6% glycerol, $10.4 \mu\text{m}^2/\text{s}$, is slightly beyond our training range. It is therefore not unexpected that our estimations lack accuracy for such low glycerol concentrations. With increasing glycerol values, the fraction of BM segments becomes even more prominent, reaching levels close to 100% from 13 to 31 % glycerol (Fig. 5a). In this range of glycerol concentrations, the inference of α remains mostly concentrated around 1.0 (Fig. 5b) and the distributions of D exhibit medians that are close to the theoretical values (Fig. 5c). For the largest glycerol concentration tested (e.g. 48%), the algorithm still predicts a massive domination of BM ($\approx 85\%$), but a small fraction of fBM appears ($\approx 15\%$) together with traces of CTRW motions (less than 1%). In addition to a majority Brownian population at $\alpha = 1$, the inference of α predicts an anomalous minority population centered on $\alpha = 0.9$ but reaching values down to 0.3. The inference of D remains very good compared to its theoretical value. Therefore our algorithm classifies the bead motions as close to fully BM up to 31% glycerol with inferred D values that match their theoretical values predicted from Stokes-Einstein's law. For higher concentrations, however - here 48% glycerol, the motions become more complex, with a significant population of weakly anomalous (fBM) motion. In opposition to the results obtained with our method, the estimations of α and D obtained with standard non-linear fits show much broader distributions, with medians of anomalous exponents centred around 0.8 to 0.9 (Fig. 5b, orange). The medians of the estimated

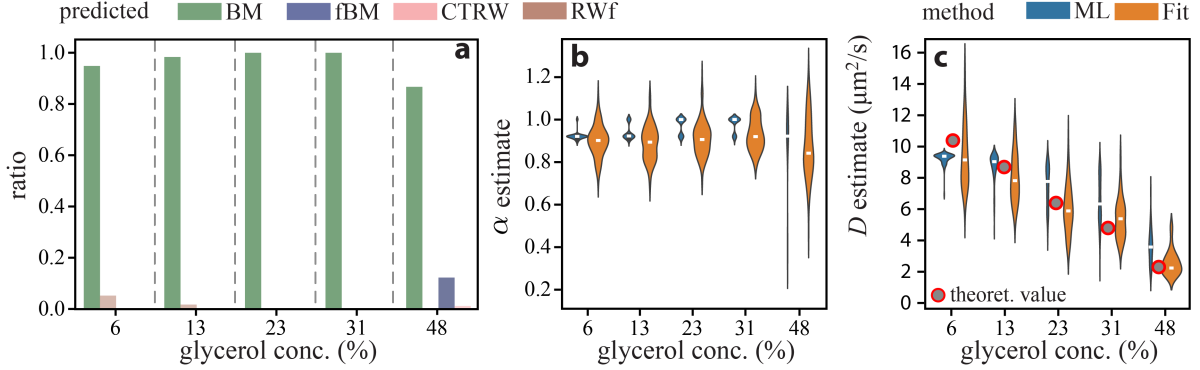


Figure 5: Validation of our machine learning approach to experimental data. The motion of calibrated fluorescent beads in glycerol solutions of increasing concentration was monitored for 1 seconds by FCS (see section 6.2 for details). This recording was cut into 10 consecutive overlapping segments of duration 0.5 s, with a 0.1 s shift. Our machine learning approach was then used for the model classification task (a) and the parameter estimation task (b-c) on each segments. For increasing glycerol concentration, we plot in (a) the ratio of 0.5 s segments motions that were predicted as BM (green), fBM (indigo), CTRW (pink) or RWf (brown). We also show the estimations of the anomalous exponent α (b) and of the diffusion coefficient D for BM cases (c) with comparison between the estimations given by our ML method and a classical non-linear fit.

diffusion coefficients (Fig. 5c, orange) with standard non-linear fit are however slightly closer to the expected theoretical values. Taken together, these data confirm that our ML methodology is more adapted than the standard non-linear fit for short and individual FCS measurements such as those used in these experiments.

We then pushed the analysis further and carried out segmentation of the FCS measurements. To this end, we projected the decision regions of our classification algorithm on a two-dimensional representation. Figure 6 shows the results of this projection as a ternary diagram where the green region shows the zone where the algorithm decides that the motion is BM, whereas the brown region shows where the decision is RWf and purple regions show where the decision is fBM or CTRW. These regions locate positions where the probability of following one model of motion is larger than the probability of following any of the other two motions. To locate the experimental FCS measurements in this 2d-plane, we projected a given experiment as a trajectory made of the classifications given by the successive sliding windows in this ternary coordinate system (full lines with full circles). With low glycerol concentrations (fig. 6 a-d), most of the segments are located or at least end up in the BM domain. In all cases, the trajectories oscillate between the BM and the RWf domain but every incursion into the RWf domain are associated with $\alpha > 0.9$. We conclude that the beads actually conforms to effective BM motion along the whole trajectories. Inspection of the trajectories obtained with larger glycerol concentrations (48%), confirms the results of fig. 5a. These trajectories remain in the center of the triangle, indicating that classification is harder than the other glycerol concentrations (the difference of probabilities between two models is smaller). In addition, the trajectories are more spread out over the regions than for the other concentrations, so that a trajectory can switch classification regions several times, and not only after the first segments, as seen with 6% glycerol. This suggests that with 48% glycerol, the bead motions change and become more complex, in particular with the emergence of a marked heterogeneity of the motion conditions either along time or along the explored space. Alternatively, the addition of such large amounts of glycerol could change the physical properties of the imaged volume to a point that the amount of misclassifications made by our algorithm would increase.

5 Discussion

The current study is a first step to widen the applicability of Fluorescence Correlation Spectroscopy (FCS) by using machine learning for FCS recording analysis. We propose a method that is robust enough to be generic regardless of the specific technical characteristics of the setup under consideration. Depending on the laboratory or even on the specific experiment, the value of the beam waists (in

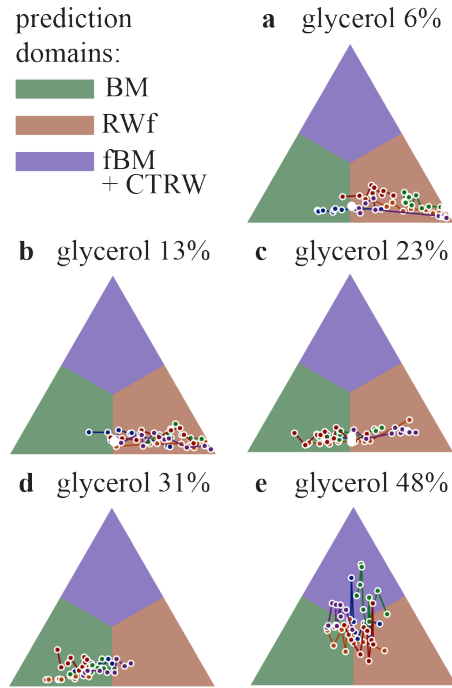


Figure 6: Segmentation of bead motions in glycerol solutions. A ternary diagram is used to map the decision regions of the classification algorithm as a 2d representation: an FCS recording which projection falls in the green region is classified as BM by the algorithm, whereas it is classified as RWf whenever its projections is located in the brown region. Finally, we aggregate fBM and CTRW trajectories, and indifferently locate fBM and CTRW in the indigo region. The successive 0.5 s segments of motion taken from the 1 s FCS recordings of the bead motions of fig. 5 were projected as trajectories using the ternary diagram coordinate (full lines). The initial segment is shown as a white dot. Each panel corresponds to a different concentration of glycerol: 6 (a), 13 (b), 23 (c), 31 (d) or 48 % (e).

