PAPER • **OPEN ACCESS**

Change-point detection in anomalous-diffusion trajectories utilising machine-learning-based uncertainty estimates

To cite this article: Henrik Seckler and Ralf Metzler 2024 *J. Phys. Photonics* **6** 045025

Journal of Physics: Photonics

PAPER

Change-point detection in anomalous-diffusion trajectories utilising machine-learning-based uncertainty estimates

Henrik Seckler[1] and Ralf Metzler[1,2,*]

[1] Institute for Physics & Astronomy, University of Potsdam, 14476 Potsdam-Golm, Germany
[2] Asia Pacific Centre for Theoretical Physics, Pohang 37673, Republic of Korea
[*] Author to whom any correspondence should be addressed.

E-mail: rmetzler@uni-potsdam.de

Keywords: diffusion, anomalous diffusion, change-point analysis, AnDi 2 challenge, Bayesian-deep learning analysis

OPEN ACCESS

RECEIVED
5 September 2024

REVISED
10 October 2024

ACCEPTED FOR PUBLICATION
17 October 2024

PUBLISHED
30 October 2024

Abstract

When recording the movement of individual animals, cells or molecules one will often observe changes in their diffusive behaviour at certain points in time along their trajectory. In order to capture the different diffusive modes assembled in such heterogeneous trajectories it becomes necessary to segment them by determining these change-points. Such a change-point detection can be challenging for conventional statistical methods, especially when the changes are subtle. We here apply *Bayesian Deep Learning* to obtain point-wise estimates of not only the anomalous diffusion exponent but also the uncertainties in these predictions from a single anomalous diffusion trajectory generated according to four theoretical models of anomalous diffusion. We show that we are able to achieve an accuracy similar to single-mode (without change-points) predictions as well as a well calibrated uncertainty predictions of this accuracy. Additionally, we find that the predicted uncertainties feature interesting behaviour at the change-points leading us to examine the capabilities of these predictions for change-point detection. While the series of predicted uncertainties on their own are not sufficient to improve change-point detection, they do lead to a performance boost when applied in combination with the predicted anomalous diffusion exponents.

1. Introduction

In ancient Rome, around 60 BC, the philosopher Lucretius gives a remarkable description in his poem '*De rerum natura*' ('*On the Nature of Things*'), observing the motion of dust particles in a ray of light in a stairwell [1]. While what he called the 'dancing' of particles is predominantly caused by air currents, he provided a surprisingly accurate explanation for the much later understood phenomenon of Brownian motion [2, 3]. He postulated that the motion must be caused by smaller unseen particles colliding into larger ones until cascading to the visible motion.

Mathematically such a motion is modelled as a path of successive random steps, called a random walk [4, 5]. As long as these increments $\Delta x(t)$ are independently, identically distributed with finite variance, they will, under the central limit theorem (CLT), lead to normal diffusion [6]. The prime example of this is the aforementioned Brownian motion, as observed by Robert Brown in 1827, when tracking pollen granules suspended in water [2, 3, 7–9]. Amongst others, normal diffusion entails a Gaussian probability density function (PDF) of the position $x(t)$ with a linear growth of the mean squared displacement (MSD) [10, 11][3]

$$\langle x^2(t)\rangle \sim 2Dt. \tag{1}$$

[3] We here consider the one-dimensional case. Future work will study the performance of the developed algorithm in higher dimensions.

In reality however, many experiments show a non-linear growth of the MSD [12–29], indicating that one or multiple conditions of the CLT are broken. Often these systems feature a power-law growth

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

of the MSD, which is referred to as anomalous diffusion with anomalous diffusion exponent α and generalised diffusion constant K_α [12, 13, 26]. A motion with slower-than-normal growth of the MSD (with $0 < \alpha < 1$) is called subdiffusive, whereas for $\alpha > 1$ it is referred to as superdiffusive. Mathematically, anomalous diffusion may be achieved in many ways, all breaking at least one condition of the CLT [12, 13, 30]. Some examples of this include long-range correlated increments in fractional Brownian motion (FBM) [31], non-identically distributed increments in scaled Brownian motion (SBM) [32, 33] or increment-distributions with infinite variance in Lévy flights [34, 35]. Additionally, introducing waiting times between jumps in a so called continuous time random walk (CTRW) allows for further options to reach anomalous diffusion [36–40]. A more detailed description of these models is provided in section 2.1.

As each of these models corresponds to a different cause of anomalous diffusion, determining the model best fitting experimental data can yield insights into the underlying diffusive mechanisms. Furthermore one may wish to determine the associated parameters such as the anomalous diffusion exponent α or the generalised diffusion constant K_α. The experimental data provided for these tasks usually consist of single particle trajectories, tracking the movement of molecules in cells [20–22, 41–44], animals [45–47] or even stock prices [48, 49].

