PAPER • OPEN ACCESS

Unsupervised learning of anomalous diffusion data: an anomaly detection approach

View the article online for updates and enhancements.

Unsupervised learning of anomalous diffusion data: an anomaly detection approach

Gorka Muñoz-Gil[1,*] ⓘ**, Guillem Guigo i Corominas**[1]
and Maciej Lewenstein[1,2]

[1] ICFO-Institut de Ciències Fotòniques, The Barcelona Institute of Science and Technology, Av. Carl Friedrich Gauss 3, 08860 Castelldefels (Barcelona), Spain
[2] ICREA, Pg. Lluís Companys 23, 08010 Barcelona, Spain

E-mail: munoz.gil.gorka@gmail.com

Abstract
The characterization of diffusion processes is a keystone in our understanding of a variety of physical phenomena. Many of these deviate from Brownian motion, giving rise to anomalous diffusion. Various theoretical models exists nowadays to describe such processes, but their application to experimental setups is often challenging, due to the stochastic nature of the phenomena and the difficulty to harness reliable data. The latter often consists on short and noisy trajectories, which are hard to characterize with usual statistical approaches. In recent years, we have witnessed an impressive effort to bridge theory and experiments by means of supervised machine learning techniques, with astonishing results. In this work, we explore the use of unsupervised methods in anomalous diffusion data. We show that the main diffusion characteristics can be learnt without the need of any labelling of the data. We use such method to discriminate between anomalous diffusion models and extract their physical parameters. Moreover, we explore the feasibility of finding novel types of diffusion, in this case represented by compositions of existing diffusion models. At last, we showcase the use of the method in experimental data and demonstrate its advantages for cases where supervised learning is not applicable.

*Author to whom any correspondence should be addressed.

(Some figures may appear in colour only in the online journal)

Stochastic diffusion processes are ubiquitous in nature, with applications in diverse fields such as physics, chemistry, biology or social sciences. The best-known model for these is Brownian motion [1], but recently, deviations from such paradigmatic phenomena have been discovered in a huge range of systems [2]. We define as anomalous diffusion any of these deviations, which are usually characterized by a power-law scaling of the mean squared displacement (MSD),

$$\langle x^2(t) \rangle \sim t^\alpha, \tag{1}$$

where α is defined as the anomalous exponent. For $\alpha = 1$ one typically recovers normal diffusion, and usually a return to the Brownian motion universality class (for exceptions, see reference [3]). Any $\alpha \neq 1$ showcases anomalous diffusion, while other features such as non-ergodicity or correlated displacements may also be signals of its appearance [4, 5].

In recent years, there has been an increasing interest in the topic, motivated by the development of single particle tracking (SPT) techniques, which open the possibility of studying the motion of particles beyond the diffraction limit [6]. Such achievements have shown that diffusion in many different biological scenarios is indeed anomalous [7]. Moreover, researchers have satisfactorily applied such theoretical framework to systems at all scales, ranging from ultracold atom experiments [5], to animals [8] or economic signals [9].

While currently both theory and experimental techniques are very well developed, connecting these two is a non-trivial task. Due to the stochastic nature of the problem, the analysis of anomalous diffusion data and its comparison to theoretical predictions often relies on the use of ensemble averages [10, 11] or extensive statistical analysis [12, 13]. Moreover, trajectories arising from SPT are usually short and cannot be compared to the long time predictions theories usually make. On top of that, the experimental data is often very noisy and rather difficult to gather, hence increasing even more the difficulty of the problem.

Lately, considerable interest has been devoted to develop and improve single trajectory methods, i.e. tools which extract the maximum amount of information from a single anomalous diffusion trajectory. Here, machine learning (ML) approaches have shown incredible success and are able to beat state of the art methods in a variety of scenarios [14]. A wide range of ML architectures have been tested: from usual neural networks [15], convolutional [16] and recurrent layers [17, 18], graph neural networks [19], Bayesian inference [20], or extreme learning machines [21]. All these methods are trained in a supervised scheme, which means that the machine learns to characterize the data by training with human-labelled data. While for some applications, such as parameter estimation, this approach is completely valid and advantageous, for others it creates an inherent problem that may deeply affect our understanding of the data analyzed.

An example of such inherent problem is the discrimination of the diffusion model of the particles, which has been usually stated as an ML classification problem [22–24]. When classifying between diffusion *modes* (i.e. normal, anomalous, confined or directed diffusion), the problem is simplified, as they probably contain all possible kinds of diffusion. However, a different problem is to classify between diffusion *models*, such as e.g. fractional Brownian motion (FBM) or continuous time random walk (CTRW) (see section 1.3 for further details on these). In this case, it is very possible that the actual model of the data analyzed is not contained in the training set, or may even be a combination of these. Usual supervised approaches

lack the tools to discriminate these cases and they associate them with the most resembling model [24].

In this work, we explore the use of unsupervised techniques for single trajectory characterization and show that they can indeed be used to analyze anomalous diffusion data coming both from simulations and various experimental setups. In particular, we focus on the use of neural network autoencoders (AEs), a typical ML architecture for unsupervised and semi-supervised approaches. Our aim is: first, to investigate if unsupervised learning yields the same results of supervised techniques previously studied; second, to study the use of unsupervised approaches to understand and compare different anomalous diffusion models beyond supervised methods; third, to explore the possibility of finding new physical processes related to anomalous diffusion by means of AEs.

The work is organized as follows: in section 1 we present the unsupervised approach of use, namely convolutional AEs (section 1.1) and anomaly detection (section 1.2). Moreover, in section 1.3 we present the anomalous data used, as well as further details on the anomalous diffusion models considered. Then, the results of the paper are presented in section 2. First, we explore the feasibility of anomaly detection in simulated data (section 2.1), both for model discrimination (section 2.1.1) and parameter estimation (section 2.1.2, but also to characterize novel types of diffusion, showcased in this case by the combination of existing models. (Section 2.1.3) we then apply the knowledge gained to analyze experimental data arising from SPT trajectories (section 2.2.)

1. Methods

1.1. Autoencoders

AEs are artificial neural networks that learn efficient encodings of high-dimensional data. They consist of two parts: an encoder that compresses the input data into a low-dimensional embedding, referred to as the *latent space*, and a decoder that recovers the original input from the compressed representation. Ideally, the output of the decoder is equal to the input of the encoder. However, due to the compression of information happening in the latent space, such an ideal scenario is often not found. Because of such compression, in order to improve the similarity between reconstructions and inputs, the AE has to learn features of the dataset during the training. Such features will then be used to plug the information gaps caused by the encoder's compression.

In our particular problem, the training data consists of a set of simulated anomalous diffusion trajectories $X = \{x^{(1)}, x^{(2)}, \ldots, x^{(n)}\}$, where each training sample x_i consist of a one dimensional single trajectory with T timesteps, $x^{(i)} = [x_1^{(i)}, x_2^{(i)}, \ldots, x_T^{(i)}]$, with $x \in \mathbb{R}$. A typical loss function for the reconstruction of such data is the mean squared error (MSE), which for a single trajectory reads

$$\text{MSE} = \frac{1}{T} \sum_{t=1}^{T} (f(x_t^{(i)}) - x_t^{(i)})^2, \tag{2}$$

where $f(x^{(i)})$ is the AE output for the input trajectory $x^{(i)}$.