x, y or z) or the total brightness can vary. Our machine-learning algorithm has been designed to accommodate a range of values for these parameters. Figures 1, 2 and 4 demonstrate the performances of our algorithm over a wide range of beam waists (from 200 to 300 nm in x, y and from 400 to 600 nm in z) on synthetic data. The limited dispersion of the resulting performance curves suggests that our method is largely independent of the exact value of the beam waists and should be applicable to a wide gamut of beam sizes. We conclude that our machine learning algorithm should be able to accommodate many experimental setups. That being said, the algorithm cannot be expected to exhibit correct performance for technical characteristics that differ significantly from the value ranges used in the training set. For instance, this would be the case for STED-FCS experiments [34], where smaller values of the beam waists are reached (from 100 down-to 30 nm) [35]. In such a case, the accuracy of our approach, trained on the current parameter ranges, will likely deteriorate. This is for instance the case with our bead experiments with 6% glycerol where the theoretical diffusion coefficient is above the range used for training the algorithm (Fig. 5c). For these cases, our algorithm delivers a deteriorated accuracy. However, it is easy to generate a new synthetic learning set with parameter ranges that are better adapted to the specificity of the setup. We provide in parallel with the current article an open-source computer code that can be directly used to generate a new learning set, and train a new version of the algorithm on this more adapted learning set (see Online Methods sec. 6.1.3). Furthermore, approaches similar to the one used here can be exploited for the analysis of other experimental methods that are derived from FCS, such as modulation of the illumination with alternating on/off periods to model fluorescent lifetime correlation spectroscopy (FLCS) [36].

The performance of our algorithm for the model classification and parameter inference tasks on FCS recording can be compared to the algorithms developed for the same tasks on single-particle tracking. To this end, the benchmark provided by the anomalous diffusion (AnDi) challenge is especially useful [37]. This collaborative open community competition has produced a fair benchmarking of the performance of more than 10 state-of-the-art algorithms on synthetic single-particle tracks (SPT). The proposed tasks included a model classification task (among 5 possible anomalous diffusion models), an inference task (anomalous diffusion exponent α) and a segmentation task in which the model class is altered along the trajectories. Because the data used in this challenge were individual single-particle trajectories, the performance of the algorithms was quantified as a function of the most critical parameter, the length L of the trajectories. The inherent discrepancy between single-particle trajectories-based experiments and FCS does not permit a rigorous and fair comparison of the performance of SPT-based methods and FCS-based ones. In particular, the elaboration of the learning datasets, or even the metrics used to evaluate performances cannot be compared. At best, a very rough and loose juxtaposition can be done. For instance, FCS data do not explicitly feature trajectory length whereas this a crucial parameter for SPT. However, since the average length of the imaged trajectories in FCS is expected to increase with the observation time T_{obs} , we use T_{obs} below as a FCS proxy for L in SPT. For the classification task, the best-in-class SPT algorithms exhibit F_1 scores ranging from 0.6 ($L = 40$) to 0.9 ($L > 500$) whereas for our FCS-based algorithm, the F_1 scores for classification varied is larger than 0.86 for $T_{\text{obs}} > 1$ second (fig. 1d). Regarding the inference task, the best SPT methods provided MAE values for α ranging from 0.35 ($L = 40$) down-to ≈ 0.10 ($L > 500$). For comparison, even if we exclude the case of incorrectly classified BM (2a, brown), the MAE of our FCS method for the estimation of α varied from 0.17 ($T_{\text{obs}} = 0.25$ s, fBM) to circa 0.05 ($T_{\text{obs}} \leq 1.5$ s, RWf). We conclude from these comparisons that our FCS-based machine-learning approach exhibits performances that are similar to the best-of-the-class SPT algorithms of the AnDi challenge. Our methods may even be a bit better for short T_{obs} than SPT methods on short L . However, SPT is expected to be more robust to optical aberrations than FCS, so model classification or parameter inference with SPT data may be more robust with real experimental conditions than FCS (but see below). These differences preclude a precise one-to-one comparison, so we only retain the general conclusion that our method on FCS data yields an accuracy that favorably compares to the best-of-the-class methods for SPT data.

In our analyses above, we did not estimate the diffusion coefficient D for anomalous diffusion (fBM, RWf and CTRW). The reason for this lies in part in practicality, since it allows to keep constant the number of free parameters for all the motions considered. But more importantly, the estimation of D in anomalous diffusion is usually a very difficult problem because errors on the estimation of α strongly deteriorate the quality of the estimation of D . This is probably due to the fact that the value of α not only impacts the value of D , but also its unit, which is $\text{distance}^2/\text{time}^\alpha$. An interesting perspective of our work is therefore to develop further methods dedicated to accurate inference of α

and D in anomalous motions. Using deep learning instead of machine learning could, for instance, be an interesting alternative. In our present case, where only the estimation of α was targeted, replacing our machine-learning algorithm (histogram gradient boosting) by a deep-learning approach did not improve the classification. We leave for future work the exploration of the possibility that, with a larger learning set and the inference of both α and D , deep learning might provide better predictions.

As a validation of our method, we applied it to experimental FCS measurement of calibrated fluorescent beads in solutions with an increasing glycerol concentration. For all the studied glycerol concentrations but the largest one (48%), our algorithm predicts that the bead motion essentially remains Brownian with a diffusion coefficient that decreases with an increase of glycerol. This is in agreement with the behavior expected from the diffusion of spherical molecules at very low Reynolds numbers in viscous fluids or from point tracers among diffusing mobile obstacles (see e.g., [38]). Our estimates for the diffusion coefficient D also agree with the values one would expect from the Stokes-Einstein’s law. However, with very large glycerol concentrations (48%), our algorithm reports a slight qualitative change in the bead motions, with close to 15% of the trajectories that depart from pure BM to exhibit mostly fBM motion. It is not clear whether such anomalous behaviors correspond to real fBM motions of the molecules or are mere classification errors caused by e.g., changes of the experimental medium at such large glycerol concentrations. Indeed, alterations of the shape of the detection volume or of the fluorophore photophysics, deviations from uniform spatial distribution of the beads, as well as optical aberrations could all potentially alter the measured signal enough to give rise to misclassifications. Future work will be needed to estimate the extent to which these deviations affect the performance of our algorithm. But in any case, these specific conditions can be accounted for in the simulation engine used to generate the synthetic data used to train the model, thus minimizing the potential classification errors due to deviation from the ideal conditions considered in the current study. Alternatively, adhoc correction methods applied to the raw FCS signal, such as proposed in ref [39] for optical aberrations, can still be applied to the raw experimental data of interest before analysis by our current algorithm.

