PAPER • OPEN ACCESS

Single trajectory characterization via machine learning

To cite this article: Gorka Muñoz-Gil *et al* 2020 *New J. Phys.* **22** 013010

View the article online for updates and enhancements.

New Journal of Physics

The open access journal at the forefront of physics

PAPER

Single trajectory characterization via machine learning

Gorka Muñoz-Gil[1,6] , **Miguel Angel Garcia-March**[1,2] , **Carlo Manzo**[3,6] , **José D Martín-Guerrero**[4] and **Maciej Lewenstein**[1,5]

1   ICFO—Institut de Ciències Fotòniques, The Barcelona Institute of Science and Technology, E-08860 Castelldefels (Barcelona), Spain
2   Instituto Universitario de Matemática Pura y Aplicada, Universitat Politècnica de València, E-46022 València, Spain
3   Facultat de Ciències i Tecnologia, Universitat de Vic—Universitat Central de Catalunya (UVic-UCC), C. de la Laura,13, E-08500 Vic, Spain
4   Universitat de València—IDAL, Department of Electronic Engineering, ETSE-UV, Avgda. Universitat, s/n E-46100 Burjassot (Valencia), Spain
5   ICREA, Lluis Companys 23, E-08010 Barcelona, Spain
6   Author to whom any correspondence should be addressed.

E-mail: gorka.munoz@icfo.es and carlo.manzo@uvic.cat

Abstract

In order to study transport in complex environments, it is extremely important to determine the physical mechanism underlying diffusion and precisely characterize its nature and parameters. Often, this task is strongly impacted by data consisting of trajectories with short length (either due to brief recordings or previous trajectory segmentation) and limited localization precision. In this paper, we propose a machine learning method based on a random forest architecture, which is able to associate single trajectories to the underlying diffusion mechanism with high accuracy. In addition, the algorithm is able to determine the anomalous exponent with a small error, thus inherently providing a classification of the motion as normal or anomalous (sub- or super-diffusion). The method provides highly accurate outputs even when working with very short trajectories and in the presence of experimental noise. We further demonstrate the application of transfer learning to experimental and simulated data not included in the training/test dataset. This allows for a full, high-accuracy characterization of experimental trajectories without the need of any prior information.

In the last decades, the research in biophysics has conveyed large efforts toward the development of experimental techniques allowing the visualization of biological processes one molecule at a time [1–4]. These efforts have been mainly driven by the concept that ensemble-averaging hides important features that are relevant for cellular function. Somehow expectedly, experiments performed by means of these techniques have shown a large heterogeneity in the behavior of biological molecules, thus fully justifying the use of these raffinate tools. Besides, experiments performed using single particle tracking [3] have revealed that even chemically-identical molecules in biological media can display very different behaviors, as a consequence of the complex environment where diffusion takes place. By way of example, this heterogeneity is reflected in the broad distribution of dynamic parameters of distinct individual trajectories corresponding to the same molecular species, such as the diffusion coefficient, well above stochastic indetermination. Typically, the trajectories are analyzed by quantifying the (time-averaged) mean square displacement (tMSD) as a function of the time lag τ [5]:

$$\overline{\delta^2(\tau)} = \frac{1}{t_{max} - \tau} \int_0^{t_{max}-\tau} [x(t' + \tau) - x(t')]^2 \mathrm{d}t'. \tag{1}$$

