

Machine-Learning Solutions for the Analysis of Single-Particle Diffusion Trajectories

Henrik Seckler, Janusz Szwabiński, and Ralf Metzler*



Cite This: *J. Phys. Chem. Lett.* 2023, 14, 7910–7923



Read Online

ACCESS |



Metrics & More

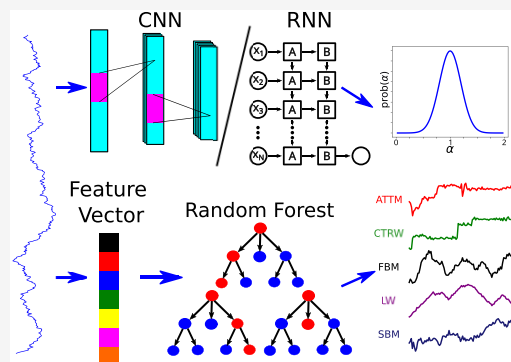


Article Recommendations



Supporting Information

ABSTRACT: Single-particle traces of the diffusive motion of molecules, cells, or animals are by now routinely measured, similar to stochastic records of stock prices or weather data. Deciphering the stochastic mechanism behind the recorded dynamics is vital in understanding the observed systems. Typically, the task is to decipher the exact type of diffusion and/or to determine the system parameters. The tools used in this endeavor are currently being revolutionized by modern machine-learning techniques. In this Perspective we provide an overview of recently introduced methods in machine-learning for diffusive time series, most notably, those successfully competing in the anomalous diffusion challenge. As such methods are often criticized for their lack of interpretability, we focus on means to include uncertainty estimates and feature-based approaches, both improving interpretability and providing concrete insight into the learning process of the machine. We expand the discussion by examining predictions on different out-of-distribution data. We also comment on expected future developments.



Single-particle tracking (SPT) refers to the observation of the microscopic motion of molecules. In 1828, Robert Brown used SPT to observe the movement of granular particles, laying the foundations of Brownian motion.¹ After advancements in theory spearheaded by Einstein, Smoluchowski, Sutherland, and Langevin, Jean Perrin was able to give a first estimate of Avogadro's number by observing particle motion in a colloid.² While SPT applies mainly to observing the movement of molecules or micron-sized tracer particles,^{3–15} similar SPT data are also garnered in systems ranging from the movement of animals^{16–18} to eye movement^{19,20} or stock dynamics.^{21,22} Understanding such trajectories and developing techniques for their analysis is thereby of vital importance in a multitude of different fields.^{9,16,17,21,23–26} Mathematically such a motion is described by a random walk, as introduced by Karl Pearson.²⁷ Here the position x_i of a particle at time t_i is obtained via a sequence of random steps Δx_i ($i = 1, \dots, T - 1$), such that $x_n = x_0 + \sum_{i=1}^n \Delta x_i$ ($n = 0, \dots, T - 1$). The simplest case, called the “Wiener process”, whose steps Δx_i are independent and identically distributed according to $(2\pi\sigma^2)^{-1/2} \exp(-\Delta x_i^2/[2\sigma^2])$ with constant waiting time $t_i - t_{i-1} = \Delta t$, will lead to a Gaussian probability density function (PDF),

$$f(x, t) = \frac{1}{\sqrt{4\pi K_1 t}} \exp\left(-\frac{x^2}{4K_1 t}\right) \quad (1)$$

where $K_1 = \sigma^2/\Delta t$. Due to the action of the Central Limit Theorem (CLT), the same PDF is reached as long as the increments are independent and identically distributed with

finite variance and finite mean waiting time.^{28,29} In particular, this entails a linear growth of the mean-squared displacement (MSD),^{30–32}

$$\langle x^2(t) \rangle \sim 2K_1 t \quad (2)$$

This type of behavior is referred to as normal diffusion, the best known example being the aforementioned Brownian motion as described by Einstein, Smoluchowski, Sutherland, and Langevin when analyzing the motion of small particles suspended in liquids or gases.^{33–36}

In practice, however, one often observes a non-Gaussian PDF and/or an MSD that grows nonlinearly in time.^{5–8,37–48} Here we focus on the frequent case of power-law growth of the MSD,

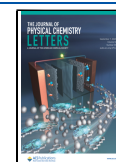
$$\langle x^2(t) \rangle \sim 2K_\alpha t^\alpha \quad (3)$$

referred to as “anomalous diffusion”, with the anomalous diffusion exponent α . A growth slower than linear ($0 < \alpha < 1$) is called subdiffusive, whereas a faster than linear growth ($\alpha > 1$) is referred to as superdiffusive, with the special case of ballistic motion for $\alpha = 2$. For such behavior to emerge, one or more of the conditions for the CLT to kick in need to be

Received: May 17, 2023

Accepted: August 25, 2023

Published: August 30, 2023



violated, as is the case when the system shows heterogeneities, long time correlations, diverging mean waiting times, and/or infinite jump variance. As an example, one may consider a granular gas with a temperature changing over time, which causes non-identically distributed increments since the increment variance is temperature dependent.^{49,50} As a random walk, such a motion is modeled by scaled Brownian motion (SBM), in which the diffusivity is time dependent.^{51,52} A diffusivity increasing with time will lead to superdiffusion, while a decreasing diffusivity will lead to subdiffusion. As another prominent example, long time correlations are often observed in biomolecules, whose crowded environments lead to strong anticorrelations (viscoelastic effects), while active motion may give rise to strong correlations. Mathematically such motion is often modeled as so-called fractional Brownian motion (FBM).⁵³ There exist plenty of other models to explain the occurrence of anomalous diffusion,^{54–59} apart from the mentioned SBM and FBM. We here also consider continuous-time random walk (CTRW), with random waiting times between successive jumps,^{29,60,61} Lévy walks (LW),^{62–66} and annealed transient time motion (ATTM).⁶⁷ We provide short descriptions of each of these models in the [Supporting Information](#).

Since each of these models describes different physical causes for anomalous diffusion, identifying the best-fitting stochastic model is an important step in unraveling the physical origin of an experimentally observed anomalous diffusion.^{4,39–43,68} Similarly determining specific parameters attributed to each model, such as the anomalous diffusion exponent α and coefficient K_α , can help quantify and/or differentiate between trajectories or systems.^{39,69} Typically this task is tackled through the use of statistical observables, aiming at quantifying the expected differences between the models.^{47,70–78} However, the stochastic nature of these models in combination with the often noisy and limited experimental data can severely hinder this process and may lead to conflicting results from different observables. For example, it has been shown that noisy data can lead to a mistaken identification as subdiffusion.^{74,79}

The rising computing power of modern processors has brought along a competing approach. Machine learning (ML) has already shown wide applicability in physical chemistry⁸⁰ and is increasingly used in a variety of fields from materials science⁸¹ to medicine⁸² or quantum chemistry.⁸³ In particular, in recent years ML has also been applied to anomalous diffusion dynamics seen in SPT data.^{84–88} Here the task of finding the best way to determine the underlying diffusion model and model parameters is left to machines trained on simulated trajectories, either by directly feeding into the machine the raw position data or by extracting relevant features from the trajectories first.

After shortly discussing classical methods, we here focus on the competing approaches utilizing ML, most notably those introduced during the so-called “Anomalous Diffusion (AnDi) Challenge”.^{87,89} To address the “black box problem”, we present a deeper look into approaches including uncertainty estimates as well as those relying on extracted features. We present tests for the limits of both approaches when applied to out-of-distribution data. We conclude with a discussion on benefits, shortcomings, and expected future developments of ML techniques to analyze anomalous diffusion data.

Classical Approach. The simplest path to the anomalous exponent is given by direct calculation of the scaling exponent

of the MSD, which, given an ensemble of N trajectories, is defined as

$$\langle x^2(t) \rangle = \frac{1}{N} \sum_{n=1}^N (x^{(n)}(t) - x^{(n)}(0))^2 \sim 2K_\alpha t^\alpha \quad (4)$$

In experiments, one often relies on time-series analysis, utilizing the time-averaged MSD (TAMSD),

$$\langle \overline{\delta^2(\Delta)} \rangle = \frac{1}{\mathcal{T} - \Delta} \int_0^{\mathcal{T}-\Delta} \langle [x(t + \Delta) - x(t)]^2 \rangle dt \quad (5)$$

with observation time $\mathcal{T}$. As long as the system is ergodic, the TAMSD for sufficiently long $\mathcal{T}$ will convey the same information as the MSD. However, for anomalous diffusion this is often not the case—for instance, when models feature diverging mean waiting times, such as CTRW or LW. This indicates that, when experimental conditions allow access to both ensemble MSD and TAMSD, the possible difference between their behaviors allows one to differentiate between ergodic and non-ergodic models.

An alternative method is provided by the p -variation test.^{71,90,91} The sample p -variation is calculated using the difference between every m th element of the trajectory,

$$V_m^{(p)} = \sum_{k=0}^{(T-1)/m-1} |x_{(k+1)m} - x_{km}|^p \quad (6)$$

Different models often show different behaviors of the p -variation. For example, for FBM we have

$$V_m^{(p)} \propto m^{p\alpha/2-1} \quad (7)$$

implying that, as a function of m , the p -variation increases for $p > 2/\alpha$ and decreases for $p < 2/\alpha$. This is in contrast to, e.g., CTRW, where the p -variation will decrease for $p > 2$ and increase for $p < 2$, regardless of anomalous exponent α . Thus, calculating the p -variation for different p values can help differentiate between models or, for some models, provide an estimate of the anomalous exponent α . However, static noise may compromise the p -variation output, as, e.g., tested for subdiffusive CTRWs.⁹² Alternatively, it is also possible to decompose the anomalous dynamics into the Moses M , Noah N , and Joseph J scaling exponents (with $\alpha/2 = J + L + M - 1$), obtained from the scaling of the cumulative absolute increments, the sum of the squared increments, and the rescaled range statistic. Each of these exponents corresponds to the violation of one of the three conditions for the CLT.^{78,93–95}

Another method is given through the use of the single-trajectory power spectral density (PSD),^{72–74}

$$S(f, \mathcal{T}) = \frac{1}{\mathcal{T}} \left| \int_0^{\mathcal{T}} dt e^{ift} x(t) \right|^2 \quad (8)$$

Of particular interest here is the coefficient of variation,

$$\gamma(f, \mathcal{T}) = \frac{\sigma(f, \mathcal{T})}{\mu(f, \mathcal{T})} \quad (9)$$

where $\sigma(f, \mathcal{T})$ and $\mu(f, \mathcal{T})$ are the mean value and variance of the PSD. In FBM, for example, $\gamma(f, \mathcal{T})$ shows distinct behavior for subdiffusion ($\gamma(f, \mathcal{T}) \approx 1$), superdiffusion ($\gamma(f, \mathcal{T}) \approx \sqrt{2}$), and normal diffusion ($\gamma(f, \mathcal{T}) \approx \sqrt{5}/2$), in the limit of high

frequencies or long observation times.⁴¹ Single-trajectory PSDs are also quite robust against static and dynamic noise.⁷⁴

The aforementioned methods cover only a fraction of possibilities. Other techniques not further specified here include the use of the velocity autocorrelation,⁴⁷ the first passage statistics,⁷⁵ the codifference,⁷⁶ or the autocovariance.⁷⁷ The applications of these statistical techniques, however, struggle when data is sparse and often require an ensemble of trajectories.^{74,79}

As an alternative approach to classify SPT data, Thapa et al. demonstrated that Bayesian inference may be used to determine the best-fitting model and its parameters directly from the position data of an SPT experiment using their mathematical description.⁹⁶ Specifically, these descriptions allow one to directly calculate the likelihood of a given trajectory for a specific model with the given parameters. These parameters are then adjusted to maximize the probability of the trajectory. The difference in the maximum likelihood is used to determine the most probable model. This method has shown great promise for processes for which the likelihood is easily calculated in closed form, such as FBM or SBM. It struggles, however, when models feature hidden waiting times, though there have been recent advances using hidden Markov processes.⁹⁷ Even so, high computational cost remains an issue for Bayesian inference, often resulting in a trade-off between computational feasibility and accuracy.