Advantages and drawbacks of the proposed method

Compared to the traditional methodology for FCS analysis, that relies of non-linear fitting of the auto-correlation function, our machine learning approach presents a number of advantages and drawbacks:

1. Our proposed method systematically performs model selection before the parameter estimation step. Model selection is in principle feasible with traditional fitting methods, using a criterion that takes into account the number of free parameters of the fitting expression (e.g., Bayesian or Akaike Information criteria). However, this is rarely done in practice and can only be used to decide between Brownian and fractional Brownian motion.
2. The traditional fitting method is limited to motion models for which there exists an analytical expression for the ACF of the photon emission intensity. Again, in practice, this restricts the motion models to BM and fBM. Our method on the other hand is applicable to any motion model that can be simulated by Monte-Carlo methods and for which the time-averaged auto-correlation estimator \hat{G} (eq. 5) can be computed numerically. In the current study we applied it to BM, fBM, CTRW and RWf but any other motion model could be accounted for after retraining of the machine learning model with synthetic data generated using this new model. This includes, e.g., Heterogeneous Diffusion processes (HDP) [40], Scaled Brownian motion (SBM) [41], or “subordinated” random walks that combine two models, for instance CTRW + fBM [42, 10]. Our open source framework available at <https://gitlab.inria.fr/aistrosight/mlfcs> provides the code needed to generate such a personalized synthetic learning set and re-train the algorithm on it.
3. The analytical expression needed for the nonlinear fitting of the auto-correlation function for the traditional methods assumes a 3D Gaussian shape for the sample illumination. It also assumes that the space (surface or volume) in which the motion takes place is isotropic and homogeneous [25]. This assumption is violated, e.g. when the illumination encompasses both a membrane (plasmic or nuclear) inside the 3D bulk of the cytosol or the nucleus. Our method can accommodate non-Gaussian illuminations or diffusion geometries that are anisotropic or non-homogeneous. This would demand retraining of the machine learning algorithm with synthetic

data generated by a simulation model that takes into account the correct illumination shape or volume geometry.

4. In contrast to the traditional analysis method, our approach delivers very good performance even when applied to a single FCS recording of short duration (only 250 ms). This is particularly useful for monitoring rapid changes in patterns or motion parameters that occur at hertz frequencies. One advantage is to reduce the risk of bleaching, which is a problem when measuring FCS in a cell. In addition, it avoids the need for the user to accumulate and average a large number of successive recordings from one second to the next.
5. Our method could be extended to the analysis of mixed populations of molecules sharing the same motion model but with different parameters. For example, a population of molecules moving according to BM with two diffusion coefficients. This would demand to train the algorithm using synthetic data from mixed populations, for instance. Future work will be needed to assess the realism of this proposal.

6 Methods

6.1 Machine learning methods

The goal of our machine learning approach is to (i) learn to predict the class of motion M of the random walkers among the set of possible motions $\mathcal{M} = \{\text{BM}, \text{fBM}, \text{CTRW}, \text{RWf}\}$ (classification task), and (ii) estimate the value of the parameters θ_M of this motion, i.e. D for BM and α for fBM, CTRW and RWf.

6.1.1 Auto-correlation functions

Our analysis starts with the collection of photon emission times, $\{\Gamma(t), t \leq T_{\text{obs}}\}$ that constitutes the raw data of an FCS experiment (see sec. [SI.1.3](#)). T_{obs} is the total measurement duration. Let $E = (\mathcal{S}, \mathcal{F}, \mathbb{P})$ be a probability space with sample space \mathcal{S} , event space \mathcal{F} and probability function \mathbb{P} . In case of a stationary process (true for BM and a fBM), Γ is $L^2([0, T_{\text{obs}}], E)$, in the sense that $\|\Gamma\|_{L^2} = \int_0^{T_{\text{obs}}} \mathbb{E} [|\Gamma(s)|^2] ds < +\infty$. In this case, Γ admits an auto-correlation function [\[43\]](#) denoted $\{G(\tau), \tau \in [0, T_{\text{obs}} - \tau]\}$ that depends on the auto-correlation lag τ but not on time t :

$$G(\tau) = \frac{\langle \Gamma(t)\Gamma(t+\tau) \rangle - \langle \Gamma(t) \rangle \langle \Gamma(t+\tau) \rangle}{\sqrt{\langle \Gamma^2(t) \rangle \langle \Gamma^2(t+\tau) \rangle}} \quad (4)$$

where $\langle \cdot \rangle$ denotes ensemble averaging.

To introduce time binning, we first define a few notations:

- Number of photons emitted between t_a and t_b : $I[t_a, t_b] = \int_{t_a}^{t_b} \Gamma(s) ds$
- Bin interval: $\Delta\tau = \frac{T_{\text{obs}}}{L}$, where L the length of the binned vector
- Binned value of I : $(I[i])_{i \in [0, L-1]} = (I[i\Delta\tau, (i+1)\Delta\tau])_{i \in [0, L-1]}$

Using these notations, we estimate the ensemble-average $\langle \Gamma(t) \rangle$ of eq. [4](#) by its time-average $\bar{I} = \frac{1}{L} \sum_{i=0}^{L-1} I[i]$ and its second moment $\langle \Gamma^2(t) \rangle$ by \bar{I}^2 , since $\langle \Gamma^2(t) \rangle = \langle \Gamma(t) \rangle^2$ for a Poisson process. This leads to an approximation of G by its time-averaged auto-correlation estimator \hat{G} [\[44\]](#):

$$\hat{G}(\tau) = \frac{1}{L - \tau/\Delta\tau} \sum_{i=0}^{L-1-\tau/\Delta\tau} \frac{I[i] I[i + \tau/\Delta\tau] - \bar{I}^2}{\bar{I}^2} \quad (5)$$

In case Γ is not stationary but still $L^2([0, T_{\text{obs}}], E)$, i.e. for the CTRW in our case, the auto-correlation function eq. [\(4\)](#) is not defined, but it is still possible to construct a partial auto-correlation function [\[45, 46, 47\]](#) for every associated $t \in [0, T_{\text{obs}}]$, denoted as $\{G_t(\tau), t, \tau \in [0, T_{\text{obs}} - \tau]\}$. The

partial auto-correlation function of such a non-stationary process is a quantity characterizing the autocorrelation function of the stationary process associated to the non-stationary process for every t , defined by :

$$G_t(\tau) = \frac{\langle \Gamma(t)\Gamma(t+\tau) \rangle - \langle \Gamma(t) \rangle \langle \Gamma(t+\tau) \rangle}{\sqrt{\langle \Gamma^2(t) \rangle \langle \Gamma^2(t+\tau) \rangle}}, \quad (6)$$

In theory, the partial auto-correlation function of a non-stationary process cannot be estimated by time averaging, but only by ensemble averaging [48]. This is not suitable in our case since we want to produce estimations for a handfull of molecules, as expected in FCS. However, we still used the time-averaging of eq. 5 as a feature to quantify the auto-correlation of non-stationary processes based on the *ansatz* that this feature is still good enough for machine learning algorithms. This *ansatz* originates from the hypothesis that the process Γ exhibits periodicity at long times, which would mean that the mean on t of its partial auto-correlation function

$$m(\tau) = \lim_{T_{\text{obs}} \rightarrow \infty} \frac{1}{T_{\text{obs}} - \tau} \int_0^{T_{\text{obs}} - \tau} G_s(\tau) ds \quad (7)$$

exists and is finite. In this case, the quantity $\hat{G}(\tau)$ from eq. (5) is also a good estimator for non-stationary processes.