The AE used consists of a stack of convolutional and fully-connected layers. Moreover, we consider features such as batch normalization, as well as a global average pooling and upsampling before and after the latent space. An schematic representation of such AE is presented in figure 1(b). In this scheme, connections exist between subsequent layers, as typically occurs in most neural networks architectures, but extra connections are created between not
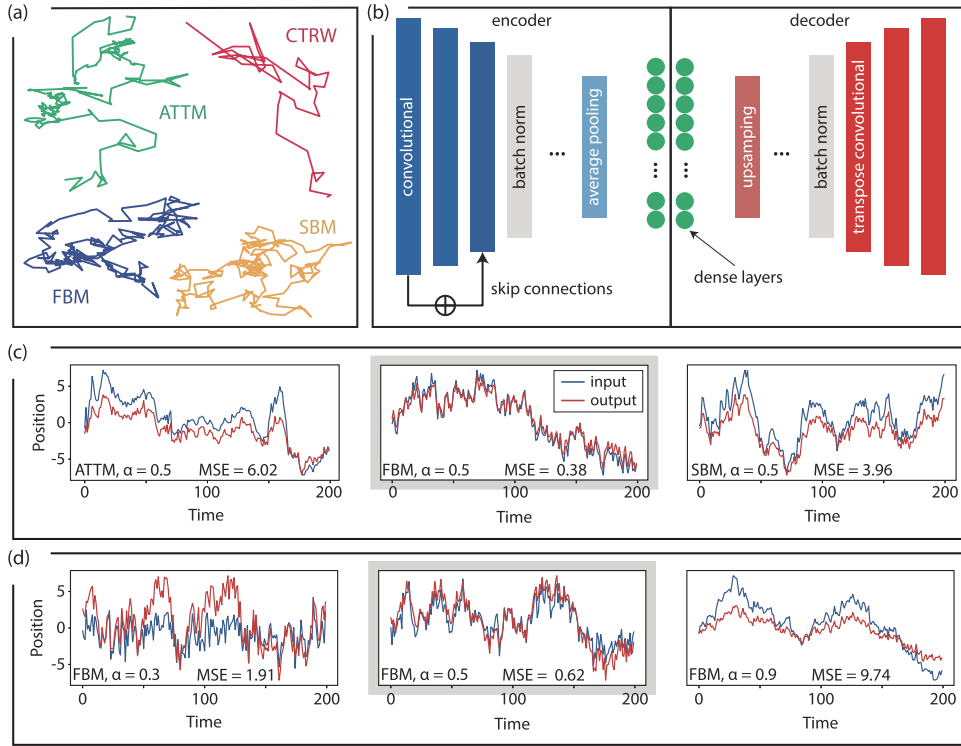
Figure 1. (a) Representative two dimensional trajectories of the diffusion models considered in this work. All trajectories have anomalous exponent $\alpha = 0.7$. (b) Schematic representation of the AEs used, showcasing some of the layers composing them, as well as the *skip connections* feature. (c) Input and output of an AE trained with FBM at $\alpha = 0.5$ (highlighted in gray), predicting trajectories of ATTM, FBM and SBM with $\alpha = 0.5$, respectively. (d) Input and output of the same AE, now with trajectories of FBM at three different exponents, $\alpha = [0.3, 0.5, 0.9]$. In both (c) and (d), the best reconstruction (lowest MSE) occurs for trajectories of the model/exponent with which the AE was trained.

subsequent ones. Such a special feature is often referred to as *skip connections*, and is widely used in residual neural networks [25]. Heuristically, we find that the use of such connections boosts the power of the AE, increasing vastly its accuracy. Moreover, due to the stochastic nature of the trajectories, any decrease in the dimensionality of the trajectory (i.e. to the size of the latent space) will cause the loss of necessary information for its reconstruction. In that sense, the reconstruction error will always have a lower bound related to the size of the latent space [26]. In our particular implementation, the latent space consists of a fully connected layer whose size will depend on the length of the trajectories T. A detailed description of the architecture is presented in appendix A as well as in the repository of reference [27].

Importantly, in order to avoid the effect of overfitting of the AEs, and to ensure that their predictions are related to the learnt features and not the memorization of the data, all the results presented in this manuscript correspond to predictions over sets of trajectories not used in the training.

1.2. Anomaly detection

Anomaly detection refers to the problem of detecting instances or samples from a certain dataset that differ in some sort from the rest. In other words, the goal is to detect outliers, a.k.a. anomalies, which while having similarities with the rest of the dataset, also have special features that crucially make them different. In recent years, ML methods have shown great capabilities in dealing with such problem [28], with interesting application in physics [29].

Particularly, AEs offer a powerful and versatile architecture for anomaly detection while relying on unsupervised or semi-supervised learning [30]. As commented in the previous section, the goal of an AE is to reconstruct an input instance with minimal error. In this sense, anomalies can be detected by setting a threshold to this error, in such a way that any value above indicates that the input reconstructed is indeed an anomaly of the dataset. In this particular work, the prediction error is given by the MSE, equation (2). Heuristically, we find that such a metric, in conjunction with the data preprocessing presented in section 1.3, yields good results and suffices for the task at hand. Nevertheless, other interesting approaches, based on the reconstruction probabilities [31] rather than absolute values are left to be explored in future works.

1.3. Anomalous diffusion data

As commented, the datasets considered consist of trajectories arising from different anomalous diffusion models. In particular, we consider four different models, with which we aim to cover a large variety of real-world phenomena: annealed transient time motion (ATTM) [32], CTRW [33], FBM [34], Lévy walk [35], and scaled Brownian motion (SBM) [36]. Moreover, these models allow us to explore the whole range of anomalous diffusion exponents α: ATTM and CTRW are subdiffusive models ($\alpha \leqslant 1$) and FBM and SBM cover the whole range (i.e. $0 < 1 < 2$). A complete review on the properties of each model can be found in references [2, 14].

The trajectories that we analyzed here were generated via the ANDI dataset python package [14]. Moreover, each trajectory is normalized such that the standard deviation of their displacements is equal to 1. This ensures that all trajectories have a similar diffusion coefficient $D \approx 1$, but the anomalous exponent is unaltered. Moreover, to avoid very slow and immobile trajectories, we will restrict the dataset to $\alpha \geqslant 0.1$. Finally, all the results presented in this work corresponds to AE trained with trajectories of only 20 time steps, if nothing different is stated. While similar results are found for longer trajectories, our goal is to showcase the applicability of the method for experimental scenarios, where such short trajectories are often the only ones accessible.

In figures 1(c) and (d) we briefly showcase the expected results when dealing with anomaly detection with anomalous diffusion trajectories. Depending on the features of the input data, the AE will do a better or worst reconstruction. For instance, in figure 1(c) we showcase that an AE trained with FBM yields better results, seen both visually and with the values of the MSE, when reconstructing trajectories of this same model (highlighted in gray), rather that trajectories of ATTM (left) or SBM (right). Similarly, in figure 1(d), we showcase that the MSE is lower when dealing with trajectories of the same exponent the AE was trained with. In the next sections we will deeply analyze this behaviour and understand the power and limitations of the proposed anomaly detection method.