The tools to decipher the stochastic processes underlying measured data and the associate parameters are various, ranging from the use of several statistical observables [27, 50–61] to Bayesian inference [62–64] or machine learning [65–79]. The latter demonstrated its potential in the *Anomalous Diffusion (AnDi-) Challenge* held in 2020 [80, 81]. In this challenge participants were tasked with the determination of the anomalous diffusion model and its anomalous diffusion exponent α from computer-generated single-particle trajectories. Especially when data is sparse, machine-learning models tended to outperform the conventional ones.

In a recent paper we showed that using *Bayesian Deep Learning*, one can expand the machine learning solutions to the *AnDi-Challenge* to include an uncertainty estimate [65, 66]. This estimate allows for an easier assessment whether predictions are reliable and improves the interpretability of the algorithm.

We here aim to explore the applicability of this machine-learning model to diffusion trajectories in which the walker may change between diffusive modes. Such a behaviour may occur, for instance, when the tracked animal or cell switches between a searching and feeding state, as an effective behaviour in an heterogeneous environment or due to transient interactions with other objects [46, 82–87]. If one desires to differentiate between these diffusive modes it is imperative to be able to segment the trajectory by identifying the points at which the changes occur [88–92].

Applications of machine-learning techniques to this problem are explored in the second *AnDi-Challenge* [93], whose results are expected to be published in the near future; note however that the setup differs strongly from the one described herein and can thereby not be compared directly. Already published works suggest that a point-wise prediction, in which one provides a prediction of the diffusion parameters at each point in time, using machine learning may be a promising solution [94, 95].

In this work we expand on previous works by training a *Bayesian Deep Learning* model to provide a point-wise prediction of not only the anomalous diffusion exponent but also its uncertainty in this prediction for a given single-particle trajectory, see figure 1 for a schematic representation. We subsequently aim to extract the change-points between the diffusive modes from the sequences of predicted anomalous diffusion exponents and their uncertainties, that exhibit a peak around the change-point. Apart from the obvious advantage from using an additional time series, we also expect this to improve change-point detection in cases when the anomalous diffusion exponent itself does not change significantly, which could lead to a failed change-point detection if only the point-predictions for α were used. However, even for such a small change in α the uncertainty will typically show a more pronounced signal and thus improve detection.

The paper is structured as follows. We begin with an explanation of the utilised anomalous diffusion data set (section 2.1), machine learning procedures (section 2.2) as well as evaluation metrics (section 2.3). Subsequently we present the obtained results in section 3—this includes the performance of the point-wise prediction (section 3.1) as well as a detailed evaluation of change-point extraction methods (section 3.2), followed by a short consideration of parameter predictions for the diffusive modes (section 3.3). We finish with a discussion of the results in section 4.