As a final step, we normalize the feature $\hat{G}(i\Delta t)$ obtained from eq. (5) by dividing it by the mean of its first five elements and reduce dimensionality by keeping only the first $K < \frac{L}{2}$ values of the sequence, using log sampling of the delay τ .

6.1.2 Machine learning methods

Learning set. A central concern in this work is that our machine learning methods must be robust to the variety of setups used in experimental labs and, in particular, must be able to be generalized to a range of beam waists ω_{xy} and ω_z . To this aim we generated a learning set comprising more than 2.5 million simulated FCS experiments of various duration and beam waists, in the following way:

- We first set the value of the motion parameters with uniform sampling: $\alpha \sim \mathcal{U}((0, 1))$ for fBM and CTRW and $\alpha \sim \mathcal{U}(\{0.495, \dots, 0.970\})$ for RWf (see section SI.1.2 below) and $D \sim \mathcal{U}((0, 10])$.
- Using the algorithms described in section SI.1.1, we then generated four sets of simulated trajectories using the sampled α and D : one with fBM motion, one with CTRW motion, one with RWf and one with BM motion (for BM, we set $\alpha = 1$).
- For each resulting set of trajectories, we sampled the corresponding set of photon emission times for 3 seconds, using the thinning algorithm of section SI.1.3. The process of photon time sampling was repeated with all possible pairs of beam waists among $\omega_{xy} \in \{200, 225, 250, 275, 300\}$ nm and $\omega_z \in \{400, 500, 600\}$ nm, resulting in 15 FCS simulations per set of trajectories.
- In order to analyze the performance of our machine learning algorithms depending on the duration of the FCS experiment, every 3s FCS simulation described above was split into non-overlapping segments of duration $T_{\text{obs}} \in \{0.1, 0.25, 0.5, 0.75, 1, 1.25, 1.5, 2\}$ seconds, and every one of the 60 resulting segments was used in the learning set. With this procedure, the number of examples in the learning set was larger for short T_{obs} than longer ones (e.g., 10 times more examples with $T_{\text{obs}} = 0.1$ s compared to $T_{\text{obs}} = 1.0$ s). This allowed us to invest more learning effort on shorter observation times than larger ones.
- Finally, we computed the estimator of the auto-correlation \hat{G} from eq. (5) for each of the simulation fragments above.

We repeated this process 945 times (i.e., 945 samplings of the motion parameters), yielding a learning set of 3,402,000 simulated FCS experiments in total. This learning set was then split into a test set (315 sampled parameter values, i.e. 1,134,000 simulations, 4.8% of the total) and a training set (the rest of the simulations) using uniform distribution.

Learning algorithm. The initial feature associated with each simulated FCS experiment is a vector of size 1,003, comprising the 1,000 log-sampled values of \hat{G} , plus the values of ω_{xy} , ω_z and T_{obs} used

for this simulation. Therefore, the values of these parameters ω_{xy} , ω_z and T_{obs} are not inferred by our algorithm but must be provided by the user. We used these features to train a classifier C with the Histogram Gradient Boosting Classifier of scikit-learn[49] (`sklearn.ensemble.HistGradientBoostingClassifier`) with default parameters. The classifier yields the predicted model probability for the simulation: $C(\hat{G}, T_{\text{obs}}, \omega_{xy}, \omega_z) = (\mathbb{P}_{\text{BM}}, \mathbb{P}_{\text{fBM}}, \mathbb{P}_{\text{CTRW}}, \mathbb{P}_{\text{RWf}})$.

In a second phase, we trained regressors to determine α and D (Histogram Gradient Boosting Regressor of scikit-learn with default parameters), individually for each pair of ω_{xy} and ω_z and each candidate model. For each pair of values (α, D) , this resulted in $5 \times 3 \times 4 = 60$ classifiers, $(R_{\omega_{xy}, \omega_z, M})$. The input to these classifiers is also the vector of size 1,003: $(\hat{G}, T_{\text{obs}}, \omega_{xy}, \omega_z)$. For example $R_{225, 600, \text{fBM}}$ is trained on data with beam waist diameter of $\omega_{xy} = 225$ nm, $\omega_z = 600$ nm and with diffusion model fBM. These regressors are trained to predict α and D :

$$R_{\omega_{xy}, \omega_z, M}(\hat{G}, T_{\text{obs}}, \omega_{xy}, \omega_z) = \begin{cases} \hat{\alpha} & \text{if } M \in \{\text{fBM}, \text{CTRW}, \text{RWf}\} \\ \hat{D} & \text{if } M = \text{BM} \end{cases} \quad (8)$$

The final stage consolidates the classification and the regression tasks above using a last Histogram Gradient Boosting Regressor that takes into account the regression estimation for all beam waists and model classes. This final regressor \bar{R} learns to predict α and D taking as input the output of the above classifier C and the outputs of the 45 corresponding regressors (vector of size $3+60+3=66$): $(R_{\omega_{xy}, \omega_z, M})$:

$$\begin{aligned} \bar{R}\left(C(\hat{G}, T_{\text{obs}}, \omega_{xy}, \omega_z), \left(R_{\omega_{xy}, \omega_z, M}(\hat{G}, T_{\text{obs}}, \omega_{xy}, \omega_z)\right), T_{\text{obs}}, \omega_{xy}, \omega_z\right) \\ = \begin{cases} \hat{\alpha} & \text{if } M \in \{\text{fBM}, \text{CTRW}, \text{RWf}\} \\ \hat{D} & \text{if } M = \text{BM} \end{cases} \end{aligned} \quad (9)$$

For inference or testing, we determine the model class according to the maximal value of \mathbb{P}_M estimated by the classifier C and the estimation of the parameter value ($\hat{\alpha}$ or \hat{D}) according to the prediction of the final regressor \bar{R} .

6.1.3 Code availability

The entirety of the code used in the present article is available as an open source framework at <https://gitlab.inria.fr/aistrosight/mlfcs>. The repository also offers the possibility to download the trained algorithm for use on a local computer. In addition, we provide the code needed to generate a personalised synthetic learning set and to train the algorithm on it. The repository also proposes a simple interface based on a Jupyter notebook that allows the user to upload their own data (either as direct FCS recordings or the derived auto-correlation functions) and use our trained algorithm for the classification and inference tasks. The gitlab repository comes with a medium-size test set of synthetic trajectories, that can be used to test the performance of the algorithm. The whole training set used in the present manuscript (more than 3.4 million synthetic trajectories), or the experimental FCS measurements of the beads represent a considerable volume of data. The corresponding files are too large to be made available on an open access server, but they can be obtained from the authors upon request.