2. Results

We present here the results on applying anomaly detection to anomalous diffusion trajectories. In particular, we focus here on two important aspects of the trajectories: their diffusion model

and their anomalous exponents. Our aim is to show that an AE can learn to reconstruct trajectories arising from different anomalous diffusion conditions. More importantly, after training, the AE must be capable of reconstructing new, unseen trajectories. As commented previously, the reconstruction error will then be used as a measure of how related the new trajectories are w.r.t. the ones used for training. In order to benchmark the method in a controlled scenario, we will first present the results on simulated trajectories, and then show its applicability to data from various experimental conditions.

2.1. Simulated trajectories

2.1.1. Detecting changes of diffusion model.   Our first result relates to the ability of the method to differentiate between distinct diffusion models. We start with the simplest scenario, in which an AE is trained with a single diffusion model. Then, we compare its error (as given by the MSE, equation (2)) when reconstructing trajectories of other models. In figure 2(a) we show the results when training AEs with subdiffusive trajectories with the models indicated in the legend. Each line corresponds to the averaged error over ten different AE and the shaded area represents their respective standard deviations. Then, we reconstruct trajectories with each of the AE for different models, as shown in the *x*-axis. Importantly, in all cases, the error of the AE is minimal in the model it was trained with (highlighted with a black rounded circle). Moreover, we can already infer, based on the reconstruction error, similarities between some of the models. For instance, the AEs trained with ATTM and SBM trajectories share a very similar behaviour. This is to be expected as they are the most resembling from the pool of chosen models. Indeed, both are so-called random diffusivity models, i.e. showing Brownian-like diffusion but with non-trivial changes of the diffusion coefficient. We note that, even in the case of supervised and model specific algorithms, their discrimination is very challenging [14]. In the case of the FBM trajectories, we see that the error made by AEs is an order of magnitude larger than the rest of the models (note the logarithmic scale). This showcases the strong differences between FBM and the rest of models, being the former the only ergodic model, as well as the only with correlated displacements.

To further understand the presented results, we study how the MSE behaves for the different models at different anomalous exponents α, as shown in figures 2(b)–(e). Each panel corresponds to an AE trained with the indicated model, while each line corresponds to trajectories of the different models (as indicated by the legend in figure 2(a)). These AE were trained in the whole range of α showcased. We see here that when $\alpha \to 1$, trajectories of all models collapse to the same MSE, no matter the AE used. This is an important result, as even if their microscopical behaviour may be different, all models converge to the Brownian motion universality class at $\alpha = 1$ (up to logarithmic corrections [37]). Hence, the AEs are able to learn reliable features from the anomalous diffusion trajectories and not just trivial patterns.

Such results also allow us to further compare the different models considered. Again, we see that the non-ergodic models tend to share very similar MSE. For instance, the AEs trained with ATTM (figure 2(b)) and SBM (figure 2(e) are able to correctly reconstruct all models but FBM. However, for CTRW trajectories the best results are found when trained with the latter model (figure 2(d), as the characteristic trapping times of CTRW may be difficult to reproduce if not learnt in the training. As previously seen, the error made by the FBM trained AE shows errors an order of magnitude larger when reconstructing other models, which nevertheless compare to the error of FBM at $\alpha = 1$.

2.1.2. Detecting parameter changes within a diffusion model.   While the AEs trained with ATTM, CTRW and SBM show an almost constant MSE at all α for their own model, the one trained with FBM tends to minimize the error for a subset of exponents. Taking advantage
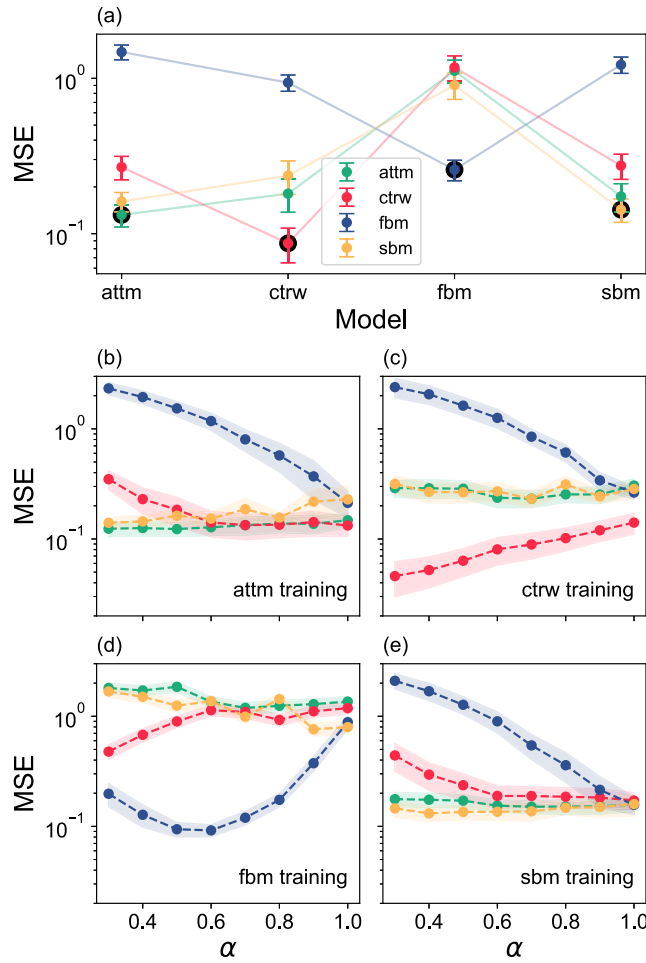
Figure 2. (a) MSE vs the diffusion model of the input trajectories. Each line corresponds to an AE trained with the indicated model. (b)−(e) MSE vs anomalous exponent α. Each panel corresponds to an AE trained with the indicated diffusion model, while the color of the lines indicate the model of the input trajectories, following the legend of (a).

that both FBM and SBM are able to generate both subdiffusive and superdiffusive trajectories, we trained AEs in $\alpha \in [0, 2]$. In figures 3(a) and (b) we show such results for FBM and SBM, respectively. The line color indicates the model of the input trajectories. Interestingly, the FBM trained AE shows the same MSE behaviour stated above, but now the minimal MSE occurs at a different exponent. Surprisingly, in the subdiffusive range, the MSE is lower for SBM than for FBM (and also in the highest part of the superdiffusive range). This does not occur in the SBM trained AE, where the MSE is almost constant for all α and is always minimal for the model it was trained with. Again, in this case, the minimal error for FBM trajectories is found at $\alpha = 1$.