The calculation of this quantity—expected to scale linearly for a Brownian walker in a homogeneous environment—has provided a ubiquitous evidence of anomalous behaviors in biological systems, characterized by an asymptotic nonlinear scaling of the tMSD curve $\overline{\delta^2} \sim \tau^\alpha$. More experiments have shown that the anomalous exponent can vary from particle to particle (figure 1(a)) as a consequence of molecular interactions and that these changes can be experienced by the same particle in space/time [6]. Several methods have been
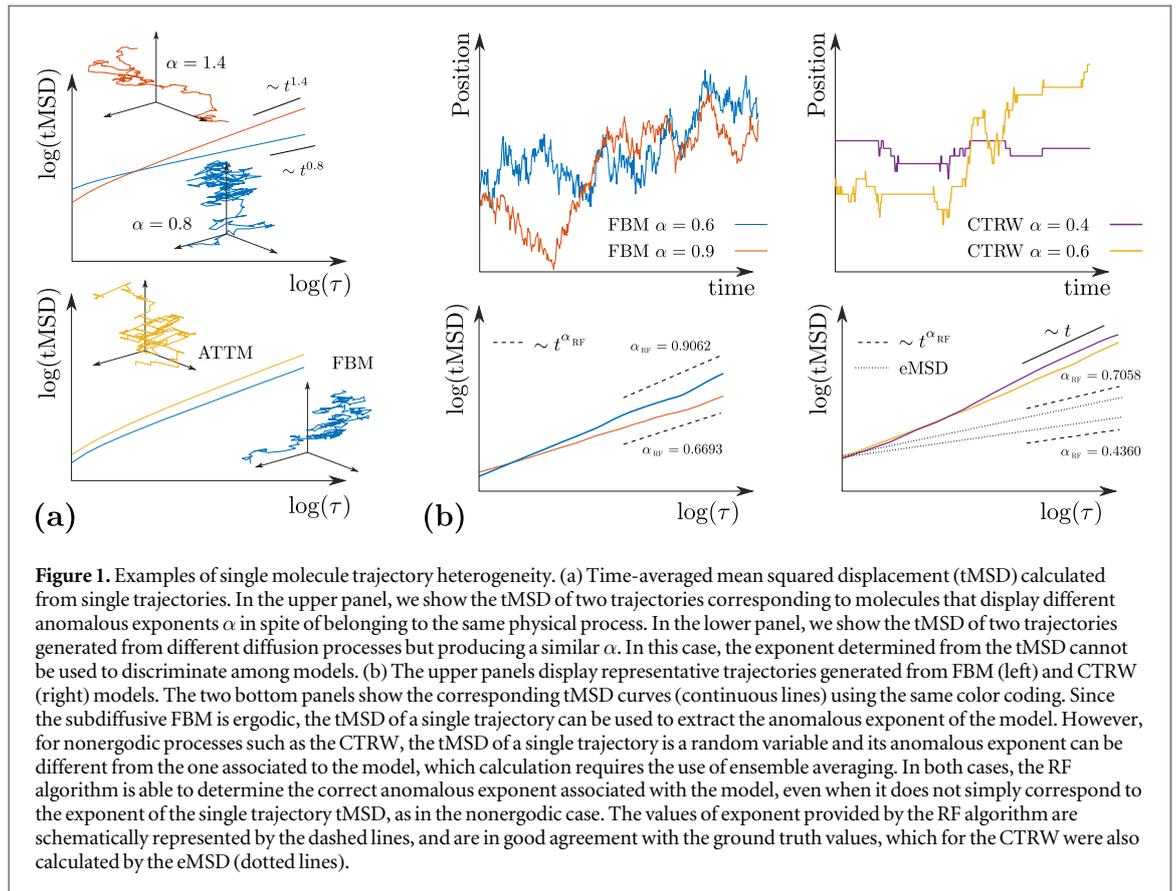
Figure 1. Examples of single molecule trajectory heterogeneity. (a) Time-averaged mean squared displacement (tMSD) calculated from single trajectories. In the upper panel, we show the tMSD of two trajectories corresponding to molecules that display different anomalous exponents α in spite of belonging to the same physical process. In the lower panel, we show the tMSD of two trajectories generated from different diffusion processes but producing a similar α. In this case, the exponent determined from the tMSD cannot be used to discriminate among models. (b) The upper panels display representative trajectories generated from FBM (left) and CTRW (right) models. The two bottom panels show the corresponding tMSD curves (continuous lines) using the same color coding. Since the subdiffusive FBM is ergodic, the tMSD of a single trajectory can be used to extract the anomalous exponent of the model. However, for nonergodic processes such as the CTRW, the tMSD of a single trajectory is a random variable and its anomalous exponent can be different from the one associated to the model, which calculation requires the use of ensemble averaging. In both cases, the RF algorithm is able to determine the correct anomalous exponent associated with the model, even when it does not simply correspond to the exponent of the single trajectory tMSD, as in the nonergodic case. The values of exponent provided by the RF algorithm are schematically represented by the dashed lines, and are in good agreement with the ground truth values, which for the CTRW were also calculated by the eMSD (dotted lines).
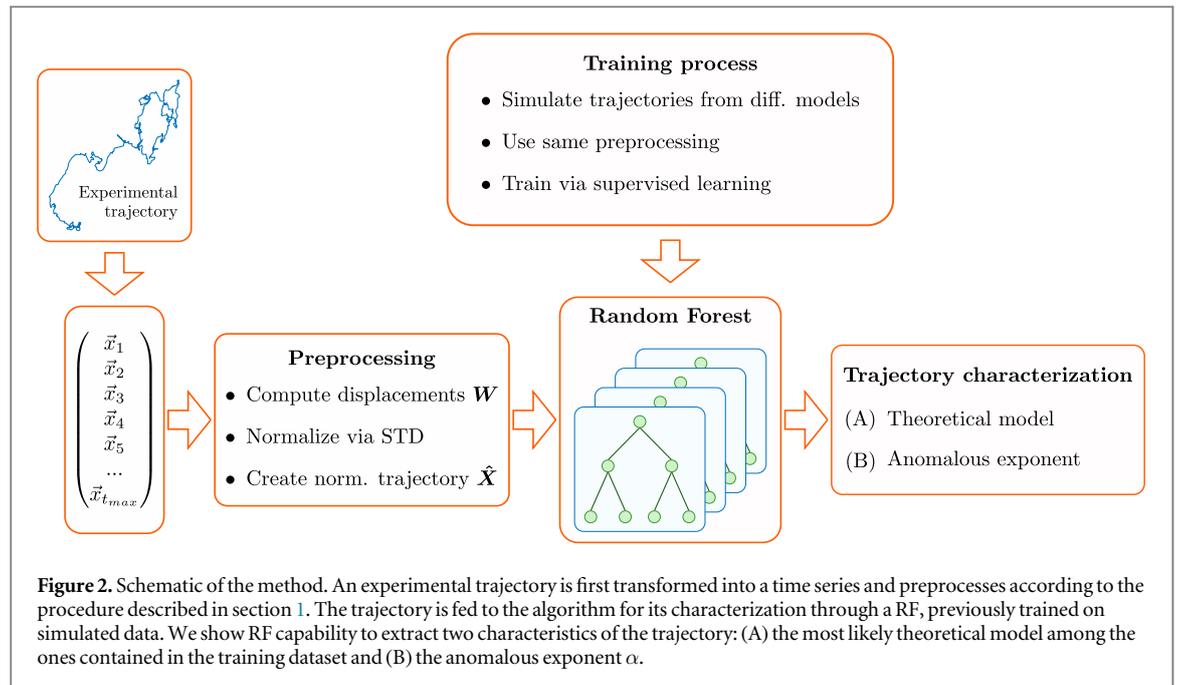
proposed to accurately estimate this exponent for single trajectories [7, 8] in the presence of experimental limitations, such as optical diffraction and the finite length of the trajectory.