The Anomalous Diffusion Challenge. ML in recent years has grown into a strongly competing class of approaches. In 2019, Granik et al., using a convolutional neural network, demonstrated that one can differentiate between simulated Brownian motion, CTRW, and sub- or superdiffusive FBM trajectories.⁸⁴ In the same year, Bo et al. used a similar procedure to determine the anomalous diffusion exponent of FBM trajectories via a recurrent neural network.⁸⁵ Similarly, in 2020, Muñoz-Gil et al. demonstrated that a random tree forest can differentiate between CTRW, LW, FBM and ATTM and provide an estimate for the anomalous diffusion exponent.⁸⁶ In all these cases, it was shown that ML can achieve better accuracy than conventional methods, especially when the available data is sparse. It should be noted, however, that these approaches all suffer from the often-quoted “black box problem”, outputting answers without explanations as to how these are obtained,⁹⁸ as detailed below.

Among ML approaches, the mentioned strategies utilizing convolutional neural networks,^{84,99} recurrent neural networks,^{85,88} and random tree forests⁸⁶ already differ significantly. In an effort to compare the performance of different techniques, in 2020 Muñoz-Gil et al. launched the AnDi-Challenge.^{87,89} Reported in 2021, the goal of the AnDi-Challenge was to provide a competitive comparison of different available methods to decode anomalous diffusion.^{87,100} The AnDi-Challenge also continues to serve as a benchmark to quickly assess the performance of newly developed or improved methods.^{94,95,97,101–111}

The challenge consisted of three tasks: (i) inference of the anomalous diffusion exponent, (ii) classification of the diffusion models, and (iii) segmentation of trajectories. For tasks (i) and (ii), participants were given a set of trajectories, each randomly generated from one of five different anomalous diffusion models with a randomly chosen anomalous diffusion exponent. For task (iii), the model and/or exponent changed at a given point in the trajectory. Participants were required to predict the change point, in addition to the model and the

anomalous diffusion exponent in both segments. To emulate experimental data, all trajectories were corrupted by white Gaussian noise of varying strength. Moderately sized training data sets as well as the code necessary to generate further labeled data are freely available in a repository.¹¹²

In total, 15 teams participated in the AnDi-Challenge, using a variety of different methods. While most teams used some form of ML, the more traditional approaches were represented by teams using Bayesian inference^{97,111,113} and scaling analysis as well as feature engineering, primarily based on a decomposition method using the Moses, Noah, and Joseph exponents.^{94,95}

Several different ML techniques were used, some of which were applied to the raw position data.^{85,101–105} Other methods relied on features extracted from the input trajectories^{91,99,106–109,114} or used a combination of both strategies.^{115,116} The techniques using raw data focused on deep learning (DL),^{84,85,101–105} while the feature-based methods also included other ML methods such as gradient boosting,^{99,108,114} random forests,^{91,99,108,114} and extreme learning machines.¹⁰⁷ In the AnDi-Challenge, the ML methods outperformed the classical approaches, with top results obtained by DL, achieving an accuracy of 88% for model classification and a mean absolute error (MAE) of 0.14 for the regression of the anomalous diffusion exponent for 2D trajectories.¹⁰⁰ For comparison, the more traditional Bayesian inference—with a limited amount of processes for which the likelihood function was derived at that point—achieved an accuracy of 53% and MAE of 0.20 in the challenge.¹⁰⁰ Classical observables such as the above-mentioned decomposition method using the scaling exponents M , N , and J scored 51% accuracy with an MAE of 0.31.¹⁰⁰

The Raw Data Approach of Deep Learning. Following the increasing availability of high computational power along with increasingly more-detailed data sets, more and more ML approaches rely on highly complex architectures involving thousands of parameters. With DL we refer to neural networks with many hidden layers, often resulting in several hundreds of thousand of fitting parameters (weights).¹¹⁷ The complexity of these models allows them to directly learn from massive amounts of raw data with little to no need for human-engineered preprocessing. Specifically for the analysis of anomalous diffusion, this entails directly learning from the position-time series of the recorded trajectories. To speed up training and reduce the required data volume, the input data undergo minimal preprocessing via a normalization of their standard deviation. Since diffusion models only rely on the increments of a process and no additional relevant information is included on the absolute positions, the trajectories are often also converted to the increment process.^{85,101–105,112} In inhomogeneous, static environments, this condition may, of course, no longer hold.

The DL solutions, as presented in the AnDi-Challenge and newly developed ones since, mostly utilize convolutional^{84,102–105} and recurrent^{85,101,102,118} neural network architectures. In convolutional neural networks, best known for their applications in image classification, the layers consist of one or several convolutional kernels that are slid along the input tensor.¹¹⁹ Stacking multiple such layers makes it possible to detect correlations in the sequence. Recurrent neural networks, most notably the so-called “long short-term memory” (LSTM) networks, are specifically designed for time sequence data, making them useful for tasks such as speech recognition,

translation, or sequence forecasts.¹²⁰ Layers typically consist of a single recurrent unit applied successively to each time step, with outputs of the previous time step included as additional inputs for the next time step. Figure 1 shows simplified

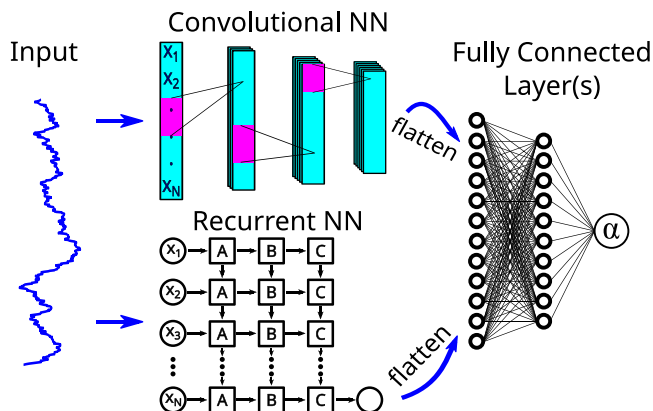


Figure 1. Schematic representation of convolutional and recurrent neural network architectures for the analysis of single-particle trajectories. In both cases, input data consist of normalized trajectory positions (or increments) $x_1, \dots, x_N$. In a convolutional neural network, a kernel is slid along the input data, generating outputs for each region. Usually each layer consists of multiple kernels with identical sizes but different weights, each generating a new data sequence, depicted in the figure as an additional dimension. We show here a convolutional neural network with three layers utilizing 3, 5, and 6 kernels, respectively. In a recurrent neural network, as depicted in the lower half, the data are passed in sequence through a recurrent unit, with the output of the previous time step included as input in the next step (vertical connections in the figure). Here we depict a stacked recurrent neural network consisting of three layers with weight matrices A , B , and C , respectively. For both recurrent and convolutional networks, the resulting output is usually flattened into a one-dimensional array and passed through one or multiple fully connected layers, ending, e.g., in a prediction of the anomalous diffusion exponent α or diffusion coefficient K_α .

schematic representations of both architectures. Other notable architectures that have been shown to be applicable to SPT data analysis include graph neural networks^{115,116} and transformer/encoder networks.^{103,104}

In the AnDi-Challenge, all top results were achieved by DL,^{84,100–103,106} though notably one of these did rely on extracted features rather than the raw trajectories.¹⁰⁶ Overall, DL, and ML in general, showed great promise in the community challenge. However, one should not dismiss the shortcomings of such methods, which are most often criticized for their lack of interpretability.⁹⁸ To that end, we discuss added uncertainty estimations, as well as feature-based approaches, in the following.

Qualifying Deep Learning by Including Uncertainties. Classical DL models only provide point estimates of the output and do not furnish any concrete information on the reliability of this estimate. In extreme cases, this also means that these methods will provide outputs on data that have nothing to do with the learned problem. Even within the desired task, different inputs may provide the same point estimate but underlie massively different uncertainties. As an example, Figure 2 depicts two diffusion trajectories which, fed into a neural network, would both be assigned an anomalous diffusion exponent of $\alpha = 1$, even though one input contains considerably more information (data points) than the other.

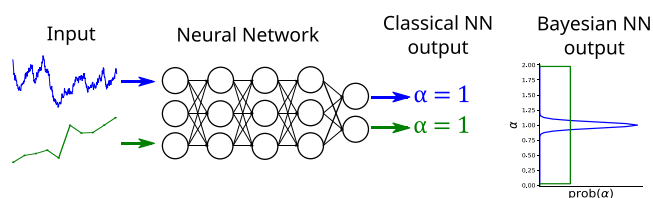


Figure 2. Uncertainty problem of DL. The same anomalous diffusion exponent is predicted for two trajectories of different lengths when fed into a classical neural network—despite their differing amounts of information. The two cases can be distinguished only when the probability distribution of possible output anomalous diffusion exponents is considered, instead of a point estimate. Such an estimate can be provided by Bayesian neural networks. Figure adapted from ref 10. Copyright 2022, The Authors.

To reveal the difference, one would need to output a probability distribution of the α values instead. Examining such a distribution reveals that, while the first trajectory is roughly a Brownian motion, the prediction of the second is just the result of obtaining no relevant information, with the uniform distribution of the anomalous diffusion exponent over the interval $\alpha \in [0, 2]$ leading to a point estimate of $\alpha = 1$.