6.2 Experimental data

To evaluate our estimation method, we tested it on experimental data. We carried out FCS measurements with calibrated fluorescent beads. FCS measurements were performed using a confocal microscope (Nikon A1R) with a 488 nm diode laser (LBX-488, Oxxius). Experiments were conducted with polystyrene nanobeads (Fluoro-Max G40, Thermo Fisher) with an average diameter of 40 nm and diluted in a water-glycerol mixture to modulate the viscosity and, consequently, the diffusion coefficient. The sample was placed in a glass bottom dish (0.16-0.19 mm, P35G-1.5-20-C MatTek) and FCS measurements were acquired using a 40x NA = 1.25 water immersion objectif (CFI Apo LWD Lambda S). The pinhole was set at 1.2 Airy Units, and the temperature of sample was kept at 24.9°C during the experiments. The experimental values of ω_{xy} and ω_z were measured during calibration on Atto488

in distilled water that is expected to exhibit Brownian motion with a theoretical diffusion coefficient of $400 \mu\text{m}^2/\text{s}$. We calibrated the values of ω_{xy} and ω_z using classical fit of the ACF of the experimental recording assuming a BM. The resulting beam waists were $\omega_{xy} = 214 \text{ nm}$ and $\omega_z = 522 \text{ nm}$. The output signal from the sample was collected with a photon counting module (SPCM-CD, Excelitas), and time tagging was carried out by a time-correlated single photon counting module (HydraHarp 400, PicoQuant). Bead solutions were diluted to reach a concentration of 10^{11} particles/mL, resulting in approximately 2 individual beads on average within the focal volume.

Assuming that the beads in glycerol solutions are spherical objects and the flows are dominated by the viscous effect, the Reynolds numbers is very small ($Re \ll 1$). Then, the theoretical value of their diffusion coefficient can be estimated using the Stokes-Einstein formula $D = kT/(6\pi \cdot \eta_0 \cdot r_{\text{beads}})$ where η_0 is the viscosity of the glycerol solution and r_{beads} the bead radius. We estimated the dependence of the viscosity η_0 to glycerol concentration according to Ref [50, 51, 52]. Using $r_{\text{beads}} = 20 \text{ nm}$ in the Stokes-Einstein formula then yields a theoretical estimate for the bead diffusion coefficient.

7 Acknowledgments

The authors acknowledge financial support by the French Agence Nationale de la Recherche (ANR), project ABC4M under reference ANR-20-CE45-0023.

8 Declaration of Interest

The authors declare no competing interests.

9 Author Contributions

N.Q. developed the computer code for synthetic data generation, algorithm training and exploitation, analyzed the synthetic and experimental data, and drafted the paper. J-M.R contributed analytic tools and participated to writing and improving the computer code from initial prototypes. P.L, H.T., and A.H. performed the experiments with beads in glycerol and contributed analytic tools. H.T., A.H. and L.H. participated in the writing of the manuscript. L.H. supervised the experimental work and contributed analytic tools. H.B. designed the research, led its development, contributed to data analysis, and wrote the manuscript.

SI Supplementary Information

SI.1 Generation of synthetic FCS data

SI.1.1 Models of random motion

We note W_i the waiting time between the $i - 1^{th}$ and the i^{th} jumps of the random walker and consider $(W_i)_{(i \geq 1)}$ the associated i.i.d family of random variables of density λ . We associate it with the jump time J_i of the i^{th} jump:

$$J_i = \sum_{n=1}^i W_n \quad \text{with } i \geq 1 \quad (\text{SI.1})$$

Let $\Delta_i \in \mathbb{R}^d$ be the vector in space representing the i^{th} displacement in space. We note $(\Delta_i)_{(i \geq 1)}$ the corresponding family of random variables, of law ΔX_i . The position of the particle in the d -dimensional-space at time t , $r(t)$, with initial position $r_0 \in \mathbb{R}^d$ is

$$r(t) = r_0 + \sum_{i \geq 0} \Delta_i \mathbb{1}_{\{J_i < t\}} \quad (\text{SI.2})$$

Consider a walker located at position x at time s , that has arrived there at time $J_i = t - s$. With these notations, the next jump of the walker will happen at time $J_{i+1} = t - s + W_{i+1}$, and its new position will be $x + \Delta_{i+1}$.

In the current study, we focus on three motion models, that we define below for the spatial dimension $d = 1$:

- Brownian motion (BM) [53] is a stationary process with independent Gaussian increments: $\lambda = \delta_{dt}$, with dt the simulation time step. For BM, $(\Delta_i) = \left(\mathcal{N} \left(0, \sqrt{2Ddt} \right) \right)$ is an i.i.d. Gaussian random variable family $\forall i$, $\alpha = 1$.
- Fractional Brownian motion (fBM) [2, 31, 54, 55], which is also a stationary Gaussian process but different from white noise due to the temporal auto-correlation of its increments: $\lambda = \delta_{dt}$, $\mathbb{E}[\Delta_i \Delta_j] = D (|idt|^\alpha + |jdt|^\alpha + |idt - jdt|^\alpha)$, $\alpha \leq 1$.
- Continuous time random walk (CTRW) [32, 56], which also has Gaussian distributed jumps, but which is not a stationary process if the distribution of its residence times is heavy-tailed, for instance according to a power-law: $\lambda(t) = \frac{\alpha}{\epsilon} \left(\frac{\epsilon}{\epsilon+t} \right)^{\alpha+1}$, $(\Delta_i) = \mathcal{N} \left(0, \sqrt{2Ddt} \right)$, $\forall i$, $\alpha < 1$. We used $\epsilon = 10^{-7}$ throughout this work.
- Random walk on Fractal (RWf) [57], in particular on generalized Sierpinski carpets [58]. RWf is a process evolving on a 3d grid where only the vertices of a given fractal are attainable to the random walker. At each time step dt , the walker chooses uniformly on all edges connected to its position a candidate for the next move. If the vertice of the edge chosen is accessible (i.e. it is part of the fractal) the walker accepts the move, else the walker stays at its initial position (blind-ant algorithm [33]). Each step is of size $dx = C_{\text{fractal}} \sqrt{D}$. With C_{fractal} a normalization constant that depends on the chosen fractal pattern, and $\alpha \leq 1$ (more details in section SI.1.2).

In this study, BM, fBM and CTRW trajectories were simulated in $d = 3$ space dimensions by simulating a $d = 1$ independent random walk for each of the 3 dimensions, whereas the RWf model was inherently 3d because of the chosen 3d-fractal. The random walks were simulated in a sphere Ω of diameter $\{\Omega_x, \Omega_y, \Omega_z\}$ centered on $(x, y, z) = (0, 0, 0)$. Their initial location was uniformly distributed in Ω . To keep a constant density of walkers in Ω , some form of boundary condition has to be imposed at the surface of the sphere. We rejected reflective boundaries because they induce artificial correlations that strongly impact the auto-correlation signal. Instead, we used the following condition: whenever a walker leaves the sphere, we remove it from the simulation and replace it by a new walker, the initial location of which is chosen at random over the surface of the sphere.