In some cases, this heterogeneity leads to exotic effects, such as the breaking of ergodicity observed in several cellular systems [9–12]. Nonergodicity implies the nonequivalence of time and ensemble averages of the mean squared displacement (MSD). In the nonergodic case, $\overline{\delta^2(\tau)}$ remains random even in the long measurement times, i.e. the diffusion coefficients are irreproducible but the distribution of the tMSD is universal [13]. For nonergodic models, the determination of the anomalous exponent at the single trajectory level does not fully characterize the model, since the tMSD can have different scaling when averaged with respect to the time or to the ensemble (figure 1(b)). Therefore, it requires the calculation of the ensemble-averaged MSD (eMSD) over a rather large number of trajectories [5].

The emergence of anomalous behavior has also been widely studied from the theoretical point of view and conceptually-different models have been proposed [5]. However, the fact that models with different physical properties can produce the same tMSD exponent, strongly limits the unambiguous determination of the underlying dynamics, based only on the evaluation of the tMSD (figure 1(a)). In order to solve this ambiguity, a large effort has been made to classify experimental data showing anomalous transport. As an example, the use of alternative estimators [14, 15] has been proposed to determine whether the pioneering results of Golding and Cox [16] were arising from a continuous-time random walk (CTRW) [17] or fractional Brownian motion (FBM) [18]. This search for a better classification between CTRW and FBM often relied in the determination of the (non)ergodicity of the data [15, 19–21], since CTRW is consistent with weak ergodicity breaking [22]. The experimental evidence of nonergodic behavior has boosted the proposal of new theoretical frameworks consistent with these features [23, 24].

In this scenario, determining whether a single-molecule trajectory (or a short segment of it) displays normal or anomalous behavior by its tMSD scaling exponent and associating the motion to a specific physical model are elements of paramount importance to gain insight about the biophysical mechanism underlying anomalous diffusion, thus providing a detailed picture of a variety of phenomena. Recent works in this direction have focused on classification schemes based on optimization procedures [8], power spectral density [25], or Bayesian approaches [26, 27].

Surprisingly, in spite of the fast rise of machine learning methods, little efforts have been made to characterize single trajectories. A few attempts in this sense have been mainly directed to qualitatively

Figure 2. Schematic of the method. An experimental trajectory is first transformed into a time series and preprocesses according to the procedure described in section 1. The trajectory is fed to the algorithm for its characterization through a RF, previously trained on simulated data. We show RF capability to extract two characteristics of the trajectory: (A) the most likely theoretical model among the ones contained in the training dataset and (B) the anomalous exponent α.

discriminate among confined, anomalous, normal or directed motion [28, 29], without extracting quantitative information of classifying with respect to the underlying physical model.

This paper presents a machine learning algorithm based on the Random forest (RF) architecture that efficiently and robustly characterizes single trajectories at different levels: first, obtaining the discrimination among several diffusion models; then, providing the estimation of the exponent that characterizes the anomalous diffusion, thus inherently classifying between normal and anomalous (sub- and super-) diffusion. The algorithm allows to accurately tackle these challenging problems even when dealing with short and noisy trajectories.

In comparison to previously reported methods, the algorithm we present in this paper allows for the characterization of a trajectory without prior knowledge about the process from which the single trajectory has been extracted. Most of the works existent in the literature focus their studies on anomalous ergodic trajectories (usually FBM related processes), whereas our method is robust against the appearance of nonergodicity. While showing similar accuracy as others on ergodic trajectories [8, 27], to the best of our knowledge, our method represents the first attempt to extract the anomalous exponent for nonergodic processes through single-trajectory characterization.

1. Machine learning method

In this section, we outline the main parts of the trajectory characterization algorithm, consisting of: the machine learning algorithm that takes the form of a RF architecture; the simulated dataset; and the preprocessing applied to the dataset before being analyzed by the RF. Figure 2 shows a schematic representation of the pipeline.

1.1. Random forest

RF is an architecture based on decision trees. A decision tree is an efficient non-parametric method widely used for classification and regression problems [30]. The basic idea consists in producing recursive binary splits of the input space, so that the samples with the same label are grouped together. The criterion to produce the splits is based on a homogeneity measure (usually, the information entropy) of the target variable within each of the obtained groups. In regression problems, a commonly used criterion is to select the split that minimizes the mean squared error; this recursive process continues until some stopping rule is satisfied, e.g. a common one is to consider that a tree node can be split if it contains more than a given number of samples; therefore, the minimum number of samples required to split a tree node should be adjusted in order to control the size of the tree, thus preventing overfitting. Once a decision tree is obtained, the output for unseen samples is computed just passing them through the nodes of the tree, where a decision is made with respect to which direction to take. Finally, a terminal tree node is reached, where the output is obtained.