To change the predictions to a probability distribution, we need to model two types of uncertainty.^{121,122} *Aleatoric* uncertainty refers to the uncertainty inherent in the data caused, for instance, by measurement noise or an inherent stochasticity of the system. This uncertainty remains even for a perfect model obtained from an infinite amount of data and therefore must be included in the output of the neural network model and trained by utilizing an appropriate loss function.^{123,124} As no model is perfect, it is insufficient to consider the aleatoric uncertainty alone. Namely, to account for the difference between training and test data, or an insufficient amount of training data in the first place, one needs to introduce a second uncertainty measure. *Epistemic* (or *systematic*) uncertainty can be included by considering the weights of the neural network themselves as uncertain quantities. Formally, the probability $p(\theta|\mathcal{D})$ of the weights θ , given data $\mathcal{D}$, is given by Bayes's rule,¹²⁵

$$p(\theta|\mathcal{D}) = \frac{p(\mathcal{D}|\theta)p(\theta)}{p(\mathcal{D})} \quad (10)$$

To obtain the final probability $p(y|x_i, \mathcal{D})$ for some output y given the input x_i , we combine the aleatoric uncertainty, represented by the probability $p(y|x_i, \theta)$, for one set of weights θ , with the epistemic uncertainty by marginalization over the weights. The resulting integral is usually approximated through Monte Carlo sampling,^{126,127}

$$p(y|x_i, \mathcal{D}) = \int d\theta p(y|x_i, \theta)p(\theta|\mathcal{D}) \approx \frac{1}{M} \sum_{m=1}^M p(y|x_i, \theta_m) \quad (11)$$

where θ_m is sampled from the distribution $p(\theta|\mathcal{D})$ for a sufficient number M of discrete points. As an exact calculation of $p(\theta|\mathcal{D})$ (via eq 10) quickly becomes computationally infeasible for deep neural networks, one uses approximations to generate the samples θ_m . Various methods, summarized under the term *Bayesian Deep Learning*, have been proposed, the simplest of which is to train an ensemble of neural networks, known as *deep ensembles*.¹²⁸ Other ways to generate samples include *MC-Dropout*,^{129,130} in which one uses dropout

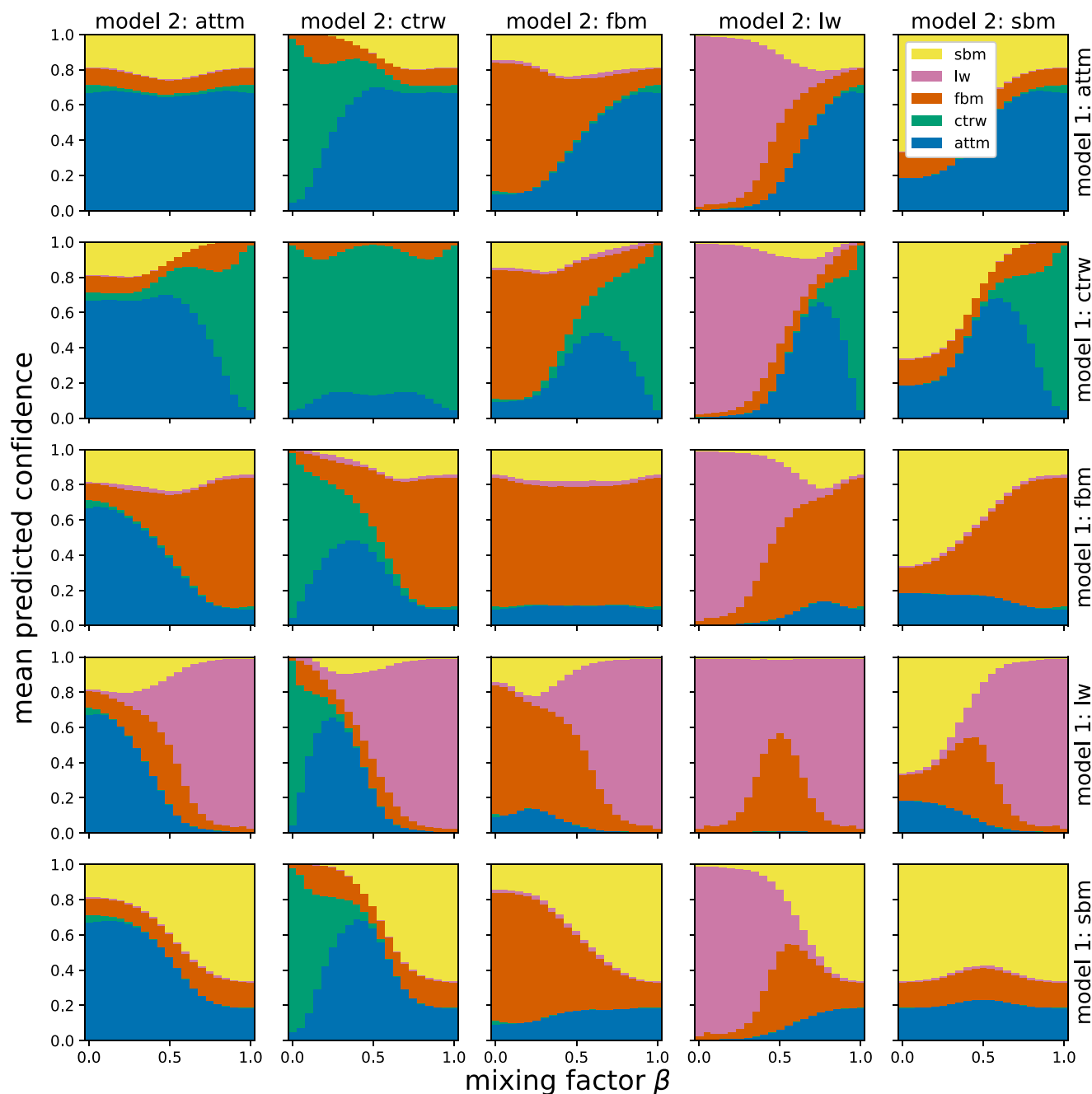


Figure 3. ML classification for a superposition of two models. The panels depict the dependence of the mean confidence assigned by the neural network when presented with a mixture of two models on the mixing factor β . The depicted results are obtained from two-dimensional trajectories with 100 data points each.

to generate multiple samples from the same neural network, and *Stochastic Weight Averaging Gaussian* (SWAG),^{131,132} which approximates $p(\theta|\mathcal{D})$ by a Gaussian distribution, obtained by interpreting a stochastic gradient descent¹³³ as an approximate *Bayesian Inference* scheme.

Recently it was demonstrated that, based on *Multi-SWAG*, a combination of SWAG and deep ensembles, one can add informative uncertainty predictions to the DL solution for the analysis of single-particle anomalous diffusion trajectories.¹¹⁰ The introduced method maintains the performance of the top AnDi-Challenge competitors, while it provides a well-calibrated uncertainty estimate with expected calibration errors^{134,135} of only 0.0034 for the regression of α and

0.45% for the classification of the diffusion model. On top of this, it was demonstrated¹¹⁰ that the added error prediction improves the interpretability of the deep neural networks, demonstrating in detail that the predicted behavior can be linked to properties of the underlying diffusion models. Figure S1 shows an example of how error predictions can be analyzed when inferring the anomalous diffusion exponent.

To further elaborate on the study in ref 110, we now discuss the results obtained from the *Multi-SWAG* approach when confronting the introduced networks with previously unseen out-of-distribution data. First, we examine the outputs when feeding the network with a superposition of two models, the increments of which are obtained by the weighted sum of the

increments of two models with random anomalous diffusion exponents. With the mixing factor β , we then obtain

$$\Delta x_{\text{new}} = \beta \Delta x_{\text{model1}} + (1 - \beta) \Delta x_{\text{model2}} \quad (12)$$

Based on two-dimensional trajectories of length 100, Figure 3 shows the dependence of the mean confidences, that is, the mean value of the predicted model probabilities over 2×10^5 input trajectories, on the mixing factor β for different model combinations, represented by the rows and columns in the panel grid. For the convenience of the reader, the panels include the redundant case of swapped models 1 and 2, which results in a symmetry with respect to the panel grid diagonal, i.e., superposition of a model with itself. In most cases, we see a smooth transition of the confidence from the marginal cases on the left and right, which are the normal predictions for pure trajectories of models 2 and 1, respectively. A notable exception, however, is the behavior for superpositions with CTRW, as these often show high probabilities for ATTM. Since ATTM could be considered a combination of CTRW and Brownian motion, often showing the jumping motion of CTRW interspersed with Brownian motion, this is not unexpected. Moreover, we see that superpositions with LW often show high probabilities for FBM, which can be explained due to the similarity of LW with highly correlated FBM. Analogous 1D behavior can be seen in Figure S2.

As another example we confront the trained neural network with trajectories obtained from the mobile-immobile model (MIM).^{136–138} In the MIM, trajectories switch between mobile and immobile states, with mean residence times τ_m and τ_{im} . At equilibrium, the fraction of time a test particle spends in the mobile phase is given by $f_m = \tau_m / (\tau_m + \tau_{im})$. This model provides information about the immobilized fraction of the particle motion. Moreover, it includes a continuous transition between a normal-diffusive ($\alpha = 1$) CTRW on the one side for a low fraction of the mobility ($f_m \rightarrow 0$) and Brownian motion on the other side for a high fraction of the mobility ($f_m \rightarrow 1$). The results, depicted in Figure 4, confirm that the method correctly classifies the two extremes as CTRW for low mobility and as Brownian motion for high mobility,

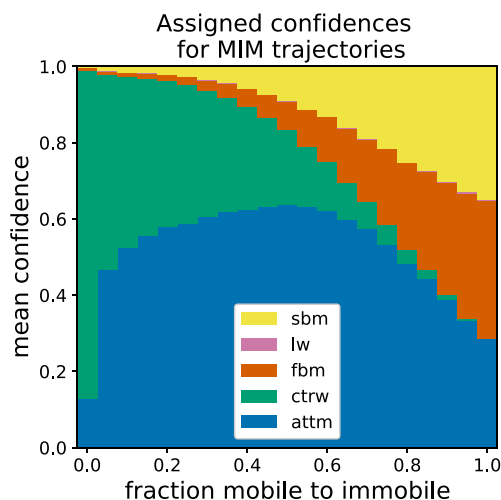


Figure 4. Classification for mobile-immobile model (MIM) trajectories for different fractions of mobility. The MIM effectively converges to CTRW for low mobile fractions f_m and to Brownian motion for $f_m \approx 1$. The depicted results are obtained from one-dimensional trajectories with 250 data points each.

which for this method is represented as a split probability among SBM, FBM, and ATTM (all three models that can exhibit Brownian motion). In between these two limits we see high confidences for ATTM, which is not surprising, as ATTM is the only model, of those considered here, that mimics the phase switching behavior of a MIM trajectory.

Feature-Based Classification of Single-Particle Trajectories. As demonstrated by the AnDi-Challenge,¹⁰⁰ DL methods perform excellently in the analysis of the diffusion models and outperform the more traditional approaches to SPT data. However, the choice of a suitable classification method is usually more subtle than simply looking at its performance. The availability of tools and libraries for DL makes it relatively easy to quickly create effective predictive models. But due to their complexity, those models are black boxes providing (almost) no insight into the decision-making processes. In the previous section, we showed how confidences can be established to judge the validity of the provided output. Here we consider the interpretability of the parameters in the ML approach. To give an example for the complexity in DL, consider ResNet18, one of the simplest deep residual network architectures used in ref 88 for trajectory classification. This network originally had 11,220,420 parameters. The authors were able to reduce this number to 399,556, with a positive impact on the accuracy of the resulting classifier. Although this is an impressive achievement, the interpretation of all those remaining parameters is, of course, practically elusive.

The trade-off between a model's accuracy and its interpretability is one of the reasons for feature-based attempts for the classification of diffusion models.^{86,91,99,108,114,139} These feature-based methods are statistical learning algorithms that do not operate on raw data. Instead, each data sample is characterized by a vector of human-engineered features or attributes. Those vectors are then used as inputs for a classifier (see Figure 5 for a workflow of the method). In some sense, those methods may be treated as a kind of extension to the statistical techniques usually used for classification purposes. Instead of conducting a testing procedure based on one statistic, we can turn all of them into features and use them to train the model. This could be of particular importance in situations when single statistics yield inconclusive results or when testing results based on different statistics significantly differ from each other.¹⁴⁰ Automated feature-based analysis can thus be used in addition to DL methods to learn more about the values of specific features and their relative importance in categorizing input data.

Feature engineering, i.e., the process of extracting attributes from raw data, is not a trivial task. It requires domain expertise to pinpoint which features may be valuable for the process that generated the given set of data. It may also be time and resource consuming, as testing the impact of newly created features on the predictions involves repetitive trial-and-error work. It has been already shown that classifiers, which were trained with a popular set of features, may not generalize well beyond the situations encountered in the training set.⁹⁹ Thus, careful attention must be paid to the choice of the attributes. They should cover all important characteristics of the process, but, at the same time, they should contain the minimal amount of unnecessary information, as each redundant piece of data causes noise in the classification and may lead to overfitting (see ref 141 for a general discussion concerning the choice of features).