The FBM model offers an interesting playground to study the suitability of anomaly detection in anomalous diffusion. Within the same diffusion model, FBM shows two completely different behaviours, depending on the value of α. For $0 < \alpha < 1$, FBM is anti-persistent, hence showing negative correlations, while it is persistent and positively correlated for
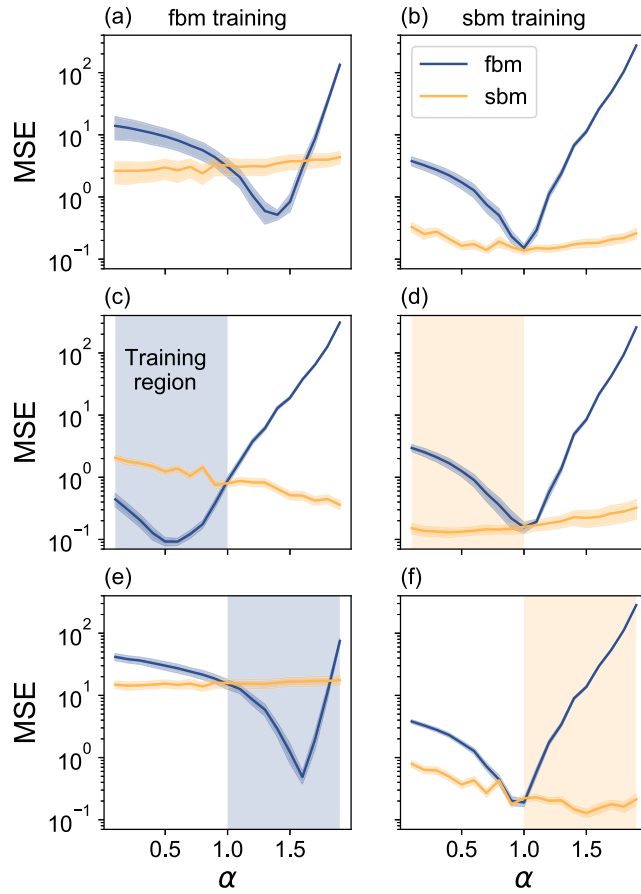
Figure 3. MSE vs anomalous exponent α for AEs trained with FBM (left) or SBM (right) trajectories in different ranges of α (as indicated by the shaded regions): (a) and (b) $\alpha \in [0.1, 2)$; (c) and (d) $\alpha \in [0.1, 1]$; (e) and (f) $\alpha \in [1, 2)$.

$1 < \alpha < 2$. This makes trajectories for different α very different from each other, hence the difficulty of an AE to correctly reconstruct trajectories at all α. Such behaviour is highlighted in figures 3(b)–(e), where AEs were trained with FBM and SBM in the subdiffusive (b) and (c) and superdiffusive (d) and (e) ranges. While the information learnt by the SBM trained AE in any of these two ranges is sufficient to correctly reconstruct trajectories for all α, the one trained with FBM only yields low errors for the range it was trained on.

In order to understand further this behaviour, we analyze the reconstructed trajectories output by the AE trained with two models, FBM and CTRW. We note that the results found for CTRW are analog to the ones found with SBM and ATTM. We proceed to input trajectories to the AE with different α. Then, we calculate the anomalous exponent of the output trajectories by either fitting an ensemble average of the MSD (for CTRW) or by performing a time averaged MSD (for the FBM). A linear fit in the logarithmic scale outputs the anomalous exponent $\alpha_f \approx \alpha$.

We present results for AEs trained at three different α, as shown in figure 4. For this test, we consider trajectories only of the model the AE was trained with (i.e. FBM and CTRW in
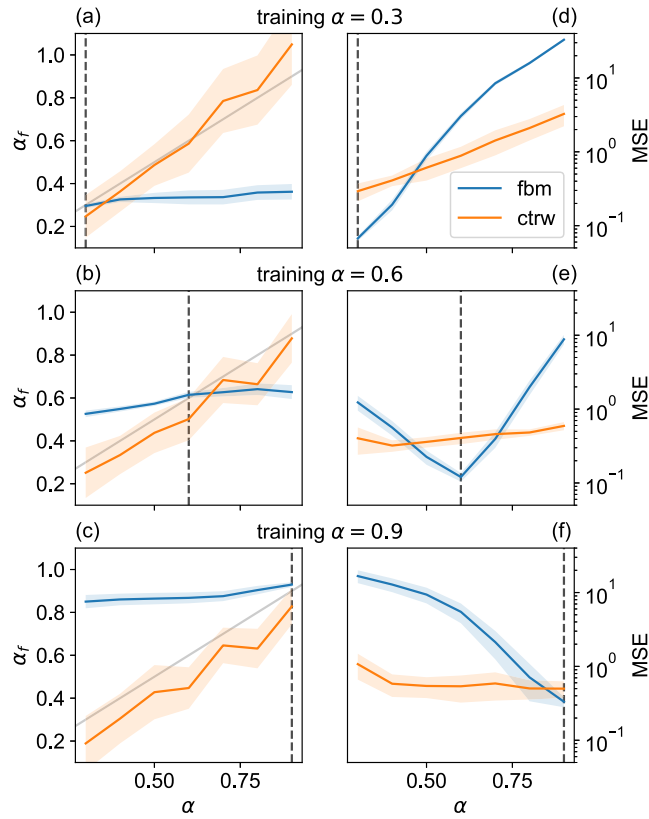
Figure 4. (a)–(c) Fitted anomalous exponent α_f of the reconstructed trajectories vs the ground-truth exponent α of the input trajectory. Gray line corresponds to $\alpha_f = \alpha$. (d)–(f) MSE vs anomalous exponent α. For all plots, we consider AEs trained with FBM (blue) or ATTM (orange). Each row corresponds to AEs trained at different α, indicated in the title and the vertical dashed line.

this case). Moreover, to allow for more accurate estimation of α_f via the ensemble and time averaged MSD, the AEs trained with trajectories of $T = 200$ time steps. In the left column, we show the fitted α_f w.r.t. to the exponent of the input trajectories. In the right column we show the reconstruction error for these same AEs, again as a function of the exponent of the input trajectories. In all cases, for the CTRW trained AE (orange), no matter the exponent it was trained with, the output trajectories will maintain the exponent of the input (orange line in figures 4(a)–(c)). Surprisingly, even if the MSE increases for αs different of the one used in the training (see e.g. figure 4(d)), the AE is able to reproduce a key feature of the CTRW trajectories as is the anomalous exponent. Oppositely, the outputs of the AE trained with FBM are transformed such that their α_f equals the one used for training (blue line in figures 4(a)–(c)). This hence causes that the MSE grows significantly when dealing with trajectories of different α, as we saw already in figure 3 and again highlighted in the right column of figure 4.

We note here that while both CTRW and FBM trajectories can be characterized by the anomalous exponent, the source of anomalous diffusion complete differs from one another. In the case of the CTRW, the exponent is related to the distribution of waiting or trapping times of the particle. However, in the case of FBM, the exponent is connected to the displacement's

correlations. The previous results showcase that the later phenomena is a crucial feature for the FBM trained AE, as once a certain correlation is learnt (i.e. one α is learnt), all the reconstructed trajectories share that same correlation. However, in the case of CTRW, the AE is able to reconstruct trajectories without 'loosing' their characteristic exponent, even with different α of the one used from training. The reason of this difference is still unknown and further exploration of such phenomena is currently ongoing with more sophisticated AE.