A RF is a tree-based ensemble method, which builds several decision tree models independently and then computes a final prediction by combining the outputs of the different individual trees [31]. In particular, the

ensemble is produced with single trees built from samples drawn randomly with replacement (bootstrap) from the training set. An additional randomness is added when splitting a tree node because the split is chosen among a random subset of the input variables, selected in this case without replacement, instead of the greedy approach of considering all the input variables. Due to this randomization, the bias of the ensemble is slightly higher than that of a single tree, but the variance is decreased and the model is more robust to variations in the dataset.

RF is a very powerful, state-of-the-art technique for both regression and classification problems, usually outperforming not only single decision trees but also sophisticated models, as shown in a thorough comparison study [32].

1.2. Training and test datasets

The training dataset is built out of numerical simulations of trajectories from various kinds of theoretical models. As a natural choice, we included three of the best-known and used models that can give rise to anomalous diffusion: CTRW [17], FBM [18] and Lévy walks (LW) [33]. In addition, we included trajectories from the annealed transient time model (ATTM) [23]. In the ATTM, a diffuser performs a random walk but stochastically changes the diffusion coefficient at random times. Both the diffusion coefficient and the time at which the diffusivity changes are drawn from distributions with a power law behavior [23]. Its time-averaged MSD shows a linear scaling, but the model has a regime in which it displays weak ergodicity breaking. The ATTM has been shown to reproduce the features observed for the diffusion of a cell membrane receptor [12], one of the experimental datasets analyzed.

1.3. Preprocessing

Our aim is to design a method that can be used to accurately characterize heterogeneous trajectories without having to calculate other parameters or using *a priori* knowledge. In order to be able to analyze data coming from any possible spatiotemporal scale, we designed a preprocessing procedure that properly rescale the data. We implemented the following procedure, chosen among other plausible normalization techniques as it gives rise to the best results in terms of accuracy:

1. We use one of the models above to simulate the trajectory of a particle during t_{max} time steps. The result is a vector of positions, $X = (x_1, x_2, \ldots, x_{t_{max}})$.

2. This vector is transformed into a vector of distances traveled in an interval of time T_{lag}, i.e. $W = (\Delta x_1, \Delta x_2, \ldots, \Delta x_{J-1})$,, where $J = t_{max}/T_{lag}$. We define Δx_i as

$$\Delta x_i = |x_{iT_{lag}} - x_{(i+1)T_{lag}}|. \tag{2}$$

3. To normalize the data, we divide W by its standard deviation to get a new vector $\hat{W}$.

4. Then, we do a cumulative sum of $\hat{W}$ to construct a normalized trajectory $\hat{X}$.

Summarizing, the previous procedure generates a new trajectory which is constructed via the normalized displacements of the original trajectory. This makes that the magnitudes of the resulting trajectories are comparable, no matter what were their original values. While the RF could be trained using $\hat{W}$, our results show that training with $\hat{X}$ gives indeed much better results. The same preprocessing is applied to both the simulated and experimental trajectories used in sections 2 and 3.

2. Trajectory characterization as a machine learning problem

We will use our method to characterize single trajectories according to two different schemes: (A) discrimination among diffusion models; (B) prediction of the anomalous exponent α, that inherently implies classification as normal or anomalous diffusion. For each of these problems, we created a dataset of $1, 2 \cdot 10^5$ trajectories with $t_{max} = 10^3$, divided into a training and test set with ratio 0.8/0.2, respectively. The results presented in all the figures and the values of the accuracy discussed in the text correspond to the ones measured in the test set, ensuring that the RF does not present overfitting in any of the problems considered. The different classes considered in each problem have an equal number of trajectories, hence allowing us to use the accuracy as a measure of the goodness of the RF. For technical details and an practical example of the implementation, we refer the reader to the appendix and to the code repository [34].

Figure 3. Benchmarking the RF algorithm: (a) accuracy of the RF when discriminating among models as a function of the preprocessing parameter T_{lag}. Blue triangles = CTRW versus FBM, Red circles = CTRW, LW, FBM, ATTM, and Yellow squares = CTRW versus ATTM. (b) Accuracy of the model discrimination as a function of the length of the trajectories t_{max}. (c) Histograms of the error in the prediction of the anomalous exponent for different values of the length trajectory t_{max} and noise variance σ_n. Y-axis is percentage of trajectories $\bar{N}$ with given error ε when predicting the value of α. Each histogram corresponds to the highlighted points in (d) and (e). (d) MAE of the anomalous exponent prediction as a function of the length of the trajectories t_{max}. (e) MAE in anomalous exponent prediction as a function of the variance of the normal noise variance. The RF was trained with a dataset without noise and then asked to classify trajectory of such models with normal noise given by (3). $T_{lag} = 0$, i.e. raw trajectories, were used for (b)–(e).

2.1. Discrimination among diffusion models

In order to predict the diffusion model underlying a certain trajectory, we construct a RF whose input is the normalized trajectory $\hat{X}$, and the output is a number between 0 and $N - 1$ corresponding to the different models, with N the total number of models used in the training. Figure 3(a) shows the accuracy of the RF. Each line corresponds to a training dataset made up of different models. In the absence of data preprocessing (point marked as 'Raw' in the x-axis), the RF shows large accuracy. The accuracy drops significantly as T_{lag} increases, likely as a consequence of the removal of microscopical properties of the model, such as short-time correlations, hence preventing the RF from learning important features of them. This might lead to the conclusion that the filtering introduced by the preprocessing steps only limits the time resolution. This is obviously true for simulated data, obtained at the same scale, for which preprocessing is unnecessary. However, when dealing with experimental data of unknown spatiotemporal scale, such a preprocessing is of fundamental importance to be able to apply the same architecture and training dataset, in spite of the little loss of performance.