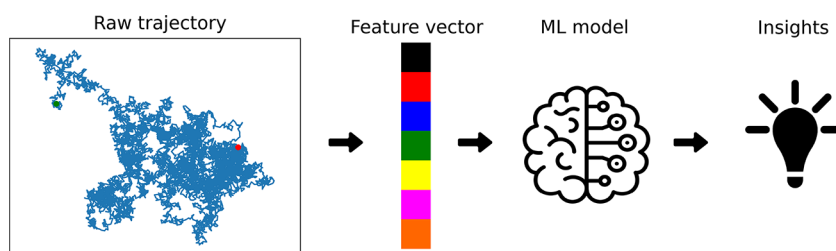


Figure 5. Schematic workflow of the feature-based method: a set of features is extracted from raw trajectories and used as input to the classification or regression model. Analysis of the impact of the features on the outcome gives insights into the decision-making process of the model.

Once the appropriate set of features is identified, the choice of an actual classification algorithm is of secondary importance. Very often, random forest^{86,91,99,114,139} or gradient boosting^{86,91,99,108,114} methods are used, because they offer a reasonable compromise between the accuracy of the results and their interpretability. Both algorithms fall into the category of ensemble learning, i.e., methods that generate many classifiers and aggregate their results. In both cases, decision trees¹⁴² are used as the basic classifier. In a random forest, several trees are constructed from the same training data. For a given input, the predictions of individual trees are collected, and then their mode is taken as the output. In the case of gradient boosting, the trees are not independent. Instead, the single classifiers are built sequentially from the mistakes committed by the ensemble (see Figure 6). In terms of interpretability, both algorithms are placed between single decision trees (which are easy to interpret) and DL (with the black box problem).

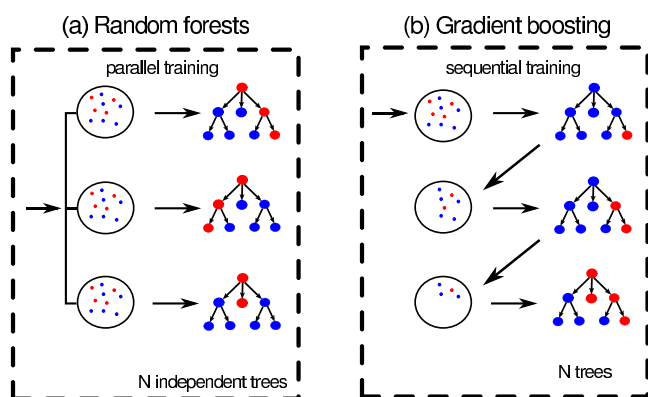


Figure 6. Schematic comparison between random forest (left) and gradient boosting methods (right). In the random forest, N independent trees are built in parallel from random subsets of the input data set. In gradient boosting, the next tree is constructed from the residuals of the ensemble and added to it.

In the AnDi-Challenge, the feature-based contribution was outperformed by the winning teams using DL (73% accuracy versus 88% for the winners). However, the authors of the feature-based method further elaborated on their set of features to achieve 83% accuracy on the same validation set.¹⁰⁸ This was based on a mixture of characteristics tailor-made to the diffusion processes (e.g., MSD, anomalous diffusion exponent, diffusion coefficient) and problem-agnostic ones (e.g., detrending moving average, kurtosis). All features used in refs 87 and 108 are summarized in Table 1.

The authors were able to assess the importance of the features in the overall classification and to calculate the

Table 1. Features Used to Characterize Single-Particle Trajectories^a

Original Features
Anomalous exponent
Diffusion coefficient
Asymmetry
Efficiency
Empirical velocity autocorrelation function
Fractal dimension
Maximal excursion
Mean maximal excursion
Mean Gaussianity
Mean-squared displacement ratio
Kurtosis
Statistics based on p -variation
Straightness
Trappedness
Additional Features
D'Agostino–Pearson test statistic
Kolmogorov–Smirnov statistic against χ^2 distribution
Noah exponent
Moses exponent
Joseph exponent
Detrending moving average
Average moving window characteristics
Maximum standard deviation

^aThe original set of features was used in the AnDi-Challenge and achieved 73% accuracy. With the additional features, the performance of the classifier increased to 83%. The definitions of the features may be found in Appendix B and in ref 108.

contribution of each attribute to the classification of every single trajectory, giving some insight into the decision-making process of the classifier.¹⁰⁸ The results achieved with a simple gradient boosting method indicate that the feature-based ML, overshadowed somewhat by DL approaches in recent years, constitutes a serious alternative to the state-of-the-art approaches. It should also be mentioned that better interpretability is not the only benefit related to feature-based methods. Compared to DL, they usually work better on small data sets and are computationally (and thus also financially) cheaper; see ref 99 for a short comparison. Additionally, in the case of SPT data, they naturally allow for the simultaneous analysis of trajectories of different lengths.

Testing the Limitations of Machine Learning. During the AnDi-Challenge, the competitors were provided with large training data sets and tested on data generated from the same distributions as those used for the training. This practice gives an undeniable advantage to ML in general and to DL

especially. Additionally, the artificial data used in the AnDi-Challenge considered only white Gaussian noise, which may not be sufficient to account for all of the noise sources present in experimental data. To address these problems, we here test ML models to analyze their performance when confronted with (a) data corrupted with dynamic noise and (b) the task of determining the anomalous exponent for models not included in the training data. These tests should indicate (a) how robust the methods are to different noise types and (b) how well the learned determination of α can be generalized to other models.

Dynamic noise stems from the finite exposure time needed to generate each data point. In contrast to the additive Gaussian white noise, this error is characterized by temporal integration, which for discrete time steps is replaced by a sum,

$$\bar{x}(t) = \frac{1}{\tau_e} \int_0^{\tau_e} x(t - \xi) d\xi \rightarrow \frac{1}{n_e} \sum_{j=0}^{n_e-1} x(t - j\Delta t) \quad (13)$$

where $\tau_e = n_e \Delta t$ is the exposure time consisting of n_e time steps of length Δt .^{74,143} To test the ML models, we generate data sets in the same manner as in the AnDi-Challenge but with added dynamic noise of different exposure times containing 10,000 trajectories each. In Table 2 we see the results when

Table 2. Performance of ML Models, When Confronted with Data Corrupted by Dynamic Noise of Different Strengths, As Characterized by the Exposure Length n_e ^a

dynamic noise, n_e	MAE		accuracy (%)	
	DL	feature	DL	feature
1	0.207	0.23	78	71
2	0.221	0.23	69.8	71
5	0.229	0.22	59.3	68
10	0.232	0.22	55.2	65
20	0.235	0.22	53.5	65

^aThe case $n_e = 1$ corresponds to no dynamic noise. The table shows the performance for the DL-based method introduced in ref 110 as well as a feature-based method utilizing the features introduced in ref 108.

confronting the DL model introduced in ref 110 and a feature-based model utilizing the features from ref 108 with data corrupted by dynamic noise. For DL, the determination of the anomalous exponent appears to be robust to the influence of dynamic noise, resulting in only a slight performance decrease from a MAE of 0.207 to 0.235 with increasing dynamic noise, as characterized by exposure time steps n_e , considering that a slight performance loss with higher noise is to be expected. The model does, however, seem to struggle with classification for high dynamic noise, where the accuracy drops from 78% down to 53.5% for the highest considered dynamic noise, $n_e = 20$. A look at the confusion matrix in Figure 7 reveals that this is caused by misclassification of LW and CTRW—evidently the added Gaussian noise in the training data is not sufficient to account for the changes incurred by dynamic noise. The feature-based model proves more robust to the influence of dynamic noise, showing a constant MAE of ~ 0.22 – 0.23 , slightly outperforming the DL model for high dynamic noise. Even more striking are the results obtained for classification: while starting with a worse accuracy than DL ($\sim 71\%$ compared to $\sim 78\%$), the feature-based model turns out to be much less hampered by high dynamic noise levels, only decreasing the accuracy to $\sim 65\%$ (compared to $\sim 53.5\%$ for

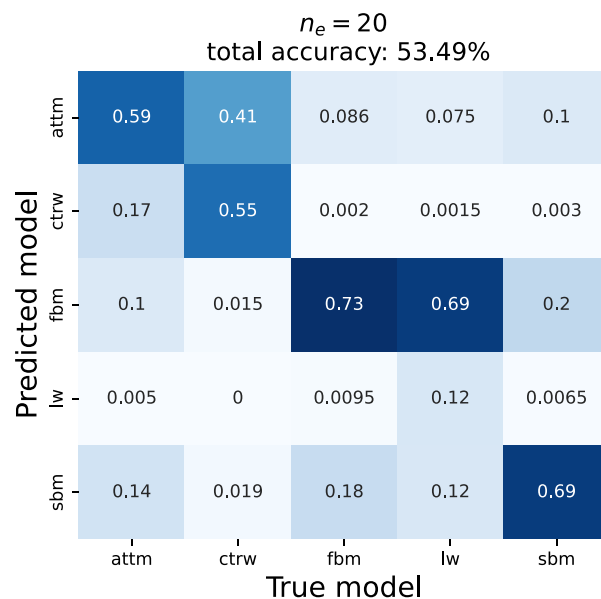


Figure 7. Confusion matrix for dynamic noise with $n_e = 20$ for the DL model introduced in ref 110. While FBM, ATMM, and SBM show behavior similar to that of the case without dynamic noise, identification of CTRW and especially LW is strongly compromised by dynamic noise.

DL) at the highest noise level. Critically, when dealing with experimental setups with high dynamic noise, for accurate classification, dynamic noise should therefore be included in the training data sets, especially when relying on a DL model.

We now consider a stochastic process not contained in the training data. The elephant random walk (ERW) is a process with infinite memory, according to which the next position of the walker is given by

$$x_i = x_{i-1} + \sigma_i \quad (14)$$

with the random variable $\sigma_i = \pm 1$.¹⁴⁴ The choice of σ_i is determined through the memory of the previous time steps by first drawing a random integer $0 \leq j < i$ and then choosing $\sigma_i = \sigma_j$ with probability p or $\sigma_i = -\sigma_j$ with probability $1 - p$. The first step σ_0 is given as $\sigma_0 = 1$ with probability q or $\sigma_0 = -1$ with probability $1 - q$, and for this work, we choose $q = 1/2$. In ref 144 it was shown that, in the limit of many steps, this leads to the long time behavior of the MSD,

$$\langle x^2(t) \rangle \simeq \begin{cases} \frac{t}{3 - 4p}, & p < 3/4 \\ t \ln(t), & p = 3/4 \\ \frac{t^{4p-2}}{(4p-3)\Gamma(4p-2)}, & p > 3/4 \end{cases} \quad (15)$$

which corresponds to normal diffusion for $p < 3/4$ and superdiffusion for $p > 3/4$ with $\alpha = 4p - 2$. We generate a data set containing 10,000 trajectories of length $T = 100$, uniformly distributed in $\alpha \in 1, 1.05, \dots, 2.0$, where for $\alpha = 1$ we choose $p < 3/4$ randomly. To eliminate peculiarities caused by the constant step size and to provide sufficiently many steps to observe long time behavior (eq 15) of the MSD, we only take every $NN = 50$ th or $NN = 200$ th data point, effectively generating trajectories of length 5,000 or 200,000 and shortening them to length 100, plus corrupting them with

white Gaussian noise. The results when confronting the models with this data set are listed in Table 3. For the DL