2.1.3. Detecting mixed diffusion models. To further understand the extent of use of unsupervised learning in anomalous diffusion data, we test the method in trajectories arising from the composition of different diffusion models. Indeed, it is expected that the diffusion of particles in complex environments, such as biological systems, is not described by a single model but the intertwine between two or more. For instance, in references [38, 39] the diffusion of particles through the cell membrane was described by a particle performing an FBM in the presence of a fractal structure and traps, respectively. The latter phenomena can be associated to a CTRW, hence leading to the intertwined diffusion of CTRW and FBM. In the small temporal scale, the dominant model would be FBM. However, in a bigger picture, traces of CTRW such as a broad distribution of trapping times will arise. Recently, other kind of diffusion model compositions have been studied, in which particles switch between different diffusion states [40, 41]. Nevertheless, we believe that such problem may be better solved with via segmentation algorithms [14, 17] and focus on the former subordinated phenomena.

In order to simplify the present example, we consider trajectories whose displacements $\Delta^i_{\gamma,\beta}$ are the weighted sum of displacements of three different diffusion models:

$$\Delta^i_{\gamma,\beta} = \gamma\beta\Delta^i_{\text{model}_1} + (1 - \gamma)\Delta^i_{\text{model}_2} + (1 - \beta)\Delta^i_{\text{model}_3}, \tag{3}$$

where $\Delta^i_{\text{model}_j}$ refers to the *i*th displacement of a trajectory of the model *j*. We note that the displacements used to generate a composite trajectory come from a single trajectory of each model, hence maintaining their properties (e.g. correlations and ageing effects). Then, the parameters γ and β allow us to change the dominant model, creating the phase space schematically represented in figure 5(a).

We start our analysis by performing anomaly detection of trajectories arising from equation (3) with AE trained with *pure* models, as shown in figures 5(b)–(d). As expected, the loss is minimized in the corner associated with the diffusion model used for training. Interestingly, the AE trained with CTRW favours trajectories of pure models, while the MSE is maximal in the regions where the models are equally weighted. In the case of the ATTM trained AE, the MSE increases when approaching the SBM corner and in the case of pure CTRW. Such a result reproduces what is expected from a theoretical point of view: the lower right sector of the phase diagram can be understood as a noisy CTRW model, which is indeed very difficult to differentiate from an ATTM trajectory in the trajectory length of work ($T = 20$). Last, the SBM trained AE seems to correctly reconstruct a much bigger region of the phase space, while favouring ATTM over CTRW. We note that the results found in the corners of the phase space are completely analogous to the ones in figure 2. As expected, most of the conclusions drawn here are also similar to the one discussed in section 2.1.1. However, this approach also allows for a much more refined analysis of the trajectories, allowing for a deeper level of understanding and the possibility of finding novel types of diffusion, as will be discussed in section 2.2.

The previous results allow us to validate the use of the composite models and their analysis with AE trained with single models. A different approach is to consider AEs trained with trajectories of different γ and β. Then, performing anomaly detection with such AE will allow us to characterize the model of the input trajectories by finding the γ and β with minimal MSE.
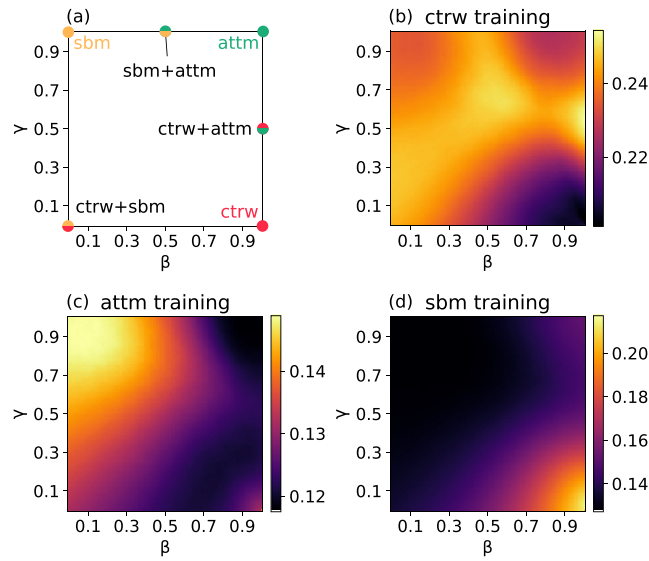
Figure 5. (a) Schematic representation of the diffusion models governing a trajectory with certain γ and β as given by equation (3). (c) and (d) MSE as a function of the trajectories γ and β for AE trained with CTRW, ATTM, and SBM trajectories in the subdiffusive range $\alpha \in [0.1, 1]$ and $T = 20$. Each point averages the results of 10 AE over 5000 trajectories.

Such approach has no particular interest applied to the considered simulated trajectories, and will hence be directly applied to experimental datasets with unknown models.

2.2. Experimental trajectories

The previous results showcased the validity of the presented method to detect variations and outliers in simulated data. However, as commented previously, the main utility of these techniques relies on their implementation in experimental data, as they provide a fast and accurate method to analyze anomalous diffusion trajectories. To showcase that unsupervised methods can be used for that aim, we present here results on the use of anomaly detection in trajectories from various SPT experiments. While the focus here is in biophysical problems, as stated in the introduction, these techniques are applicable at any scale or system where stochastic trajectories arise.

Here we consider trajectories from three different experiments and present their results in each row of figure 6. The trajectories of the two first experiments were cropped to have 200 time steps while for the third trajectories have only 20 time steps. Hence, different AEs will be used for each length. First, we analyze the trajectories of telomeres in the nucleus of mammalian cells [12, 42]. In these two references, the authors showed that the trajectories are well described by an FBM model, with a crossover in the anomalous exponent from $\alpha \sim 0.5$ at short times to $\alpha \sim 1$ for longer times. The second dataset corresponds to the trajectories of receptors diffusing in the plasma membrane of mammalian cells [43]. These trajectories have been shown to be consistent with the ATTM and $\alpha \sim 0.8$.