In addition, the accuracy heavily depends on similarities among the models to be discriminated. For example, the accuracy obtained with a RF trained only with trajectories reproducing conceptually different models such as FBM and CTRW (triangular markers in figure 3(a)) is higher than the one obtained when including in the training models with similar characteristics, such as CTRW and ATTM (red circles and yellow squares in figure 3(a)).

2.2. Anomalous exponent estimation

A first approximation toward the characterization of the anomalous exponent can be based on a regression problem, in which the output of the RF is the value of the anomalous exponent α. The nature of the regression algorithm makes that the output of the RF is the continuous value which better satisfies the constraints learn during training.

To characterize the performance of the method, we calculate the prediction error ε of a trajectory as the absolute distance between the predicted exponent and the ground truth value. The percentage of trajectories $\bar{N}(\varepsilon)$ with a given error ε is represented in the bar plots of figure 3(c) for three different cases and a subdiffusive dataset including trajectories obtained from FBM, CTRW and ATTM. The case (i) considers trajectories with $t_{max} = 10^3$ without noise, while the cases (ii) and (iii) show results for shorter and noisy trajectories (see discussion below). For case (i) the calculated mean absolute error (MAE) of the prediction of the anomalous exponent gives a value of 0.11. Moreover, the histogram showed in figure 3(c)(i) shows that for ∼80% of the trajectories, the output exponent lies within 0.1 from the true value.

2.3. Experimental scenario: short and noisy trajectories

A remarkable feature of the method is the possibility to correctly characterize very short trajectories. In figures 3(b) and (d), we show the ability of the RF to characterize short trajectories. In figure 3(b), we plot the accuracy in model discrimination as a function of the length of the trajectories, t_{max}. In figure 3(d), a similar study is done, now tracking the MAE of the RF trained to predict α. Although we observe an expected decrease of

Table 1. Process discrimination for the datasets considered in section 3. Shown is the percentage of trajectories classified as associated to each model. The results for (i) were done with $T_{lag} = 0$ and for datasets (ii) and (iii) with $T_{lag} = 1$.

Dataset	Predicted model		
	CTRW	FBM	ATTM
(i) Compartments model	89.2%	0	10.7%
(ii) Experiment 1	4.5%	86.6%	8.9%
(iii) Experiment 2	16.4%	33.2%	50.4%

performance for short trajectories, both plots show that the RF is able to characterize trajectories as short as only 10 points. Quantitatively, when comparing trajectories of 10 points with larger ones, of 1000 points, the model discrimination accuracy only decreases by a factor of 8.2%, while the MAE decreases by a factor of 18%. Panel (ii) in figure 3(e) shows the error distribution when predicting α for $t_{max} = 100$.

Importantly, one has also to take into account that the experimental trajectories have a limited localization precision, that results into Gaussian noise. Therefore, it is necessary to test the robustness to noise of the RF. To this end, we trained the RF with trajectories simulated as described before and then we try to predict the anomalous exponent of trajectories belonging to the same dataset, but whose positions X were perturbed with noise to obtain the dataset $X^{(n)}$

$$x_i^{(n)} = x_i + \mu_i(\mu, \sigma_n), \tag{3}$$

where $\mu_i(\mu, \sigma_n)$ is a random number retrieved from a Normal distribution with mean $\mu = 0$ and variance σ_n. The results obtained for training with FBM, CTRW and ATTMs are presented in figure 3(d). The RF shows a great robustness against noise. For $\sigma_n < 1$, the MAE appears almost unaffected. When increasing σ_n, we see that the MAE increases, as expected, but even for large σ_n the MAE is still reasonable.

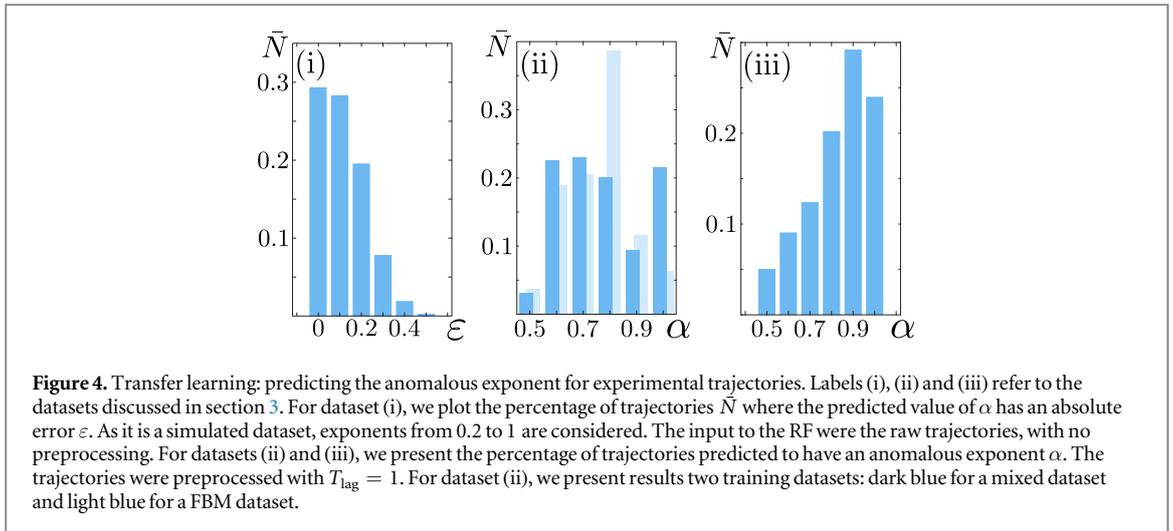## 3. Transfer learning in simulated and experimental data