Table 3. Performance of ML Models When Confronted with Data Generated from the Elephant Random Walk (ERW)^a

	MAE	
	NN = 50	NN = 200
deep-learning model	0.246	0.264
feature-based model	0.348	0.318
trained on subset of NN = 50	0.196	0.175
trained on subset of NN = 200	0.197	0.166

^aThe two cases NN = 50 and NN = 200 correspond to ERW, when one takes every 50th or 200th data point. The table shows the accuracy for the DL-based method introduced in ref 110 as well as a feature-based method utilizing the features introduced in ref 108. For reference, the table also shows what performance can be achieved when the feature-based model is trained on a subset of the ERW test data in the last two rows.

model introduced in ref 110 with this data set, we achieved an MAE of ~ 0.246 (for NN = 50). While this is a significant improvement from the unknown prediction of $\alpha \in [0.05, 2]$ (MAE of ~ 0.49), on which the model was trained, it does not improve much on the performance expected when only identifying the ERW as superdiffusive with $\alpha \in [1, 2]$ (MAE of ~ 0.25). In addition, when considering the uncertainty predictions that are provided by the method as well, we find that the predictions learned on different models are of little to no use when transferred to the ERW. Depicted in Figure 8, we

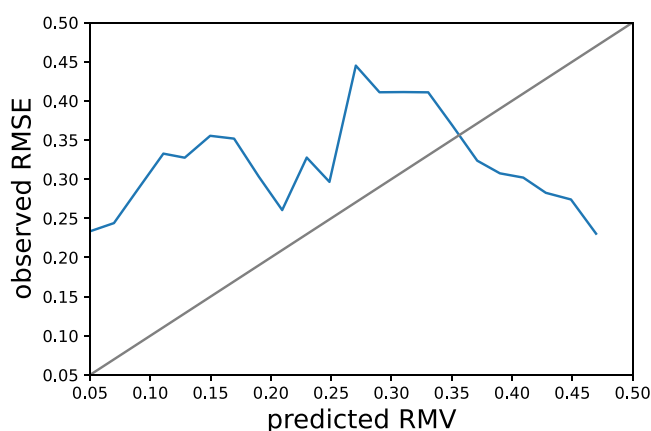


Figure 8. Confidence accuracy diagram obtained when trying to apply the DL model from ref 110 to the elephant random walk (ERW). We see a strong deviation between the predicted root-mean variance and observed root-mean-squared error, indicating that the learning error prediction from the other models does not translate well to the ERW.

see that the predicted and observed errors differ significantly. A possible reason for this can be found by closer inspection of the predicted anomalous diffusion exponent, which reveals that an unusually high number of trajectories are predicted at, or close to, a ballistic motion with $\alpha = 2$ (31% of trajectories are predicted with $\alpha \geq 1.9$ as compared to 9.5% with a true $\alpha \geq 1.9$). This might be caused by the ERW—on a single-trajectory basis in the superdiffusive regime—featuring a drift that gets eliminated only in the ensemble average. Detecting this drift, while not specifically being trained to deal with it, might lead the model to a falsely confident prediction of the ballistic

motion. For the feature-based model in Table 3, we achieved MAEs of only ~ 0.348 and ~ 0.318 for NN = 50 and NN = 200, respectively. For reference, we here also included the results one can obtain when training the feature-based model on subsets of the ERW data sets. These indicate what performance could be expected with appropriately trained models achieving a MAE of ~ 0.318 (for NN = 50) and ~ 0.166 (for NN = 200), thereby significantly improving on the performance of the models trained on the data sets of the AnDi-Challenge. In conclusion, we see that while some information can be extracted, for accurate predictions the machine needs to be trained on the appropriate model.

It should be noted, however, that classical methods, such as Bayesian inference, would similarly struggle when applied to a wrong prior, likely resulting in falsely confident predictions as well. Likewise, it is known that not considering a specific type of noise can lead to wrong predictions when using statistical observables.⁷⁴ Nevertheless, it is important to stress that ML does not circumvent the necessity of considering such cases and should be applied appropriately.

Outlook. The results of the AnDi-Challenge proved the potential of ML approaches when analyzing anomalous diffusion data. They come, however, at some price, often acting as a black box, providing answers without explanation. This lack of explainability limits their usefulness when applied to real-world problems and, inter alia, can lead to some overconfidence in the output results. Building on the AnDi-Challenge, we here presented two methods that improve the machine's explainability.

Extracting a set of statistical features, instead of using the raw position data, allows us to use easier-to-interpret ML algorithms. In addition, one can determine the importance of each feature, further improving interpretability. In a recent publication, Mangalam et al. proposed multifractal features in order to improve the classification of anomalous diffusion,¹⁴⁵ which will be examined in future publications. As an alternative to the trade-off brought by feature-based methods, we can include an uncertainty prediction in the output of deep neural networks using the Multi-SWAG Bayesian DL method on top of the ML algorithm, at no cost of accuracy. Apart from the obvious use of an added reliability estimate, analysis of these error predictions offers additional insights into the learning process of the machine. Note that this method could also be applied to feature-based DL approaches and that similar techniques for gradient boosting or random forest algorithms exist.¹⁴⁶ Despite these improvements, we showed that these methods can still be hampered by out-of-distribution test data, such as noise or models not included in the training data, possibly leading to overconfident predictions. To judge the validity of ML outputs and prepare appropriate training data sets, the analysis of experimental data using statistical methods remains necessary. Moreover, we emphasize that visual inspection as well as some intuition about the system will always present another layer of confidence or caution.

Recent work has shown that sequence-to-sequence models are appropriate to deal with trajectories changing between different diffusion models and/or diffusion exponents,^{147,148} as was the target of the third task of the AnDi-Challenge. Extending such models to include error estimates will be the subject of future work. As an example, we show the results of a preliminary model, trained on trajectories with a single change point, in Figure 9. Apart from its use for uncertainty estimation, the included error output can improve the

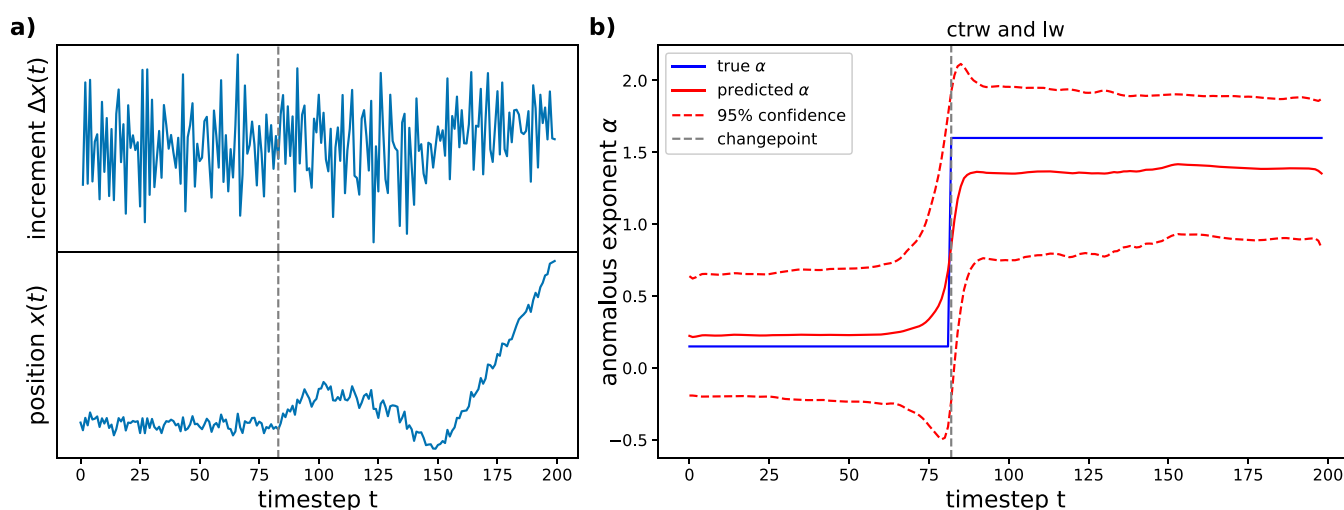


Figure 9. Example of a sequence-to-sequence prediction for a trajectory with a single change point. The trajectory shown in (a) changes from CTRW with $\alpha = 0.15$ to LW with $\alpha = 1.60$ at time step $t_{\text{change}} = 83$. Panel (b) shows the output when feeding case (a) into a trained neural network. The model used includes an uncertainty estimation, the 95% confidence interval of which is indicated by the red dashed line in the figure. Note that the predicted error spikes at the change point.

extraction of change points from the sequence, especially in cases where the anomalous diffusion exponents before and after the change point are similar, inhibiting the determination of change points only by means of the mean predictions.

Similarly to the example of superpositions of diffusion models used here, in a recent work, Muñoz-Gil et al. applied unsupervised learning to anomalous diffusion, where different neural networks are trained to reproduce trajectories generated from a specific diffusion model for each network. They showed that the differing performance in reproduction, when applied to different diffusion models than trained on, can be used to classify a single, or a superposition of, diffusion model(s).¹⁴⁹

On another note, as we saw in Figure 2, predictions on very short trajectories tend to gravitate toward the center of the prior distributions. This will limit the usefulness of single-trajectory analysis when applied to experimental data consisting of many short trajectories. Exploring the applicability of ML techniques to these kinds of data may provide an interesting research avenue in the future; see also the approach in ref 118.

The application of ML, and its comparison to conventional methods, to trajectory ensembles as well as trajectories with changing diffusion models will be the subject of the impending second AnDi-Challenge. In addition, this challenge will include video tracks of diffusing single particles, without direct access to the positions of the tracers, thereby serving as an exploration of noise types different from the simple white Gaussian type, inherent to the conversion from video tracks to particle trajectories.

Recent advances in computer vision could open a new track of research on anomalous diffusion identification. The idea is quite simple: instead of looking for custom neural network architectures for identification purposes or preparing a robust set of features, one could (at least theoretically) turn trajectories into images and feed them into well-established pretrained computer vision models that are known to excel in object recognition. The main difficulty with this approach is that one cannot simply take a plot of a trajectory as the image, since in this case the temporal structure of the data is lost.

Hence, one needs image representations of trajectories that retain the existing spatial and temporal relations.

First approaches utilizing the computer vision approach are very promising. For instance, Garibo-i-Orts et al.¹⁵⁰ used Gramian angular fields to encode trajectories as images and two well-established pretrained computer vision models (ResNet and MobileNet) for both classification of diffusion types and inference of the anomalous diffusion exponent α . Their results for short trajectories already outperform the state-of-the-art-methods. Markov transition fields¹⁵¹ or recurrence plots¹⁵² are other candidates for trajectory imaging methods that could potentially improve the performance of the classifiers. One of the benefits of the computer vision approach is that it allows one to use pretrained models, which are available in popular DL libraries like, for instance, Keras. In other words, it makes the analysis accessible to researchers lacking an extensive background in ML.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jpclett.3c01351>.