Before presenting the third dataset, we focus on the results found for the latter datasets. We start by inputting the trajectories to AEs trained with different diffusion models. In the case of the telomeres (figure 6(a)), the MSE is largely reduced when reconstructing trajectories with
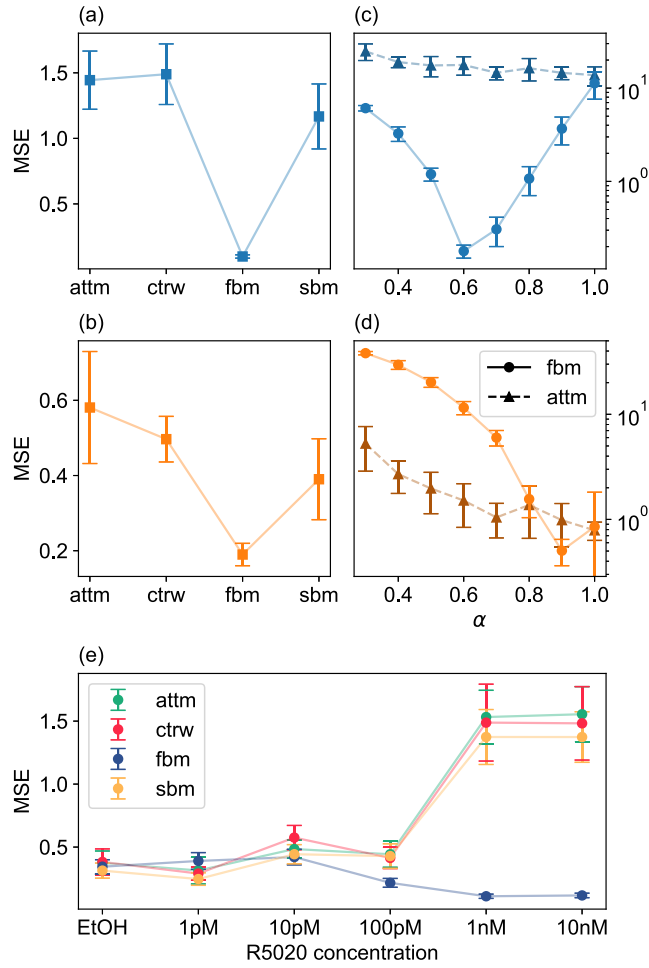
Figure 6. Anomaly detection in three experimental datasets. Each row corresponds to references [12, 42], reference [43] and reference [44], respectively. (a) and (b) MSE vs the diffusion model used to train the AE. (c) and (d) MSE. vs the anomalous exponent α for AEs trained with FBM (circles, solid line) and ATTM (triangles, dashed line). (e) MSE for four AEs trained with the indicated diffusion model at different R5020 hormone concentration.

an FBM trained AE, hence coinciding with the expected results. For the diffusion of receptors (figure 6(b)), the MSE is minimal for ATTM, but also very close to the SBM one. While these results are obviously not conclusive, they already hint at important information and are indeed very well related to what we showed for simulated trajectories above (i.e. shared behaviour between SBM and ATTM). Moreover, this is also consistent with previous ML supervised approaches (see figure 5(c) of reference [14]).

When performing anomaly detection with different anomalous exponents, the results on each dataset differ, as the assigned diffusion models are also different. For both datasets, we show in figures 6(c) and (d) the MSE w.r.t. to the anomalous exponent α of the training dataset, for AEs trained with FBM (solid line, circles) and ATTM (dashed line, triangles). For the diffusion of telomeres, there is a sharp decrease of the MSE for the FBM trained AE (figure 6(c),

circles), indicating an expected exponent of $\alpha = 0.6$, consistent with previous predictions. For the receptors' trajectories, this same AE has a minimum at $\alpha = 1$, an expected result as our previous analysis shows that these trajectories do not diffuse as FBM. Hence, the MSE is expected to be minimized at $\alpha = 1$. This result is analogous to what was found in section 2.1.1 and figures 2(b)–(e). Inputting the receptors' trajectories to an ATTM trained AE, the minimisation of the AE occurs in $\alpha \in [0.7, 1]$. While these results are in accordance with previous works, we note that AE trained with ATTM at different α are able to correctly reconstruct trajectories from the rest of the range, as shown in figures 3(d)–(f). We note again that, while the exploration of diffusion models by means of AE gives rise to great opportunities, we believe that the extraction of diffusion parameters, such as in the case the anomalous exponent, will always be better with supervised techniques.

At last, we focus on the motion of the progesterone receptor as it diffuses in the cell nucleus, under the presence of the R5020 hormone [44]. This system shows liquid–liquid phase separation above a certain hormone concentration. Interestingly, two populations coexist, one where particles diffuse freely through the cell membrane, following ATTM and $\alpha \sim 0.8$ and a second one containing the particles interacting with the chromatin and forming phase separated aggregates, following FBM and $\alpha \sim 0.4$. As the hormone concentration increases, more and more particles interact with the chromatin and hence the FBM population grows.

Importantly, such result is also found by means of anomaly detection. In figure 6(e) we show the MSE for four AEs, trained with the diffusion model indicated in the legend. While the MSE is similar for all AEs at low hormone concentrations, we see a sharp increase above 10 nM. This is consistent with the results shown in reference [44] and corresponds to the point in which most of the particles in the system phase-separate. This makes the FBM population increase drastically. Hence, the AE trained with such a model decreases its error, while the rest see a big increase. At low concentrations, the ATTM and FBM populations are balanced, hence the errors for ATTM and FBM are of the same order. Interestingly, the AEs trained with CTRW and SBM show a similar behaviour as the ATTM one, following what we already saw in simulated trajectories (see figure 2). We note here an important remark: the trajectories of this last experiment are extremely short (only 20 time steps). This makes any proper statistical analysis extremely challenging and even impossible. While the presented results may not be conclusive in differentiating the non-ergodic models, it already hints very important information that was inaccessible to non ML approaches. Moreover, being this an unsupervised method, the AE's predictions are based solely on the features learnt and not in the need of minimizing a classification based metric, as may happen in supervised learning.

To further investigate the diffusion of the PR, we use a similar method as the one presented in section 2.1.3. In this case, we will train AE with trajectories at different γ and β. For this particular example, we consider CTRW, ATTM and FBM as models 1–3, respectively (see equation (3)). This yields the phase space schematically represented by the points and labels of the first panel of figure 7(a). First, we calculate the MSE over the set of trajectories with AE trained at different γ and β. At low hormone concentrations, the minimal MSE is found with AE trained with mixed models, all with the presence of FBM. Such result is consistent to what was found in figure 6(e), where all models shared similar MSE. Interestingly, not any mix of models yields low MSE, favouring mixes with FBM. Then, at higher hormone concentrations, the lowest MSE is found with AE trained with trajectories with higher FBM weights, again, similar to what was shown in figure 6(e).

The use of such composite model's maps allows for a fast and visual inspection of anomalous diffusion data. Nevertheless, the plots presented in figure 7(a) shows the average of the
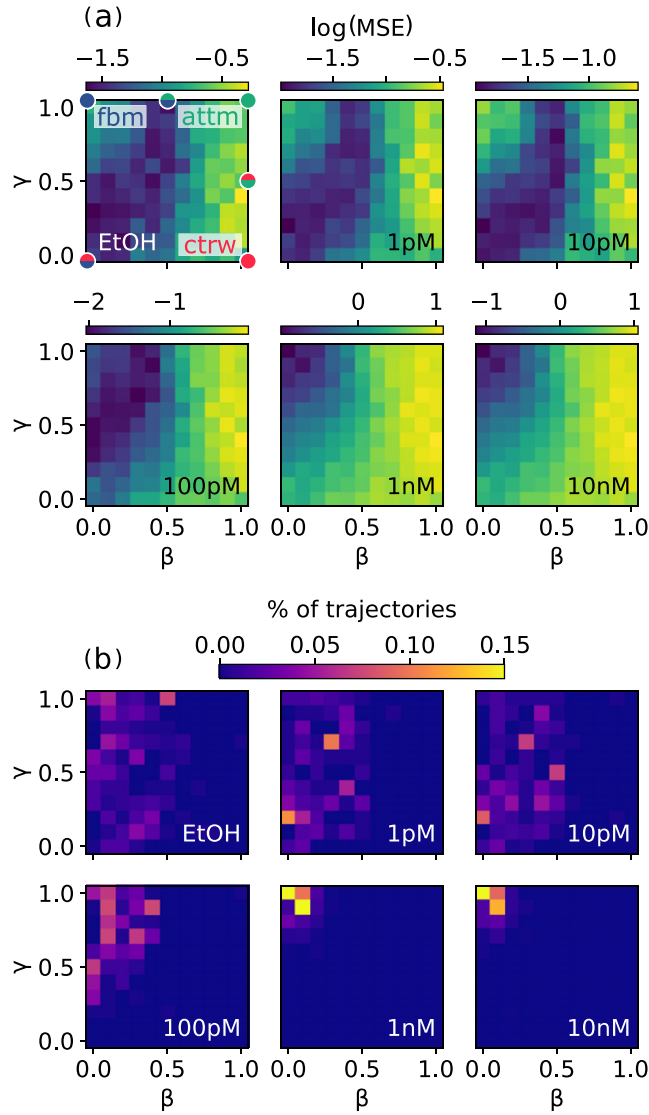
Figure 7. (a) MSE of the experimental dataset of reference [44] for AE trained with composite models (equation (3)) made of CTRW, ATTM and FBM at various γ and β. (b) Percentage of trajectories whose minimal MSE is found for the AE trained at the given γ and β. Each point corresponds to the average over 7 different AEs.