To further show the advantages of our machine learning algorithm, we applied it to three sets of trajectories different from those included in the training/test dataset. This is often referred as transfer learning, as certain architecture is trained in one setting and then applied to a different one. For this, we will consider three datasets:

(i) Simulated data coming from a recently presented model [35], describing the movement of a diffuser in a network of compartments of random size and random permeability, both drawn from universal distributions. This model shares the same subordination as the quenched trap model, i.e. a CTRW with power-law distributed trapping times and recapitulates the complexity and heterogeneity found in some biological environments. This choice allows to test the algorithm over a conceptually different model with respect to the training dataset, while having the advantage of tuning the value of anomalous exponents.

(ii) Experiment 1, reporting the motion of individual mRNA molecules inside live bacterial cells [16]. The tMSD shows anomalous diffusion with $\alpha \sim 0.7$; this behavior has been associated to FBM [14, 36].

(iii) Experiment 2, corresponding to a set of trajectories obtained for the diffusion of a membrane receptor in living cells [12]. Although the time-averaged MSD shows a nearly linear behavior, the data present features of ergodicity breaking due to changes of diffusivity [37] and have been associated to the ATTM model.

Following the scheme presented in figure 2, first we train the RF with simulated trajectories obtained with different theoretical frameworks. It should be noted that for this section, since we deal with trajectories that do not show superdiffusive behavior, we do not include the Lévy walks process in the training dataset.

3.1. Results
Following the same structure of the previous section, we start by discriminating the diffusion model that can be associated to datasets (i)–(iii). The results are reported in table 1, showing a high rate of correct classification for the dataset (i). For the experimental data in datasets (ii) and (iii), we do not dispose of ground truth values, thus we compare our results with those of previous analysis, performed with alternative methods. For the trajectories of Experiment 1, we found that the algorithm largely assign them to the FBM, in strong agreement with previously reported results based on the concept of variation [14]. The data of Experiment 2 are mainly assigned to the ATTM model. This model was shown to reproduce features observed in these data, such as subdiffusion

Figure 4. Transfer learning: predicting the anomalous exponent for experimental trajectories. Labels (i), (ii) and (iii) refer to the datasets discussed in section 3. For dataset (i), we plot the percentage of trajectories $\bar{N}$ where the predicted value of α has an absolute error ε. As it is a simulated dataset, exponents from 0.2 to 1 are considered. The input to the RF were the raw trajectories, with no preprocessing. For datasets (ii) and (iii), we present the percentage of trajectories predicted to have an anomalous exponent α. The trajectories were preprocessed with $T_{\text{lag}} = 1$. For dataset (ii), we present results two training datasets: dark blue for a mixed dataset and light blue for a FBM dataset.

and weak ergodicity breaking [12]. Moreover, a little fraction of trajectories are classified as CTRW. As previously mentioned, CTRW and ATTM share similar features (such as time subordination), increasing the difficulty in discriminating between them. This appears to be the main source of error in the results.

To obtain further insights on the study of the diffusion, we used the RF to extract the anomalous exponents. For the first dataset (i), based on simulations, we generated trajectories having a broad range of subdiffusive trajectories, namely $\alpha \in [0.2, 1]$. Then, we used the trained RF to predict the value of the anomalous exponent and evaluated the error as the absolute value of the difference between the actual and predicted α. The results are reported in the histogram of figure 4(i) and display a distribution similar to the one obtained for the training/testing data. Thus, we run the same procedure on the experimental data. For the two datasets, in figures 4(ii)–(iii) we report the values obtained for the anomalous exponent α. The histogram of the α obtained for the trajectories of Experiment 1 (dark blue) shows mainly subdiffusive values, peaked in the range 0.6– 0.8. This is in good agreement with the original paper [16], where α was estimated by means of two different approaches as 0.7 and 0.77. However, the method also classifies a percentage of the trajectories as having $\alpha = 1$. Importantly, the performance of the method can be further improved by taking advantage of the results of the model discrimination discussed above and shown in table 1. In fact, when the latter classification indicates that most of the trajectories follow a specific diffusive model, one can train the algorithm with a dataset composed only of trajectories simulated with that model. This kind of training produces exponent values in the same range, but largely reduce the fraction of those associated to $\alpha = 1$, as shown in figure 4(ii) (light blue).

Last, in figure 4(iii) we plot the distribution of exponents obtained for the Experiment 2. The subdiffusive values show a large number of occurrences in the 0.8–0.9 range, compatible with the exponent 0.84 calculated in previous studies [12]. Noteworthy, due to the nonergodic nature of the data, in the original paper α could only be calculated from the ensemble-averaged MSD, whereas the RF is able to determine this exponent from single trajectories.