Definition of anomalous diffusion models, summary of trajectory characteristics for the feature-based analysis, and Figures S1 and S2 (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Ralf Metzler – Institute of Physics & Astronomy, University of Potsdam, 14476 Potsdam-Golm, Germany; Asia Pacific Center for Theoretical Physics, Pohang 37673, Republic of Korea; orcid.org/0000-0002-6013-7020; Email: rmetzler@uni-potsdam.de

Authors

Henrik Seckler – Institute of Physics & Astronomy, University of Potsdam, 14476 Potsdam-Golm, Germany; orcid.org/0000-0002-7729-433X

Janusz Szwabiński — Hugo Steinhaus Center, Faculty of Pure and Applied Mathematics, Wrocław University of Science and Technology, 50-370 Wrocław, Poland

Complete contact information is available at:
<https://pubs.acs.org/10.1021/acs.jpclett.3c01351>

Notes

The authors declare no competing financial interest.

Biographies

Henrik Seckler studied physics at the University of Potsdam, Germany, where he is currently pursuing his doctorate. His research interests include machine learning, data assimilation, and anomalous diffusion.

Janusz Szwabiński obtained his B.Sc. in physics from University of Wrocław, Poland, and his Ph.D. in science from Saarland University in Saarbrücken, Germany. He is currently a university professor in the Department of Applied Mathematics at the Wrocław University of Science and Technology. He is mainly interested in complex systems and has a track record in multidisciplinary research, with applications in statistical physics, biology, social science, and economy.

Ralf Metzler studied physics and obtained his Ph.D. at the University of Ulm, Germany. After postdoc positions at Tel Aviv University and MIT, he held faculty positions at NORDITA (Nordic Institute for Theoretical Physics, Copenhagen, Denmark), University of Ottawa, Canada, and Technical University of Munich, Germany. Ralf is chair professor in Theoretical Physics at University of Potsdam. Among his distinctions, Ralf received a Finland Distinguished Professorship from the Finnish Academy, an Alexander von Humboldt Polish Honorary Research Scholarship, and the 2017 SigmaPhi Prize. Ralf's research interests are in nonequilibrium statistical physics and stochastic processes in complex systems, soft and bio matter, and data assimilation.

ACKNOWLEDGMENTS

H.S. thanks Timo J. Doerries for the implementation of the MIM model. J.S. was supported by NCN-DFG Beethoven Grant No. 2016/23/G/ST1/04083. R.M. acknowledges funding from the German Ministry for Education and Research (NSF-BMBF project STAXS) and the German Science Foundation (DFG project ME 1535/12-1).

REFERENCES

- (1) Brown, R. brief account of microscopical observations made in the months of June, July and August 1827, on the particles contained in the pollen of plants; and on the general existence of active molecules in organic and inorganic bodies. *Philos. Mag.* **1828**, *4*, 161–173.
- (2) Perrin, J. Brownian movement and molecular reality. *Ann. Chim. Phys.* **1909**, *18*, 5–114.
- (3) Elf, J.; Barkefors, I. Single-molecule kinetics in living cells. *Annu. Rev. Biochem.* **2019**, *88*, 635–659.
- (4) Cherstvy, A. G.; Thapa, S.; Wagner, C. E.; Metzler, R. Non-Gaussian, non-ergodic, and non-Fickian diffusion of tracers in mucin hydrogels. *Soft Matter* **2019**, *15*, 2526–2551.
- (5) Höfling, F.; Franosch, T. Anomalous transport in the crowded world of biological cells. *Rep. Prog. Phys.* **2013**, *76*, 046602.
- (6) Horton, M. R.; Höfling, F.; Rädler, J. O.; Franosch, T. Development of anomalous diffusion among crowding proteins. *Soft Matter* **2010**, *6*, 2648–2656.
- (7) Jeon, J.-H.; Tejedor, V.; Burov, S.; Barkai, E.; Selhuber-Unkel, C.; Berg-Sørensen, K.; Oddershede, L.; Metzler, R. In vivo anomalous diffusion and weak ergodicity breaking of lipid granules. *Phys. Rev. Lett.* **2011**, *106*, 048103.
- (8) Leijnse, N.; Jeon, J. H.; Loft, S.; Metzler, R.; Oddershede, L. B. Diffusion inside living human cells. *Eur. Phys. J. Spec. Top.* **2012**, *204*, 75.
- (9) Codling, E. A.; Plank, M. J.; Benhamou, S. Random walk models in biology. *J. R. Soc. Interface* **2008**, *5*, 813–834.
- (10) Gurtovenko, A. A.; Javanainen, M.; Lolicato, F.; Vattulainen, I. The devil is in the details: what do we really track in single-particle tracking experiments of diffusion in biological membranes? *J. Phys. Chem. Lett.* **2019**, *10*, 1005–1011.
- (11) Roberts, T. D.; Yuan, R.; Xiang, L.; Delor, M.; Pokhrel, R.; Yang, K.; Aqad, E.; Marangoni, T.; Trefonas, P.; Xu, K.; Ginsberg, N. S. Direct correlation of single-particle motion to amorphous microstructural components of semicrystalline poly(ethylene oxide) electrolytic films. *J. Phys. Chem. Lett.* **2020**, *11*, 4849–4858.
- (12) Erimban, S.; Daschakraborty, S. Fickian yet non-Gaussian nanoscopic lipid diffusion in the raft-mimetic membrane. *J. Phys. Chem. B* **2023**, *127*, 4939–4951.
- (13) Javanainen, M.; Martinez-Seara, H.; Metzler, R.; Vattulainen, I. Diffusion of integral membrane proteins in protein-rich membranes. *J. Phys. Chem. Lett.* **2017**, *8*, 4308–4313.
- (14) Winkler, P. M.; Regmi, R.; Flauraud, V.; Brugger, J.; Rigneault, H.; Wenger, J.; Garcia-Parajo, M. F. Optical antenna-based fluorescence correlation spectroscopy to probe the nanoscale dynamics of biological membranes. *J. Phys. Chem. Lett.* **2018**, *9*, 110–119.
- (15) Spillane, K. M.; Ortega-Arroyo, J.; de Wit, G.; Eggeling, C.; Ewers, H.; Wallace, M. I.; Kukura, P. High-speed single-particle tracking of GM1 in model membranes reveals anomalous diffusion due to interleaflet coupling and molecular pinning. *Nano Lett.* **2014**, *14*, 5390–5397.
- (16) Okubo, A. Dynamical aspects of animal grouping: swarms, schools, flocks, and herds. *Adv. Biophys.* **1986**, *22*, 1–94.
- (17) Vilks, O.; Aghion, E.; Avgar, T.; Beta, C.; Nagel, O.; Sabri, A.; Sarfati, R.; Schwartz, D. K.; Weiss, M.; Krapf, D.; Nathan, R.; Metzler, R.; Assaf, M. Unravelling the origins of anomalous diffusion: from molecules to migrating storks. *Phys. Rev. Res.* **2022**, *4*, 033055.
- (18) Bartumeus, F.; da Luz, M. G. E.; Viswanathan, G. M.; Catalan, J. Animal search strategies: a quantitative random-walk analysis. *Ecology* **2005**, *86*, 3078–3087.
- (19) Engbert, R.; Mergenthaler, K.; Sinn, P.; Pikovskiy, A. An integrated model of fixational eye movements and microsaccades. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, E765–E770.
- (20) Herrmann, C. J. J.; Metzler, R.; Engbert, R. A self-avoiding walk with neural delays as a model of fixational eye movements. *Sci. Rep.* **2017**, *7*, 12958.
- (21) Malkiel, B. G. *A random walk down wall street: including a life-cycle guide to personal investing*; W. Norton & Co: New York, 1999.
- (22) Plerou, V.; Gopikrishnan, P.; Amaral, L. A. N.; Gabaix, X.; Stanley, H. E. Economic fluctuations and anomalous diffusion. *Phys. Rev. E* **2000**, *62*, R3023.
- (23) Fernández, R.; Fröhlich, J.; Sokal, A. D. *Random walks, critical phenomena, and triviality in quantum field theory*; Springer Science & Business Media: Berlin, 2013.
- (24) Anderson, J. B. Quantum chemistry by random walk. H^2P , H^+_3 , $D^{3h}_{3h}A_1$, $H^3_2\Sigma^+_u$, $H^1_4\Sigma^+_g$, Be^1S . *J. Chem. Phys.* **1976**, *65*, 4121–4127.
- (25) Lüdtke, O.; Roberts, B. W.; Trautwein, U.; Nagy, G. A random walk down university avenue: life paths, life events, and personality trait change at the transition to university life. *J. Pers. Soc. Psychol.* **2011**, *101*, 620.
- (26) Bouchaud, J.-P.; Potters, M. *Theory of financial risk and derivative pricing: from statistical physics to risk management*; Cambridge University Press: Cambridge, 2003.
- (27) Pearson, K. The problem of the random walk. *Nature* **1905**, *72*, 294.
- (28) Mises, R. V. Fundamentalsätze der Wahrscheinlichkeitsrechnung. *Math. Z.* **1919**, *4*, 1–97.
- (29) Montroll, E. W.; Weiss, G. H. Random walks on lattices. II. *J. Math. Phys.* **1965**, *6*, 167–181.