SI.1.2 Creation of Generalized 3D Sierpinski Fractals

The generalized 3D Sierpinski fractal is a model for hierarchical and porous environments, such as found in biological systems, which features structural complexity across scales. Here, we followed the generation process proposed by ref [58]. The process begins with a cubic lattice of size N^3 , where N is the number of divisions along each axis. At each iteration, the cube is subdivided into ν^3 smaller sub-cubes, with ν the subdivision factor. A fixed number o of these sub-cubes is then removed based on a randomly predefined removal pattern, leaving $\nu^3 - o$ sub-cubes for further subdivision in subsequent iterations. The process is iterated m times for each sub-cubes, where $m = \log_\nu N$, resulting in a fractal structure characterized by self-similarity and hierarchical voids. Finally, to create a seamless geometry suitable for modelling diffusive processes or random walks, periodic boundary conditions are applied to the fractal.

To illustrate this process, consider a cube with $\nu = 3$ subdivisions (resulting in $3^3 = 27$ sub-cubes) and $o = 2$ randomly removed sub-cubes per iteration. After one iteration, $27 - 2 = 25$ sub-cubes remain, which are then subdivided into $25 \cdot 27$ smaller sub-cubes. Repeating this process for $m = 3$ iterations produces a fractal with complex occupied regions. Importantly, the same removal pattern is applied to each of the sub-cubes at every level of the hierarchy, ensuring self-similarity across scales.

This construction process is highly adaptable, allowing the control of parameters such as ν , o , and m to tailor the fractal density, connectivity, and complexity. For each removal pattern, numerous simulations of random walks are performed to estimate the effective diffusion coefficient D and the anomalous exponent α . Then, distinct removal patterns are chosen to obtain a sampling of $\alpha \in \{0.495, 0.732, 0.802, 0.813, 0.819, 0.844, 0.875, 0.889, 0.914, 0.931, 0.959, 0.970\}$. The diffusion coefficient is tuned by scaling the trajectory coordinates by the squared root of the target D value after division by the square root of the effective value of D associated to the pattern. These properties make the 3D Sierpinski fractal an ideal model for exploring diffusion in heterogeneous environments.

SI.1.3 Modelling of FCS measurements

We simulated an FCS illumination volume centered at $(0, 0, 0)$, the center of the spherical domain Ω in which the random walks occur. The point spread function (PSF) of the microscope is modelled as a 3d Gaussian with beam waists $\omega_i \ll \Omega_i$, $\forall i \in \{x, y, z\}$ [25]. In agreement with the experimental situation we considered identical beam waists in the x and y directions, i.e. $\omega_x = \omega_y := \omega_{xy}$. The illumination intensity Φ is thus given by

$$\Phi(x, y, z) = \Phi_0 e^{-2\left(\frac{x^2+y^2}{\omega_{xy}^2} + \frac{z^2}{\omega_z^2}\right)}, \quad (\text{SI.3})$$

where Φ_0 controls the illumination intensity.

The probability that a particle located at (x, y, z) emits a photon is modelled as a Poisson process with a rate proportional to the value of the illumination at this position [59, 25]. Since the particle location changes according to the random walk, we model photon emission by a single walking particle as a non-homogeneous Poisson process [60], with time-dependent rate $\mu(t) = \Phi(r(t))$.

If $\gamma = \{\gamma(t), t \geq 0\}$ is the process characterizing the times of photon emission by a single molecule, one has

$$(\gamma(t))_{t \geq 0} = \partial \mathcal{P} \left((\mu(t))_{t \geq 0} \right) \quad (\text{SI.4})$$

where $\partial \mathcal{P}$ is the process of the jump times of a Poisson process, i.e. if $(\Theta_i)_{i \geq 0}$ are the jump times of \mathcal{P} , then $\partial \mathcal{P} = \sum_{i \geq 0} \delta_{\Theta_i}$. Now, to retrieve $\Gamma = \{\Gamma(t), t \geq 0\}$, the counting process characterizing the emission times of photons from an FCS experiment with N molecules, we sum the N processes characterizing each molecule $\Gamma(t) = \sum_{n=1}^N \gamma_n(t)$. By the additive property of Poisson processes ($\partial \mathcal{P}(\sigma) + \partial \mathcal{P}(\nu) = \partial \mathcal{P}(\sigma + \nu)$ [61]), the intensity of the system can be modelled as the sum of the intensities of the N processes:

$$\bar{\mu}(t) = \sum_{n=1}^N \mu_n(t) \quad \text{and} \quad \Gamma = \partial \mathcal{P} \left((\bar{\mu}(t))_{t \geq 0} \right) \quad (\text{SI.5})$$

We simulate the process of photon emission by all the N molecules, Γ , by thinning [62]. We suppose that its rate $\bar{\mu}(t)$ is bounded for $t \in [0, T_{\text{obs}}]$ by its maximal value $\|\bar{\mu}\|_\infty < +\infty$. We first sample the

photon emission times that would be expected from a Poisson process with constant (homogeneous) rate $\|\bar{\mu}\|_\infty$: $\tilde{\Gamma} = \partial\mathcal{P} \left((\|\bar{\mu}\|_\infty)_{t \geq 0} \right)$. We refer to those as candidate emission times \tilde{T}_i :

$$\tilde{\Gamma} = \sum_{i \geq 0} \delta_{\tilde{T}_i} \sim \partial\mathcal{P} (\|\bar{\mu}\|_\infty) \quad (\text{SI.6})$$

We then reject some of the candidate emission times \tilde{T}_i to adapt them to $\bar{\mu}(t)$: we associate with every candidate emission time \tilde{T}_i , a uniformly-distributed random variable $U_i \sim \mathcal{U}(0, \|\bar{\mu}\|_\infty)$ and reject every \tilde{T}_i for which $U_i > \bar{\mu}(\tilde{T}_i)$. The emission times that were not rejected thus define the photon emission times of our initial process:

$$\Gamma = \sum_{i \geq 0} \delta_{\tilde{T}_i} \cdot \mathbb{1}_{U_i \leq \bar{\mu}(\tilde{T}_i)} \sim \partial\mathcal{P} \left((\bar{\mu}(t))_{t \geq 0} \right) \quad (\text{SI.7})$$

The output of the simulation is the resulting collection of the times the N random walkers emitted photons, $(T_i)_{i \geq 0}$. Note that the above process is currently in continuous time, but it will be binned later during pre-processing.

SI.1.4 FCS simulation parameters

For the simulations of the current paper, we used the following parameter values:

- Size of the spatial domain \mathcal{D} (μm): $(\mathcal{D}_x, \mathcal{D}_y, \mathcal{D}_z) = (1.05, 1.05, 2.4)$
- Time step $dt = 1 \mu\text{s}$
- Beam waist (μm): $\omega_{xy} \in \{0.200, 0.225, 0.250, 0.275, 0.300\}$, $\omega_z \in \{0.500, 0.600, 0.700\}$.
- Mean number of random walkers in the illuminated volume $v(= 4/3\pi\omega_{xy}^2\omega_z)$: $n = 5$
- Maximal illumination $\Phi_0 = 6 \times 10^4$