4. Conclusions

We have presented a machine learning method, based on a RF architecture, which is capable to analyze a single trajectory and to determine the theoretical model that describes it at best. Moreover the same method is used for predicting its anomalous exponent with high accuracy, and thus classify the motion as normal or anomalous. The method does not need any prior information over the nature of the system from which the trajectory is obtained. It acts as a blackbox, which we train with a dataset of simulated trajectories, and then it is used to characterize the trajectory of interest. In particular, its spatial scale is not of any relevance, as we devised a preprocessing strategy which rescales trajectories to obtain comparable estimators from very different systems. The method requires a minimal amount of information. First, because it performs extremely well even with surprisingly short trajectories. Second, because it is robust with respect to the presence of a large amount of thermal noise, and can thus be applied even with low localization precision. We showcase the suitability of our method by applying it to two experimental datasets by means of transfer learning. Overall, this method can be of large application to characterize experiments from several research areas. In contrast to other methods, it can determine the type of diffusion and the anomalous exponent also for nonergodic models, without the need of performing ensemble averages. We note that recent works show how other machine learning architectures, such as convolutional neural networks and long-short term memory neural networks, are also capable of doing single

Table A1. Hyperparameters for the training of the different models exposed in this work.

	Models in dataset	Range of α's	t_{max}
Figure 3(a)	See caption	[0.2, 2]	10^3
Figure 3(b)	CTRW/LW/FBM/ATTM	[0.2, 2]	See *x*-axis
Figure 3(c)(i)	CTRW/LW/FBM/ATTM	[0.5, 1]	10^3
Figure 3(c)(ii)	CTRW/LW/FBM/ATTM	[0.5, 1]	10^2
Figure 3(c)(iii)	CTRW/LW/FBM/ATTM	[0.5, 1]	10^3
Figure 3(d)	CTRW/LW/FBM/ATTM	[0.5, 1]	See *x*-axis
Figure 3(e)	CTRW/LW/FBM/ATTM	[0.5, 1]	10^3
Figure 4(i)	CTRW/FBM/ATTM	[0.5, 1]	10^3
Figure 4(ii) dark blue	CTRW/FBM/ATTM	[0.5, 1.2]	10^3
Figure 4(ii) light blue	FBM	[0.5, 1.2]	300
Figure 4(ii) dark blue	CTRW/FBM/ATTM	[0.5, 1.2]	200

trajectory characterization [29, 38]. The development of these methods and of other deep learning architectures may help to avoid the preprocessing procedure and could lead to increase the accuracy on the problems described in this work.

Acknowledgments

Appendix

A.1. Hyperparameters
We summarize here the different parameters used to train the RF models whose results are presented in this work. Common to all are the number of ensembled trees (100), the train/test dataset ratio (0.8/0.2) and the size of training set (1, 2 · 10^5 trajectories). For the details of each figure, see table A1.

ORCID iDs

Gorka Muñoz-Gil ● https://orcid.org/0000-0001-9223-0660
Carlo Manzo ● https://orcid.org/0000-0002-8625-0996
José D Martín-Guerrero ● https://orcid.org/0000-0001-9378-0285

References

[1] Miller H, Zhou Z, Shepherd J, Wollman A J M and Leake M C 2018 Single-molecule techniques in biophysics: a review of the progress in methods and applications *Rep. Prog. Phys.* **81** 024601
[2] Sahl S J, Hell S W and Jakobs S 2017 Fluorescence nanoscopy in cell biology *Nat. Rev. Mol. Cell Biol.* **18** 685–701
[3] Manzo C and Garcia-Parajo M F 2015 A review of progress in single particle tracking: from methods to biophysical insights *Rep. Prog. Phys.* **78** 124601
[4] Höfling F and Franosch T 2013 Anomalous transport in the crowded world of biological cells *Rep. Prog. Phys.* **76** 046602
[5] Metzler R, Jeon J-H, Cherstvy A G and Barkai E 2014 Anomalous diffusion models and their properties: non-stationarity, non-ergodicity, and ageing at the centenary of single particle tracking *Phys. Chem. Chem. Phys.* **16** 24128–64
[6] Weron A, Burnecki K, Akin E J, Solé L, Balcerek M, Tamkun M M and Krapf D 2014 Ergodicity breaking on the neuronal surface emerges from random switching between diffusive states *Sci. Rep.* **7** 5404
[7] Kepten E, Bronshtein I and Garini Y 2013 Improved estimation of anomalous diffusion exponents in single-particle tracking experiments *Phys. Rev.* E **87** 052713

[8] Burnecki K, Kepten E, Garini Y, Sikora G and Weron A 2015 Estimating the anomalous diffusion exponent for single particle tracking data with measurement errors—an alternative approach *Sci. Rep.* **5** 11306

[9] Jeon J-H, Tejedor V, Burov S, Barkai E, Selhuber-Unkel C, Berg-Sørensen K, Oddershede L and Metzler R 2011 *In vivo* anomalous diffusion and weak ergodicity breaking of lipid granules *Phys. Rev. Lett.* **106** 048103

[10] Weigel A V, Simon B, Tamkun M M and Krapf D 2011 Ergodic and nonergodic processes coexist in the plasma membrane as observed by single-molecule tracking *Proc. Natl Acad. Sci. USA* **108** 6438–43