MSE over the trajectories of the dataset. However, one of the goals of this work is to analyze the diffusion at the single trajectory level. To do so with the previous method, we look at the prediction for each trajectory with the AEs trained with composite models. Then, we find the AE that yields the lowest MSE. In figure 7(b) we show the percentage of trajectories whose minimal MSE is found at some γ and β. As expected, such approach is analogous to the one of figure 7(a) but allows for a single trajectory based analysis. As before, we see that below the critical hormone concentration (1 nM), there is big heterogeneity in the data. As explained in reference [44], in this cases two diffusion modes coexist, one related to the anomalous diffusion

of PR through the cell membrane and the other related to phase separated PR interacting with the chromatin. Only above the 1 nM hormone concentration, the diffusion is dominated by the chromatin interaction, hence most of the trajectories have minimal MSE with AE trained with pure FBM trajectories (i.e. $\gamma = 1$ and $\beta = 0$).

3. Conclusions

We have shown that AEs can learn to reconstruct trajectories arising from various anomalous diffusion models. By using the reconstruction error as a metric, we presented a method based on anomaly detection able to characterize the diffusion at the single trajectory level. Particularly, we focus on: (1) the discrimination of the theoretical diffusion model that better describes the trajectory's features; (2) the scaling factor of the MSD, defined as the anomalous exponent, usually used to evidentiate the arising of anomalous diffusion. More importantly, the training is completely unsupervised. In previous ML approaches, single trajectory characterization was based on either a classification problem (for diffusion models) or a regression problem (anomalous exponent). This makes the discovery of novel types of diffusion very challenging or even impossible, as the machine is restricted either to a pool of models or a range of parameters set by the training data.

Anomaly detection avoids such restrictions, as the training is solely based on the correct reconstruction of the given trajectories. Hence, any relevant information learnt by the AE is directly related to the features of the trajectories, rather than a preassigned label. This method then allows to discover novel types of diffusion, as the machine is not biased at assigning a predefined label to the trajectories. Contrarily, these new types will be heralded by an increased reconstruction error. More importantly, the method is successful at characterizing experimental trajectories, even when these are extremely short or noisy.

This work paves the way for the study of anomalous diffusion with unsupervised ML. While an extended theoretical framework for anomalous diffusion exists nowadays, its connection with the phenomena arising in real physical systems is still challenging. Indeed, it is not clear to what extent the diffusion in complex scenarios, as e.g. crowded environments or the presence of non-trivial interactions, can be fully described by an idealist theoretical model. In that sense, combinations of existing theories, as the ones considered in this work, or data-driven models, may indeed give a better description of the actual physical phenomena. As shown, anomaly detection allows us to test such approaches w.r.t. to experimental data in a reliable and interpretable form.

Acknowledgments

Data availability statement

The data that support the findings of this study are openly available at the following URL/DOI: https://github.com/gorkamunoz/Unsupervised-learning-of-anomalous-diffusion.

Appendix A. Autoencoder architecture

In this appendix we present some of the technical details related to the AEs used. The complete architectures as well as some examples of use can be in found in reference [27]. In this work, we used two different AE architectures: a small AE for short trajectories (20 points) and a larger one for long trajectories (200 points). In both architectures the encoder is made of 1D convolutional layers with ReLU as activation function and the decoder is made of 1D transpose convolutions without activation function. Due to the scale of the data (i.e. not a restricted range and being both positive or negative), we found that not considering any activations in the decoder yield better results.

The encoder in the small AE is made of three building blocks; each building block consists of two convolutional layers with increasing number of channels, followed by a pooling operator. In the first two blocks, the pooling operation is max pooling with stride 2, which reduces the dimensions by half. In the last block, the pooling operation is global average pooling over all the channels. Batch normalization is performed after the convolutions in each building block before pooling. The last two layers of the encoder are linear layers. The last layer has 4 output neurons, meaning that the size of the latent space is 5 times smaller than the input.

The architecture of the decoder is almost identical to that of the encoder, but performs essentially the inverse operation. It has the same number of layers (2 linear layers followed by 3 building blocks). Each block has two transpose convolutional layers followed by batch normalization and max unpooling.

The large AE for long trajectories (200 points) has a similar structure as the small one, the main difference being that each building block has three convolutional layers instead of two. Additionally, the output of the first layer in each block is added to the output of the third layer, following the *skip connection* proposed in the ResNet architecture [25]. Such feature improves dramatically the learning of deep convolutional networks with many layers and avoids the vanishing gradient problem. The extra layers and parameters allow for a better approximation of the larger trajectories. In this AE the latent space has 20 neurons, 10 times smaller than the input (table 1).

The number of epochs for the training varied depending of the dataset characteristics, and was set such as to find the convergence of the MSE loss function. In general, around 200 epochs were needed. The optimization of the parameters was done via the adaptive moment estimation (Adam) method [45].

Table 1. Training dataset characteristics for every AE used in this work. The second and third column describe the diffusion models and exponents considered for the training of the AE. The number of trajectories per exponent and model is always balanced. The fourth column presents the total number of trajectories considered. In all cases, we chose a learning rate of 10^{-3} and an L2 regularization parameter of 10^{-5}.