References

- [1] J. R. Hosking, “Modeling persistence in hydrological time series using fractional differencing,” *Water resources research*, vol. 20, no. 12, pp. 1898–1908, 1984.
- [2] B. B. Mandelbrot and J. W. Van Ness, “Fractional brownian motions, fractional noises and applications,” *SIAM review*, vol. 10, no. 4, pp. 422–437, 1968.
- [3] R. Metzler and J. Klafter, “The random walk’s guide to anomalous diffusion: a fractional dynamics approach,” *Physics Reports*, vol. 339, no. 1, pp. 1–77, 2000.
- [4] J.-H. Jeon, V. Tejedor, S. Burov, E. Barkai, C. Selhuber-Unkel, K. Berg-Sørensen, L. Oddershede, and R. Metzler, “In vivo anomalous diffusion and weak ergodicity breaking of lipid granules,” *Phys. Rev. Lett.*, vol. 106, p. 048103, Jan 2011.
- [5] I. Izeddin, V. Récamier, L. Bosanac, I. I. Cissé, L. Boudarene, C. Dugast-Darzacq, F. Proux, O. Bénichou, R. Voituriez, O. Bensaude, M. Dahan, and X. Darzacq, “Single-molecule tracking in live cells reveals distinct target-search strategies of transcription factors in the nucleus,” *eLife*, vol. 3, p. e02230, june 2014.
- [6] D. Krapf, “Chapter five - mechanisms underlying anomalous diffusion in the plasma membrane,” in *Lipid Domains* (A. K. Kenworthy, ed.), vol. 75 of *Current Topics in Membranes*, pp. 167–207, Academic Press, 2015.
- [7] M. Woringer and X. Darzacq, “Protein motion in the nucleus: from anomalous diffusion to weak interactions.,” *Biochemical Society transactions*, vol. 46, pp. 945–956, Aug. 2018.
- [8] A. Sabri, X. Xu, D. Krapf, and M. Weiss, “Elucidating the origin of heterogeneous anomalous diffusion in the cytoplasm of mammalian cells,” *Phys. Rev. Lett.*, vol. 125, p. 058101, Jul 2020.

- [9] Y. Meroz and I. M. Sokolov, “A toolbox for determining subdiffusive mechanisms,” *Physics Reports*, vol. 573, pp. 1–29, 2015. A toolbox for determining subdiffusive mechanisms.
- [10] J. Li, “Role of ergodicity, aging, and gaussianity in resolving the origins of biomolecule subdiffusion,” *Phys. Chem. Chem. Phys.*, vol. 24, pp. 16050–16057, 2022.
- [11] F. Höfling and T. Franosch, “Anomalous transport in the crowded world of biological cells,” *Reports on Progress in Physics*, vol. 76, no. 4, p. 046602, 2013.
- [12] M. Woringer, I. Izeddin, C. Favard, and H. Berry, “Anomalous diffusion in living cells: bridging the gap between experiments and models through collaborative challenges,” *Frontiers in Physics*, vol. 8, p. 134, 2020.
- [13] C. Manzo and M. F. Garcia-Parajo, “A review of progress in single particle tracking: from methods to biophysical insights,” *Reports on progress in physics*, vol. 78, no. 12, p. 124601, 2015.
- [14] H. Shen, L. J. Tauzin, R. Baiyasi, W. Wang, N. Moringo, B. Shuang, and C. F. Landes, “Single particle tracking: from theory to biophysical applications,” *Chemical reviews*, vol. 117, no. 11, pp. 7331–7376, 2017.
- [15] H.-J. Cheng, C.-H. Hsu, C.-L. Hung, and C.-Y. Lin, “A review for cell and particle tracking on microscopy images using algorithms and deep learning technologies,” *biomedical journal*, vol. 45, no. 3, pp. 465–471, 2022.
- [16] M. Fournier, P. Leclerc, A. Leray, D. Champelovier, F. Agbazahou, F. Dahmani, G. Bidaux, A. Furlan, and L. Hélot, “Combined SPT and FCS methods reveal a mechanism of RNAP II oversampling in cell nuclei,” *Scientific Reports*, vol. 13, p. 14633, Sept. 2023.
- [17] V. Briane, C. Kervrann, and M. Vimond, “Statistical analysis of particle trajectories in living cells,” *Phys. Rev. E*, vol. 97, p. 062121, Jun 2018.
- [18] A. Weron, J. Janczura, E. Boryczka, T. Sungkaworn, and D. Calebiro, “Statistical testing approach for fractional anomalous diffusion classification,” *Phys. Rev. E*, vol. 99, p. 042149, Apr 2019.
- [19] J. Krog, L. H. Jacobsen, F. W. Lund, D. Wüstner, and M. A. Lomholt, “Bayesian model selection with fractional brownian motion,” *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2018, p. 093501, sep 2018.
- [20] S. Thapa, M. A. Lomholt, J. Krog, A. G. Cherstvy, and R. Metzler, “Bayesian analysis of single-particle tracking data using the nested-sampling algorithm: maximum-likelihood model selection applied to stochastic-diffusivity data,” *Phys. Chem. Chem. Phys.*, vol. 20, pp. 29018–29037, 2018.
- [21] J. Janczura, P. Kowalek, H. Loch-Olszewska, J. Szwabiński, and A. Weron, “Classification of particle trajectories in living cells: Machine learning versus statistical testing hypothesis for fractional anomalous diffusion,” *Phys. Rev. E*, vol. 102, p. 032402, Sep 2020.
- [22] H. Loch-Olszewska and J. Szwabiński, “Impact of feature choice on machine learning classification of fractional anomalous diffusion,” *Entropy*, vol. 22, no. 12, 2020.
- [23] N. Granik, L. E. Weiss, E. Nehme, M. Levin, M. Chein, E. Perelson, Y. Roichman, and Y. Shechtman, “Single-particle diffusion characterization by deep learning,” *Biophysical Journal*, vol. 117, no. 2, pp. 185–192, 2019.
- [24] D. Han, N. Korabel, R. Chen, M. Johnston, A. Gavrilova, V. J. Allan, S. Fedotov, and T. A. Waigh, “Deciphering anomalous heterogeneous intracellular transport with neural networks,” *eLife*, vol. 9, p. e52224, mar 2020.
- [25] O. Krichevsky and G. Bonnet, “Fluorescence correlation spectroscopy: the technique and its applications,” *Reports on Progress in Physics*, vol. 65, no. 2, p. 251, 2002.
- [26] E. L. Elson, “Fluorescence correlation spectroscopy: past, present, future,” *Biophysical journal*, vol. 101, no. 12, pp. 2855–2870, 2011.