[11] Tabei S M A, Burov S, Kim H Y, Kuznetsov A, Huynh T, Jureller J, Philipson L H, Dinner A R and Scherer N F 2013 Intracellular transport of insulin granules is a subordinated random walk *Proc. Natl Acad. Sci. USA* **110** 4911–6

[12] Manzo C, Torreno-Pina J A, Massignan P, Lapeyre G J, Lewenstein M and Garcia Parajo M F 2015 Weak ergodicity breaking of receptor motion in living cells stemming from random diffusivity *Phys. Rev.* X **5** 011021

[13] Akimoto T and Yamamoto E 2016 Distributional behavior of diffusion coefficients obtained by single trajectories in annealed transit time model *J. Stat. Mech.* 123201

[14] Magdziarz M, Weron A and Burnecki K 2009 Fractional Brownian motion versus the continuous-time random walk: a simple test for subdiffusive dynamics *Phys. Rev. Lett.* **103** 180602

[15] Magdziarz M and Weron A 2011 Anomalous diffusion: testing ergodicity breaking in experimental data *Phys. Rev.* E **84** 051138

[16] Golding I and Cox E C 2006 Physical nature of bacterial cytoplasm *Phys. Rev. Lett.* **96** 098102

[17] Klafter J and Sokolov I M 2011 *First Steps in Random Walks* (Oxford: Oxford University Press) (https://doi.org/10.1093/acprof:oso/9780199234868.001.0001)

[18] Mandelbrot B and Van Ness J 1968 Fractional Brownian motions, fractional noises and applications *SIAM Rev.* **10** 422–37

[19] Deng W and Barkai E 2009 Ergodic properties of fractional Brownian–Langevin motion *Phys. Rev.* E **79** 1–7

[20] Meroz Y and Sokolov I M 2015 A toolbox for determining subdiffusive mechanisms *Phys. Rep.* **573** 1–29

[21] Lanoiselée Y and Grebenkov D S 2016 Revealing nonergodic dynamics in living cells from a single particle trajectory *Phys. Rev.* E **93** 052146

[22] Bouchaud J-P 1992 Weak ergodicity breaking and aging in disordered systems *J. Phys.* I **2** 1705–13

[23] Massignan P, Manzo C, Torreno-Pina J A, García-Parajo M F, Lewenstein M and Lapeyre G J 2014 Nonergodic subdiffusion from Brownian motion in an inhomogeneous medium *Phys. Rev. Lett.* **112** 1–5

[24] Chubynsky M V and Slater G W 2014 Diffusing diffusivity: a model for anomalous, yet Brownian, diffusion *Phys. Rev. Lett.* **113** 098302

[25] Krapf D, Marinari E, Metzler R, Oshanin G, Xu X and Squarcini A 2018 Power spectral density of a single Brownian trajectory: what one can and cannot learn from it *New J. Phys.* **20** 023029

[26] Hinsen K and Kneller G R 2016 Communication: a multiscale Bayesian inference approach to analyzing subdiffusion in particle trajectories *J. Chem. Phys.* **145** 0–5

[27] Thapa S, Lomholt M A, Krog J, Cherstvy A G and Metzler R 2018 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.* **20** 29018–37

[28] Dosset P, Rassam P, Fernandez L, Espenel C, Rubinstein E, Margeat E and Milhiet P-E 2016 Automatic detection of diffusion modes within biological membranes using back-propagation neural network *BMC Bioinf.* **17** 197

[29] Kowalek P, Loch-Olszewska H and Szwabiński J 2019 Classification of diffusion modes in single particle tracking data: feature based versus deep learning approach *Phys. Rev.* E **100** 032410

[30] Breiman L, Friedman J H, Olshen R A and Stone C J 1984 *Classification and Regression Trees* (Belmont, CA: Wadsworth Publishing Company) (https://doi.org/10.1201/9781315139470)

[31] Breiman L 2001 Random forests *Mach. Learn.* **45** 5–32

[32] Fernández-Delgado M, Cernadas E, Barro S and Amorim D 2014 Do we need hundreds of classifiers to solve real world classification problems? *J. Mach. Learn. Res.* **15** 3133–81

[33] Zaburdaev V, Denisov S and Klafter J 2015 Lévy walks *Rev. Mod. Phys.* **87** 483–530

[34] Muñoz Gil G 2019 RF-single-trajectory-characterization https://github.com/gorkamunoz/RF-Single-Trajectory-Characterization https://doi.org/10.5281/zenodo.3533014

[35] Muñoz Gil G, García-March M A, Manzo C, Celi A and Lewenstein M 2019 Diffusion through a network of compartments separated by partially-transmitting boundaries *Frontiers Phys.* **7** 31

[36] Molina-García D, Pham T M, Paradisi P, Manzo C and Pagnini G 2016 Fractional kinetics emerging from ergodicity breaking in random media *Phys. Rev.* E **94** 052147

[37] Balcerek M, Loch-Olszewska H, Torreno-Pina J A, Garcia-Parajo M F, Weron A, Manzo C and Burnecki K 2019 Inhomogeneous membrane receptor diffusion explained by a fractional heteroscedastic time series model *Phys. Chem. Chem. Phys.* **21** 3114–21

[38] Bo S, Schmidt F, Eichhorn R and Volpe G 2019 Measurement of anomalous diffusion using recurrent neural networks *Phys. Rev.* E **100** 010102(R)