Figures	Diffusion models	Exponents	# of trajectories	Batch size
2(b), 5(c)	ATTM	0.1, 0.2, …, 1	80.000	256
2(c), 5(b)	CTRW	0.1, 0.2, …, 1	80.000	256
2(d), 3(c)	FBM	0.1, 0.2, …, 1	80.000	256
2(e), 3(d), 5(d)	SBM	0.1, 0.2, …, 1	80.000	256
3(a)	FBM	0.1, 0.2, …, 1.9	160.000	512
3(b)	SBM	0.1, 0.2, …, 1.9	160.000	512
3(e)	FBM	1, 1.1, …, 1.9	80.000	512
3(f)	SBM	1, 1.1, …, 1.9	80.000	512
4(a) and (d)	CTRW, FBM	0.3	8000	64
4(b) and (e)	CTRW, FBM	0.6	8000	64
4(c) and (f)	CTRW, FBM	0.9	8000	64
4(c) and (f)	CTRW, FBM	0.9	8000	64
7	Mix (see section 2.1.3)	0.8	10 000	64

ORCID iDs

Gorka Muñoz-Gil https://orcid.org/0000-0001-9223-0660

References

[1] Metzler R 2019 Brownian motion and beyond: first-passage, power spectrum, non-Gaussianity, and anomalous diffusion *J. Stat. Mech.* 114003

[2] 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

[3] Wang B, Anthony S M, Bae S C and Granick S 2009 Anomalous yet Brownian *Proc. Natl Acad. Sci. USA* **106** 15160–4

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

[5] Kindermann F, Dechant A, Hohmann M, Lausch T, Mayer D, Schmidt F, Lutz E and Widera A 2017 Nonergodic diffusion of single atoms in a periodic potential *Nat. Phys.* **13** 137–41

[6] 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

[7] Höfling F and Franosch T 2013 Anomalous transport in the crowded world of biological cells *Rep. Prog. Phys.* **76** 046602

[8] Paiva L R, Marins A, Cristaldo P F, Ribeiro D M, Alves S G, Reynolds A M, DeSouza O and Miramontes O 2021 Scale-free movement patterns in termites emerge from social interactions and preferential attachments *Proc. Natl Acad. Sci. USA* **118** e2004369118

[9] Plerou V, Gopikrishnan P, Nunes Amaral L A, Gabaix X and Eugene Stanley H 2000 Economic fluctuations and anomalous diffusion *Phys. Rev.* E **62** R3023

[10] 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

[11] Kepten E, Weron A, Sikora G, Burnecki K and Garini Y 2015 Guidelines for the fitting of anomalous diffusion mean square displacement graphs from single particle tracking experiments *PLoS One* **10** e0117722

[12] Krapf D *et al* 2019 Spectral content of a single non-Brownian trajectory *Phys. Rev.* X **9** 011019

[13] Thapa S, Wylomanska A, Sikora G, Wagner C, Krapf D, Kantz H, Chechkin A and Metzler R 2020 Leveraging large-deviation statistics to decipher the stochastic properties of measured trajectories *New J. Phys.* **23** 013008

[14] Muñoz-Gil G *et al* 2021 Objective comparison of methods to decode anomalous diffusion (arXiv:2105.06766)

[15] Gentili A and Volpe G 2021 Characterization of anomalous diffusion classical statistics powered by deep learning (CONDOR) *J. Phys. A: Math. Theor.* **54** 314003

[16] Granik N, Weiss L E, Nehme E, Levin M, Chein M, Perlson E, Roichman Y and Shechtman Y 2019 Single-particle diffusion characterization by deep learning *Biophys. J.* **117** 185–92

[17] Argun A, Volpe G and Bo S 2021 Classification, inference and segmentation of anomalous diffusion with recurrent neural networks *J. Phys. A: Math. Theor.* **54** 294003

[18] Li D, Yao Q and Huang Z 2021 WaveNet-based deep neural networks for the characterization of anomalous diffusion (WADNet) (arXiv:2106.08887)

[19] Verdier H, Duval M, Laurent F, Cassé A, Vestergaard C L and Masson J-B 2021 Learning physical properties of anomalous random walks using graph neural networks *J. Phys. A: Math. Theor.* **54** 234001

[20] Park S, Thapa S, Kim Y, Lomholt M A and Jeon J-H 2021 Bayesian inference of Lévy walks via hidden Markov models (arXiv:2107.05390)

[21] Manzo C 2021 Extreme learning machine for the characterization of anomalous diffusion from single trajectories (arXiv:2105.02597)

[22] 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 Bioinform.* **17** 197

[23] 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

[24] Muñoz-Gil G, Garcia-March M A, Manzo C, Martín-Guerrero J D and Lewenstein M 2020 Single trajectory characterization via machine learning *New J. Phys.* **22** 013010

[25] He K, Zhang X, Ren S and Sun J 2016 Deep residual learning for image recognition *Proc. IEEE Conf. Computer Vision and Pattern Recognition* pp 770–8

[26] Epstein B and Meir R 2019 Generalization bounds for unsupervised and semi-supervised learning with autoencoders (arXiv:1902.01449)

[27] Muñoz-Gil G and Guigo G 2021 Unsupervised Learning of Anomalous Diffusion data: Code Repository https://github.com/gorkamunoz/unsupervised-learning-of-anomalous-diffusion

[28] Chalapathy R and Chawla S 2019 Deep learning for anomaly detection: a survey (arXiv:1901.03407)

[29] Kottmann K, Huembeli P, Lewenstein M and Acín A 2020 Unsupervised phase discovery with deep anomaly detection *Phys. Rev. Lett.* **125** 170603

[30] Zhou C and Paffenroth R C 2017 Anomaly detection with robust deep autoencoders *Proc. 23rd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining* pp 665–74

[31] An J and Cho S 2015 Variational autoencoder based anomaly detection using reconstruction probability *Special Lecture on IE* vol 2 pp 1–18

[32] Massignan P, Manzo C, Torreno-Pina J A, García-Parajo M F, Lewenstein M and Lapeyre G J Jr 2014 Nonergodic subdiffusion from Brownian motion in an inhomogeneous medium *Phys. Rev. Lett.* **112** 150603

[33] Scher H and Montroll E W 1975 Anomalous transit-time dispersion in amorphous solids *Phys. Rev.* B **12** 2455

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

[35] Klafter J and Zumofen G 1994 Lévy statistics in a Hamiltonian system *Phys. Rev.* E **49** 4873

[36] Lim S C and Muniandy S V 2002 Self-similar Gaussian processes for modeling anomalous diffusion *Phys. Rev.* E **66** 021114

[37] Klafter J and Sokolov I M 2011 *First Steps in Random Walks: From Tools to Applications* (Oxford: Oxford University Press)

[38] 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

[39] Ali Tabei S M, 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

[40] Janczura J, Balcerek M, Burnecki K, Sabri A, Weiss M and Krapf D 2021 Identifying heterogeneous diffusion states in the cytoplasm by a hidden Markov model *New J. Phys.* **23** 053018

[41] Sabri A, Xu X, Krapf D and Weiss M 2020 Elucidating the origin of heterogeneous anomalous diffusion in the cytoplasm of mammalian cells *Phys. Rev. Lett.* **125** 058101

[42] Stadler L and Weiss M 2017 Non-equilibrium forces drive the anomalous diffusion of telomeres in the nucleus of mammalian cells *New J. Phys.* **19** 113048

[43] Manzo C, Torreno-Pina J A, Massignan P, Lapeyre G J Jr, 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

[44] Muñoz-Gil G, Romero Aristizabal C, Mateos N, de Llobet Cucalon L I, Beato M, Lewenstein M, Garcia-Parajo M F and Torreno-Pina J A 2020 Phase separation of tunable biomolecular condensates predicted by an interacting particle model https://doi.org/10.1101/2020. 09.09.289876

[45] Kingma D P and Ba J 2014 Adam: a method for stochastic optimization (arXiv:1412.6980)