- [27] Y. Jiang, B. Xu, A. Melnykov, G. M. Genin, and E. L. Elson, “Fluorescence correlation spectroscopy and photon counting histograms in finite, bounded domains,” *Biophysical Journal*, vol. 119, no. 2, pp. 265–273, 2020.
- [28] M. Weiss, H. Hashimoto, and T. Nilsson, “Anomalous protein diffusion in living cells as seen by fluorescence correlation spectroscopy,” *Biophysical journal*, vol. 84, pp. 4043–52, 06 2003.
- [29] K. A. Rose, M. Molaei, M. J. Boyle, D. Lee, J. C. Crocker, and R. J. Composto, “Particle tracking of nanoparticles in soft matter,” *Journal of Applied Physics*, vol. 127, no. 19, 2020.
- [30] M. D. Stolle and C. Fradin, “Anomalous Diffusion in Inverted Variable-Lengthscale Fluorescence Correlation Spectroscopy,” *Biophysical Journal*, vol. 116, pp. 791–806, Mar. 2019.
- [31] B. B. Mandelbrot, “A fast fractional gaussian noise generator,” *Water Resources Research*, vol. 7, no. 3, pp. 543–553, 1971.
- [32] J. H. Schulz, E. Barkai, and R. Metzler, “Aging renewal theory and application to random walks,” *Physical Review X*, vol. 4, no. 1, p. 011028, 2014.
- [33] S. Havlin and D. Ben-Avraham, “Diffusion in disordered media,” *Adv. Phys.*, vol. 36, pp. 695–798, 1987.
- [34] E. Sezgin, F. Schneider, S. Galiani, I. Urbancic, D. Waithe, B. C. Lagerholm, and C. Eggeling, “Measuring nanoscale diffusion dynamics in cellular membranes with super-resolution sted-fcs,” *Nature Protocols*, vol. 14, p. 1, 03 2019.
- [35] C. Eggeling, C. Ringemann, R. Medda, G. Schwarzmann, K. Sandhoff, S. Polyakova, V. Belov, B. Hein, C. Middendorff, A. Schönle, and S. Hell, “Direct observation of the nanoscale dynamics of membrane lipids in a living cell,” *Nature*, vol. 457, pp. 1159–62, 12 2008.
- [36] L. Yu, Y. Lei, Y. Ma, M. Liu, J. Zheng, D. Dan, and P. Gao, “A comprehensive review of fluorescence correlation spectroscopy,” *Frontiers in physics*, vol. 9, p. 644450, 2021.
- [37] G. Muñoz-Gil and al., “Objective comparison of methods to decode anomalous diffusion,” *Nature Communications*, vol. 12, 2021.
- [38] E. Dauty and A. Verkman, “Molecular crowding reduces to a similar extent the diffusion of small solutes and macromolecules: Measurement by fluorescence correlation spectroscopy,” *Journal of molecular recognition : JMR*, vol. 17, pp. 441–7, 09 2004.
- [39] S. T. Hess and W. W. Webb, “Focal Volume Optics and Experimental Artifacts in Confocal Fluorescence Correlation Spectroscopy,” *Biophysical Journal*, vol. 83, pp. 2300–2317, Oct. 2002.
- [40] A. G. Cherstvy and R. Metzler, “Population splitting, trapping, and non-ergodicity in heterogeneous diffusion processes,” *Phys. Chem. Chem. Phys.*, vol. 15, pp. 20220–20235, 2013.
- [41] J.-H. Jeon, A. V. Chechkin, and R. Metzler, “Scaled brownian motion: a paradoxical process with a time dependent diffusivity for the description of anomalous diffusion,” *Phys. Chem. Chem. Phys.*, vol. 16, pp. 15811–15817, 2014.
- [42] S. M. A. Tabei, S. Burov, H. Y. Kim, A. Kuznetsov, T. Huynh, J. Jureller, L. H. Philipson, A. R. Dinner, and N. F. Scherer, “Intracellular transport of insulin granules is a subordinated random walk,” *Proceedings of the National Academy of Sciences*, vol. 110, no. 13, pp. 4911–4916, 2013.
- [43] P. J. Brockwell and R. A. Davis, *Time series: theory and methods*. Springer science & business media, 2009.
- [44] J. D. Hamilton, *Time series analysis*. Princeton university press, 2020.
- [45] S. Lambert-Lacroix, *Fonction d’autocorrélation partielle des processus à temps discret non stationnaires et applications*. Theses, Université Joseph-Fourier - Grenoble I, July 1998.

- [46] S. Degerine and S. Lambert, “Evolutive instantaneous spectrum associated with the partial autocorrelation function for nonstationary time series,” in *Proceedings of Third International Symposium on Time-Frequency and Time-Scale Analysis (TFTS-96)*, pp. 457–460, 1996.
- [47] S. Dégerine and S. Lambert-Lacroix, “Characterization of the partial autocorrelation function of nonstationary time series,” *Journal of Multivariate Analysis*, vol. 87, no. 1, pp. 46–59, 2003.
- [48] J. A. Dix, E. F. Hom, and A. Verkman, “Fluorescence correlation spectroscopy simulations of photophysical phenomena and molecular interactions: a molecular dynamics/monte carlo approach,” *The Journal of Physical Chemistry B*, vol. 110, no. 4, pp. 1896–1906, 2006.
- [49] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, “Scikit-learn: Machine learning in Python,” *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
- [50] N.-S. Cheng, “Formula for the Viscosity of a Glycerol-Water Mixture,” *Industrial & Engineering Chemistry Research*, vol. 47, no. 9, pp. 3285–3288, 2008.
- [51] A. Volk and C. J. Kähler, “Density model for aqueous glycerol solutions,” *Experiments in Fluids*, vol. 59, p. 75, Apr. 2018.
- [52] L. W. Bosart and A. O. Snoddy, “Specific Gravity of Glycerol ¹,” *Industrial & Engineering Chemistry*, vol. 20, pp. 1377–1379, Dec. 1928.
- [53] D. T. Gillespie, “The mathematics of brownian motion and johnson noise,” *American Journal of Physics*, vol. 64, no. 3, pp. 225–240, 1996.
- [54] J.-F. Coeurjolly, “Simulation and identification of the fractional brownian motion: a bibliographical and comparative study,” *Journal of statistical software*, vol. 5, pp. 1–53, 2000.
- [55] T. Dieker, *Simulation of fractional Brownian motion*. PhD thesis, Masters Thesis, Department of Mathematical Sciences, University of Twente . . . , 2004.
- [56] F. Mainardi, M. Raberto, R. Gorenflo, and E. Scalas, “Fractional calculus and continuous-time finance ii: the waiting-time distribution,” *Physica A: Statistical Mechanics and its Applications*, vol. 287, no. 3-4, pp. 468–481, 2000.
- [57] E. W. Montroll and G. H. Weiss, “Random Walks on Lattices. II,” *Journal of Mathematical Physics*, vol. 6, pp. 167–181, Feb. 1965.
- [58] G. Baumann, R. F. Place, and Z. Földes-Papp, “Meaningful Interpretation of Subdiffusive Measurements in Living Cells (Crowded Environment) by Fluorescence Fluctuation Microscopy,” *Current Pharmaceutical Biotechnology*, vol. 11, p. 527, Aug. 2010.
- [59] M. R. Horton, F. Höfling, J. O. Rädler, and T. Franosch, “Development of anomalous diffusion among crowding proteins,” *Soft Matter*, vol. 6, no. 12, pp. 2648–2656, 2010.
- [60] T. L. J. Ng and A. Zammit-Mangion, “Non-homogeneous poisson process intensity modeling and estimation using measure transport,” *Bernoulli*, vol. 29, no. 1, pp. 815–838, 2023.
- [61] J. F. C. Kingman, *Poisson processes*, vol. 3. Clarendon Press, 1992.
- [62] P. W. Lewis and G. S. Shedler, “Simulation of nonhomogeneous poisson processes by thinning,” *Naval research logistics quarterly*, vol. 26, no. 3, pp. 403–413, 1979.