- (30) Van Kampen, N. G. *Stochastic processes in chemistry and physics*; Elsevier, 1992.
- (31) Lévy, P. *Processus stochastiques et mouvement brownien*; Gauthier-Villars: Paris, 1948.
- (32) Hughes, B. D. *Random walks and random environments*; Oxford University Press: Oxford, 1995.
- (33) Einstein, A. Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen. *Ann. Phys.* **1905**, 322, 549–560.
- (34) Von Smoluchowski, M. Zur kinetischen Theorie der Brownschen Molekularbewegung und der Suspensionen. *Ann. Phys.* **1906**, 326, 756–780.
- (35) Sutherland, W. A dynamical theory of diffusion for non-electrolytes and the molecular mass of albumin. *Philos. Mag.* **1905**, 9, 781–785.
- (36) Langevin, P. Sur la théorie du mouvement brownien. *C. R. Acad. Sci. (Paris)* **1908**, 146, 530–533.
- (37) Bouchaud, J.-P.; Georges, A. Anomalous diffusion in disordered media: statistical mechanisms, models and physical applications. *Phys. Rep.* **1990**, 195, 127–293.
- (38) Metzler, R.; Klafter, J. The random walk's guide to anomalous diffusion: a fractional dynamics approach. *Phys. Rep.* **2000**, 339, 1–77.
- (39) Golding, I.; Cox, E. C. Physical nature of bacterial cytoplasm. *Phys. Rev. Lett.* **2006**, 96, 098102.
- (40) Manzo, C.; Torreno-Pina, J. A.; Massignan, P.; Lapeyre, G. J., Jr.; Lewenstein, M.; Parajo, M. F. G. Weak ergodicity breaking of receptor motion in living cells stemming from random diffusivity. *Phys. Rev. X* **2015**, 5, 011021.
- (41) Krapf, D.; Lukat, N.; Marinari, E.; Metzler, R.; Oshanin, G.; Selhuber-Unkel, C.; Squarcini, A.; Stadler, L.; Weiss, M.; Xu, X. Spectral content of a single non-Brownian trajectory. *Phys. Rev. X* **2019**, 9, 011019.
- (42) Stadler, L.; Weiss, M. Non-equilibrium forces drive the anomalous diffusion of telomeres in the nucleus of mammalian cells. *New J. Phys.* **2017**, 19, 113048.
- (43) Kindermann, F.; Dechant, A.; Hohmann, M.; Lausch, T.; Mayer, D.; Schmidt, F.; Lutz, E.; Widera, A. Nonergodic diffusion of single atoms in a periodic potential. *Nat. Phys.* **2017**, 13, 137–141.
- (44) Sokolov, I. M. Models of anomalous diffusion in crowded environments. *Soft Matter* **2012**, 8, 9043–9052.
- (45) Saxton, M. J. Anomalous diffusion due to obstacles: a Monte Carlo study. *Biophys. J.* **1994**, 66, 394–401.
- (46) Saxton, M. J. Anomalous subdiffusion in fluorescence photobleaching recovery: a Monte Carlo study. *Biophys. J.* **2001**, 81, 2226–2240.
- (47) Burov, S.; Jeon, J. H.; Metzler, R.; Barkai, E. Single particle tracking in systems showing anomalous diffusion: the role of weak ergodicity breaking. *Phys. Chem. Chem. Phys.* **2011**, 13, 1800–1812.
- (48) Ernst, D.; Köhler, J.; Weiss, M. Probing the type of anomalous diffusion with single-particle tracking. *Phys. Chem. Chem. Phys.* **2014**, 16, 7686–7691.
- (49) Bodrova, A.; Chechkin, A. V.; Cherstvy, A. G.; Metzler, R. Quantifying non-ergodic dynamics of force-free granular gases. *Phys. Chem. Chem. Phys.* **2015**, 17, 21791–21798.
- (50) Bodrova, A.; Chechkin, A. V.; Cherstvy, A. G.; Safdari, H.; Sokolov, I. M.; Metzler, R. Underdamped scaled Brownian motion: (non-)existence of the overdamped limit in anomalous diffusion. *Sci. Rep.* **2016**, 6, 30520.
- (51) Lim, S. C.; Muniandy, S. V. Self-similar Gaussian processes for modeling anomalous diffusion. *Phys. Rev. E* **2002**, 66, 021114.
- (52) Jeon, J.-H.; Chechkin, A. V.; Metzler, R. Scaled Brownian motion: a paradoxical process with a time dependent diffusivity for the description of anomalous diffusion. *Phys. Chem. Chem. Phys.* **2014**, 16, 15811–15817.
- (53) Mandelbrot, B. B.; Van Ness, J. W. Fractional Brownian motions, fractional noises and applications. *SIAM Rev.* **1968**, 10, 422–437.
- (54) Metzler, R.; Jeon, J. H.; Cherstvy, A. G.; Barkai, E. Anomalous diffusion models and their properties: non-stationarity, non-ergodicity, and ageing at the centenary of single particle tracking. *Phys. Chem. Chem. Phys.* **2014**, 16, 24128–24164.
- (55) Marcone, B.; Nampoothiri, S.; Orlandini, E.; Seno, F.; Baldovin, F. Brownian non-Gaussian diffusion of self-avoiding walks. *J. Phys. A* **2022**, 55, 354003.
- (56) Hegde, A. S.; Chandrashekar, C. M. Characterization of anomalous diffusion in one-dimensional quantum walks. *J. Phys. A* **2022**, 55, 234006.
- (57) Vitali, S.; Paradisi, P.; Pagnini, G. Anomalous diffusion originated by two Markovian hopping-trap mechanisms. *J. Phys. A* **2022**, 55, 224012.
- (58) Sabzikar, F.; Kabala, J.; Burnecki, K. Tempered fractionally integrated process with stable noise as a transient anomalous diffusion model. *J. Phys. A* **2022**, 55, 174002.
- (59) Wang, W.; Cherstvy, A. G.; Chechkin, A. V.; Thapa, S.; Seno, F.; Liu, X.; Metzler, R. Fractional Brownian motion with random diffusivity: emerging residual nonergodicity below the correlation time. *J. Phys. A* **2020**, 53, 474001.
- (60) Hughes, B. D.; Shlesinger, M. F.; Montroll, E. W. Random walks with self-similar clusters. *Proc. Natl. Acad. Sci. U.S.A.* **1981**, 78, 3287–3291.
- (61) Weissman, H.; Weiss, G. H.; Havlin, S. Transport properties of the continuous-time random walk with a long-tailed waiting-time density. *J. Stat. Phys.* **1989**, 57, 301–317.
- (62) Lévy, P. *Théorie de l'addition des variables aléatoires*; Gauthier-Villars: Paris, 1937.
- (63) Chechkin, A. V.; Metzler, R.; Klafter, J.; Gonchar, V. Y. Introduction to the theory of Lévy flights. In *Anomalous Transport: Foundations and Applications*; Klages, R., Radons, G., Sokolov, I. M., Eds.; Springer: Berlin, 2008; pp 129–162.
- (64) Shlesinger, M. F.; Klafter, J. Lévy walks versus Lévy flights, in *On growth and form*; Springer: Dordrecht, 1986.
- (65) Zaburdaev, V.; Denisov, S.; Klafter, J. Lévy walks. *Rev. Mod. Phys.* **2015**, 87, 483.
- (66) Magdziarz, M.; Zorawik, T. Limit properties of Lévy walks. *J. Phys. A* **2020**, 53, 504001.
- (67) Massignan, P.; Manzo, C.; Torreno-Pina, J. A.; García-Parajo, M. F.; Lewenstein, M.; Lapeyre, G. J., Jr. Nonergodic subdiffusion from Brownian motion in an inhomogeneous medium. *Phys. Rev. Lett.* **2014**, 112, 150603.
- (68) Meroz, Y.; Sokolov, I. M. A toolbox for determining subdiffusive mechanisms. *Phys. Rep.* **2015**, 573, 1–29.
- (69) Makarava, N.; Benmehdi, S.; Holschneider, M. Bayesian estimation of self-similarity exponent. *Phys. Rev. E* **2011**, 84, 021109.
- (70) Metzler, R.; Tejedor, V.; Jeon, J. H.; He, Y.; Deng, W. H.; Burov, S.; Barkai, E. Analysis of single particle trajectories: from normal to anomalous diffusion. *Acta Phys. Polon., B* **2009**, 40 (5), 1315–1331.
- (71) Magdziarz, M.; Weron, A.; Burnecki, K.; Klafter, J. Fractional Brownian motion versus the continuous-time random walk: a simple test for subdiffusive dynamics. *Phys. Rev. Lett.* **2009**, 103, 180602.
- (72) Metzler, R. Brownian motion and beyond: first-passage, power spectrum, non-Gaussianity, and anomalous diffusion. *J. Stat. Mech.* **2019**, 2019, 114003.
- (73) Vilk, O.; Aghion, E.; Nathan, R.; Toledo, S.; Metzler, R.; Assaf, M. Classification of anomalous diffusion in animal movement data using power spectral analysis. *J. Phys. A* **2022**, 55, 334004.
- (74) Sposini, V.; Krapf, D.; Marinari, E.; Sunyer, R.; Ritort, F.; Taheri, F.; Selhuber-Unkel, C.; Benelli, R.; Weiss, M.; Metzler, R.; Oshanin, G. Towards a robust criterion of anomalous diffusion. *Commun. Phys.* **2022**, 5, 305.
- (75) Condamine, S.; Bénichou, O.; Tejedor, V.; Voituriez, R.; Klafter, J. First-passage times in complex scale-invariant media. *Nature* **2007**, 450, 77–80.
- (76) Slezak, J.; Metzler, R.; Magdziarz, M. Codifference can detect ergodicity breaking and non-Gaussianity. *New J. Phys.* **2019**, 21, 053008.

- (77) Maraj, K.; Szarek, D.; Sikora, G.; Wylomańska, A. empirical anomaly measure for finite-variance processes. *J. Phys. A* **2021**, *54*, 024001.
- (78) Chen, L.; Bassler, K. E.; McCauley, J. L.; Gunaratne, G. H. Anomalous scaling of stochastic processes and the Moses effect. *Phys. Rev. E* **2017**, *95*, 042141.
- (79) Martin, D. S.; Forstner, M. B.; Käs, J. A. Apparent subdiffusion inherent to single particle tracking. *Biophys. J.* **2002**, *83*, 2109–2117.
- (80) Prezhdo, O. V. Advancing physical chemistry with machine learning. *J. Phys. Chem. Lett.* **2020**, *11*, 9656–9658.
- (81) Back, S.; Yoon, J.; Tian, N.; Zhong, W.; Tran, K.; Ulissi, Z. W. Convolutional neural network of atomic surface structures to predict binding energies for high-throughput screening of catalysts. *J. Phys. Chem. Lett.* **2019**, *10*, 4401–4408.
- (82) Batra, R.; Chan, H.; Kamath, G.; Ramprasad, R.; Cherukara, M. J.; Sankaranarayanan, S. Screening of therapeutic agents for COVID-19 using machine learning and ensemble docking studies. *J. Phys. Chem. Lett.* **2020**, *11*, 7058–7065.
- (83) Dral, P. O. Quantum chemistry in the age of machine learning. *J. Phys. Chem. Lett.* **2020**, *11*, 2336–2347.
- (84) Granik, N.; Weiss, L. E.; Nehme, E.; Levin, M.; Chein, M.; Perlson, E.; Roichman, Y.; Shechtman, Y. Single-particle diffusion characterization by deep learning. *Biophys. J.* **2019**, *117*, 185–192.
- (85) Bo, S.; Schmidt, F.; Eichhorn, R.; Volpe, G. Measurement of anomalous diffusion using recurrent neural networks. *Phys. Rev. E* **2019**, *100*, 010102.
- (86) Muñoz-Gil, G.; Garcia-March, M. A.; Manzo, C.; Martín-Guerrero, J. D.; Lewenstein, M. Single trajectory characterization via machine learning. *New J. Phys.* **2020**, *22*, 013010.
- (87) Muñoz-Gil, G.; Volpe, G.; García-March, M. A.; Metzler, R.; Lewenstein, M.; Manzo, C. The anomalous diffusion challenge: single trajectory characterisation as a competition. *Emerging Topics in Artificial Intelligence 2020*, Vol. 11469; SPIE, 2020.
- (88) Gajowczyk, M.; Szwabiński, J. Detection of anomalous diffusion with deep residual networks. *Entropy* **2021**, *23*, 649.
- (89) Manzo, C.; Muñoz-Gil, G.; Volpe, G.; Garcia-March, M. A.; Lewenstein, M.; Metzler, R. Preface: Characterisation of physical processes from anomalous diffusion data. *J. Phys. A* **2023**, *56*, 010401.
- (90) Burnecki, K.; Weron, A. Fractional Lévy stable motion can model subdiffusive dynamics. *Phys. Rev. E* **2010**, *82*, 021130.
- (91) Loch-Olszewska, H.; Szwabiński, J. Impact of feature choice on machine learning classification of fractional anomalous diffusion. *Entropy* **2020**, *22*, 1436.
- (92) Jeon, J.-H.; Barkai, E.; Metzler, R. Noisy continuous time random walks. *J. Chem. Phys.* **2013**, *139*, 121916.
- (93) Mandelbrot, B. B.; Wallis, J. R. Noah, Joseph, and operational hydrology. *Water Resour. Res.* **1968**, *4*, 909–918.
- (94) Aghion, E.; Meyer, P. G.; Adlakha, V.; Kantz, H.; Bassler, K. E. Moses, Noah and Joseph effects in Lévy walks. *New J. Phys.* **2021**, *23*, 023002.
- (95) Meyer, P. G.; Aghion, E.; Kantz, H. Decomposing the effect of anomalous diffusion enables direct calculation of the Hurst exponent and model classification for single random paths. *J. Phys. A* **2022**, *55*, 274001.
- (96) Thapa, S.; Lomholt, M. A.; Krog, J.; Cherstvy, A. G.; Metzler, R. 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.* **2018**, *20*, 29018–29037.
- (97) Park, S.; Thapa, S.; Kim, Y.; Lomholt, M. A.; Jeon, J.-H. Bayesian inference of Lévy walks via hidden Markov models. *J. Phys. A* **2021**, *54*, 484001.
- (98) Szegedy, C.; Zaremba, W.; Sutskever, I.; Bruna, J.; Erhan, D.; Goodfellow, I.; Fergus, R. Intriguing properties of neural networks. *2nd International Conference on Learning Representations; ICLR*, 2014.
- (99) Kowalek, P.; Loch-Olszewska, H.; Szwabiński, J. Classification of diffusion modes in single-particle tracking data: feature-based versus deep-learning approach. *Phys. Rev. E* **2019**, *100*, 032410.
- (100) Muñoz-Gil, G.; Volpe, G.; García-March, M. A.; et al. Objective comparison of methods to decode anomalous diffusion. *Nat. Commun.* **2021**, *12*, 6253.
- (101) Argun, A.; Volpe, G.; Bo, S. Classification, inference and segmentation of anomalous diffusion with recurrent neural networks. *J. Phys. A* **2021**, *54*, 294003.
- (102) Garibo-i-Orts, Ò.; Baeza-Bosca, A.; Garcia-March, A.; Conejero, J. A. Efficient recurrent neural network methods for anomalously diffusing single particle short and noisy trajectories. *J. Phys. A* **2021**, *54*, 504002.
- (103) Li, D.; Yao, Q.; Huang, Z. WaveNet-based deep neural networks for the characterization of anomalous diffusion (WADNet). *J. Phys. A* **2021**, *54*, 404003.
- (104) Firbas, N.; Garibo-i-Orts, Ò.; Garcia-March, M. Á.; Conejero, J. A. Characterization of anomalous diffusion through convolutional transformers. *J. Phys. A* **2023**, *56*, 014001.
- (105) AL-hada, E. A.; Tang, X.; Deng, W. Classification of stochastic processes by convolutional neural networks. *J. Phys. A* **2022**, *55*, 274006.
- (106) Gentili, A.; Volpe, G. Characterization of anomalous diffusion classical statistics powered by deep learning (CONDOR). *J. Phys. A* **2021**, *54*, 314003.
- (107) Manzo, C. Extreme learning machine for the characterization of anomalous diffusion from single trajectories (andi-elm). *J. Phys. A* **2021**, *54*, 334002.
- (108) Kowalek, P.; Loch-Olszewska, H.; Łaszczuk, Ł.; Opala, J.; Szwabiński, J. Boosting the performance of anomalous diffusion classifiers with the proper choice of features. *J. Phys. A* **2022**, *55*, 244005.
- (109) Pinholt, H. D.; Bohr, S. S. R.; Iversen, J. F.; Boomsma, W.; Hatzakis, N. S. Single-particle diffusional fingerprinting: a machine-learning framework for quantitative analysis of heterogeneous diffusion. *Proc. Natl. Acad. Sci. U.S.A.* **2021**, *118*, No. e2104624118.
- (110) Seckler, H.; Metzler, R. Bayesian deep learning for error estimation in the analysis of anomalous diffusion. *Nat. Commun.* **2022**, *13*, 6717.
- (111) Thapa, S.; Park, S.; Kim, Y.; Jeon, J. H.; Metzler, R.; Lomholt, M. A. Bayesian inference of scaled versus fractional Brownian motion. *J. Phys. A* **2022**, *55*, 194003.
- (112) Muñoz-Gil, G.; Requena, B.; Volpe, G.; Garcia-March, M. A.; Manzo, C. *AnDiChallenge/ANDI_datasets: Challenge 2020 release (v.1.0)*, DOI: 10.5281/zenodo.4775311.
- (113) Krog, J.; Jacobsen, L. H.; Lund, F. W.; Wüstner, D.; Lomholt, M. A. Bayesian model selection with fractional Brownian motion. *J. Stat. Mech.* **2018**, 093501.
- (114) Janczura, J.; Kowalek, P.; Loch-Olszewska, H.; Szwabiński, J.; Weron, A. Classification of particle trajectories in living cells: machine learning versus statistical testing hypothesis for fractional anomalous diffusion. *Phys. Rev. E* **2020**, *102*, 032402.
- (115) Verdier, H.; Duval, M.; Laurent, F.; Cassé, A.; Vestergaard, C. L.; Masson, J. B. Learning physical properties of anomalous random walks using graph neural networks. *J. Phys. A* **2021**, *54*, 234001.
- (116) Verdier, H.; Laurent, F.; Cassé, A.; Vestergaard, C. L.; Masson, J. B. Variational inference of fractional Brownian motion with linear computational complexity. *Phys. Rev. E* **2022**, *106*, 055311.
- (117) LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444.
- (118) Chen, Z.; Geffroy, L.; Biteen, J. S. NOBIAS: Analyzing anomalous diffusion in single-molecule tracks with nonparametric Bayesian inference. *Front. Bioinform.* **2021**, *1*, 742073.
- (119) Fukushima, K. neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biol. Cybern.* **1980**, *36*, 193–202.
- (120) Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput* **1997**, *9*, 1735–1780.
- (121) Kiureghian, A.; Ditlevsen, O. Aleatory or epistemic? Does it matter? *Structural Safety* **2009**, *31*, 105–112.

- (122) Kendall, A.; Gal, Y. What uncertainties do we need in Bayesian deep learning for computer vision? *Adv. Neural Inf. Process. Syst.* **2017**, 30, 5580–5590.
- (123) Wang, Q.; Ma, Y.; Zhao, K.; Tian, Y. A comprehensive survey of loss functions in machine learning. *Ann. Data Sci.* **2022**, 9, 187.
- (124) Nix, D. A.; Weigend, A. S. Estimating the mean and variance of the target probability distribution. *Proc. 1994 IEEE Int. Conf. Neural Networks (ICNN'94)*, Vol. 1; IEEE, 1994.
- (125) Kolmogorov, A. N.; *Foundations of the theory of probability*; Chelsea Publishing Co.: London, 1950.
- (126) Metropolis, N.; Ulam, S. The Monte Carlo method. *J. Am. Stat. Assoc.* **1949**, 44, 335–34.
- (127) Binder, K.; *Monte Carlo methods in statistical physics*; Springer: Berlin, 1986.
- (128) Lakshminarayanan, B.; Pritzel, A.; Blundell, C. Simple and scalable predictive uncertainty estimation using deep ensembles. *Adv. Neural Inf. Process. Syst.* **2017**, 30.
- (129) Gal, Y.; Ghahramani, Z. Dropout as a Bayesian approximation: representing model uncertainty in deep learning. *Int. Conf. Machine Learning*; PMLR, 2016.
- (130) Gal, Y. *Uncertainty in deep learning*. Doctoral dissertation, Cambridge University: Cambridge, 2016.
- (131) Maddox, W. J.; Izmailov, P.; Garipov, T.; Vetrov, D. P.; Wilson, A. G. A simple baseline for Bayesian uncertainty in deep learning. *Adv. Neural Inf. Process. Syst.* **2019**, 32.
- (132) Wilson, A. G.; Izmailov, P. Bayesian deep learning and a probabilistic perspective of generalization. *Adv. Neural Inf. Process. Syst.* **2020**, 33, 4697–4708.
- (133) Bottou, L. Large-scale machine learning with stochastic gradient descent. *Proceedings of COMPSTAT'2010: 19th International Conference on Computational Statistics*, Paris, France, Aug 22–27, 2010; Keynote, Invited and Contributed Papers; Physica-Verlag HD, 2010; pp 177–186.
- (134) Naeini, M. P.; Cooper, G.; Hauskrecht, M. Obtaining well calibrated probabilities using Bayesian binning. *Proceedings of the AAAI conference on artificial intelligence*, Vol. 29. No. 1; 2015.
- (135) Levi, D.; Gispan, L.; Giladi, N.; Fetaya, E. Evaluating and calibrating uncertainty prediction in regression tasks. *Sensors* **2022**, 22, 5540.
- (136) Doerries, T. J.; Chechkin, A. V.; Metzler, R. Apparent anomalous diffusion and non-Gaussian distributions in a simple mobile-immobile transport model with Poissonian switching. *J. R. Soc. Interface* **2022**, 19, 20220233.
- (137) Van Genuchten, M. T.; Wierenga, P. J. Mass transfer studies in sorbing porous media I. analytical solutions. *Soil Sci. Soc. Am. J.* **1976**, 40, 473–480.
- (138) Deans, H. H. A mathematical model for dispersion in the direction of flow of porous media. *Soc. Pet. Eng. J.* **1963**, 3, 49–52.
- (139) Wagner, T.; Kroll, A.; Haramagatti, C. R.; Lipinski, H.-G.; Wiemann, M. Classification and segmentation of nanoparticle diffusion trajectories in cellular micro environments. *PLoS One* **2017**, 12, No. e0170165.
- (140) Weron, A.; Janczura, J.; Boryczka, E.; Sungkaworn, T.; Calebiro, D. Statistical testing approach for fractional anomalous diffusion classification. *Phys. Rev. E* **2019**, 99, 042149.
- (141) James, G.; Witten, D.; Hastie, T.; Tibshirani, R. *An introduction to statistical learning with applications in R*; Springer: New York, 2013.
- (142) Song, Y.-Y.; LU, Y. Decision tree methods: applications for classification and prediction. *Shanghai Archives of Psychiatry* **2015**, 27, 130.
- (143) Meyer, P. G.; Metzler, R. Stochastic processes in a confining harmonic potential in the presence of static and dynamic measurement noise. *New J. Phys.* **2023**, 25, 063003.
- (144) Schütz, G. M.; Trimper, S. Elephants can always remember: Exact long-range memory effects in a non-Markovian random walk. *Phys. Rev. E* **2004**, 70, 045101.
- (145) Mangalam, M.; Metzler, R.; Kelty-Stephen, D. G. Ergodic characterization of non-ergodic anomalous diffusion processes. *Phys. Rev. Res.* **2023**, 5, 023144.
- (146) Malinin, A.; Prokhorenkova, L.; Ustimenko, A. Uncertainty in gradient boosting via ensembles. *International Conference on Learning Representations*, 2020; E-print arXiv: 2006.10562.
- (147) Requena, B.; Masó, S.; Bertran, J.; Lewenstein, M.; Manzo, C.; Muñoz-Gil, G. *Inferring pointwise diffusion properties of single trajectories with deep learning*, 2023; arXiv: 2302.00410.
- (148) Martinez, Q.; Chen, C.; Xia, J.; Bahai, H. Sequence-to-sequence change-point detection in single-particle trajectories via recurrent neural network for measuring self-diffusion. *Transp. Porous Media* **2023**, 147, 679–701.
- (149) Muñoz-Gil, G.; Guigo i Corominas, G.; Lewenstein, M. Unsupervised learning of anomalous diffusion data: an anomaly detection approach. *J. Phys. A* **2021**, 54, 504001.
- (150) Garibo-i-Orts, Ó.; Firbas, N.; Sebastián, L.; Conejero, J. A. Gramian angular fields for leveraging pretrained computer vision models with anomalous diffusion trajectories. *Phys. Rev. E* **2023**, 107, 034138.
- (151) Wang, Z.; Oates, T. Imaging time-series to improve classification and imputation. *Proceedings of the 24th International Conference on Artificial Intelligence*, 2015.
- (152) Eckmann, J.-P.; Oliffson Kamphorst, S.; Ruelle, D. Recurrence plots of dynamical systems. *EPL* **1987**, 4, 973.