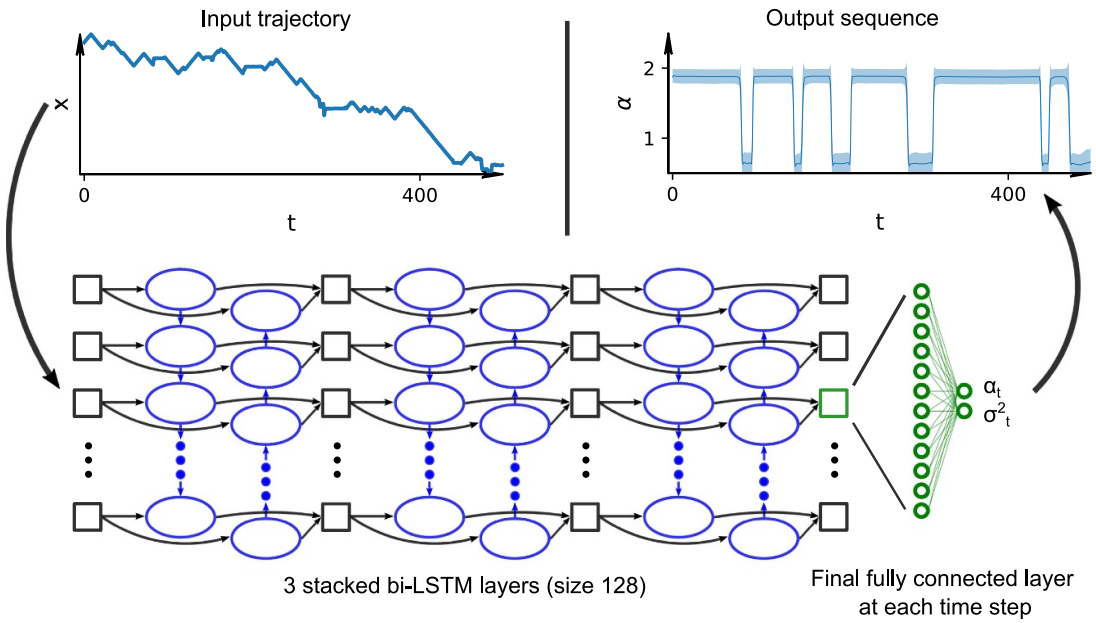

Figure 1. Visualisation of the machine learning model. The model provides a point-wise prediction of the anomalous diffusion exponent and its uncertainty given a single-particle diffusion trajectory. The output sequence in the upper right panel indicates the predicted error as a shaded area of $\pm\sigma$ around the mean. After the trajectory is converted to increments and normalised, it is fed into a stacked *bidirectional long-short-term-memory* (bi-LSTM) [96] neural network consisting of 3 layers of size 128. This network returns a time series of vectors of size 128, each of which is again passed through a single-layer fully connected network to produce the time series of estimates for the mean of the anomalous diffusion exponent α_t and its variance σ_t^2. The final predicted output time-series is obtained by combining the output of multiple such networks, sampled from a *Multi-SWAG* estimate [97].

2. Methods

2.1. Anomalous diffusion data set

The data set is generated utilising 4 different diffusion models, all yielding anomalous diffusion with $\langle x^2(t)\rangle \sim 2K_\alpha t^\alpha$. Namely, we use FBM, SBM, CTRW and Lévy Walk (LW).

FBM is the paradigmatic model for breaking the independence condition of the CLT. It is characterised by the covariance function

$$\langle x(t)\,x(t+\tau)\rangle = K_\alpha\left(|t|^\alpha + |t+\tau|^\alpha - |\tau|^\alpha\right), \tag{3}$$

which constitutes a long-time correlation of the increments in the power-law form

$$\langle \Delta x(t)\,\Delta x(t+\tau)\rangle \sim \alpha(\alpha-1)K_\alpha\tau^{\alpha-2} \tag{4}$$

for sufficiently large τ, where α is the anomalous diffusion exponent and K_α is the generalised diffusion constant [31]. This correlation entails a persistent or anti-persistent motion for $\alpha > 1$ (superdiffusion) or $\alpha < 1$ (subdiffusion) respectively.

SBM breaks the identity condition of the CLT. This is achieved by introducing a time dependent diffusivity $K(t) = \alpha K_\alpha t^{\alpha-1}$ in the Langevin equation

$$\frac{\mathrm{d}x(t)}{\mathrm{d}t} = \sqrt{2K(t)}\,\xi(t), \tag{5}$$

where $\xi(t)$ is white, zero-mean Gaussian noise [32].

In a CTRW both the jumps and the waiting times between those jumps are stochastic variables [36–38]. We here consider a CTRW with a heavy-tailed waiting time PDF $\psi(\tau)$, such that $\psi(\tau) \propto \tau^{-1-\alpha}$ for large τ with scaling exponent $0 < \alpha < 1$. The resulting diverging mean waiting time $\int_0^\infty \tau\psi(\tau)\mathrm{d}\tau = \infty$ leads to a subdiffusive behaviour. The spatial displacements are drawn from a zero-mean Gaussian PDF.

LWs can be considered as a special case of a CTRW [39, 40]. As before it features a power-law waiting time PDF $\psi(\tau) \propto \tau^{-1-\kappa}$, but, instead of only jumping after a waiting time, the walker travels with a constant speed v in one direction for the duration of one waiting time. After this time a new random direction is

chosen. As shown in [40] this leads to superdiffusion with anomalous diffusion exponent

$$\alpha = \begin{cases} 2 & \text{if } 0 < \kappa < 1 \text{ (ballistic diffusion)} \\ 3 - \kappa & \text{if } 1 < \kappa < 2 \text{ (superdiffusion)}. \end{cases} \tag{6}$$

We use the `andi-data sets` Python package for the implementation of these models [98].

We generate multiple data sets consisting of different numbers N of trajectories. Each trajectory consists of $T = 500$ data points and will randomly switch between two diffusive modes, each following one of the four diffusion models described above. For each mode, we choose the anomalous diffusion exponent α from a uniform random distribution $\alpha \in \{0.05, 0.1, \ldots, 2.0\}$, with restrictions based on the underlying diffusion model. Note that LWs may only be superdiffusive $\alpha \geqslant 1$, CTRWs only subdiffusive $\alpha \leqslant 1$, and ballistic ($\alpha = 2$) FBMs are not considered here. Thus, in each simulated trajectory we simulate sequences of two individual anomalous diffusion models, each with a given α and K_α. Each of the two diffusive modes is generated with an increment standard deviation of 1 and subsequently multiplied by a factor taken from a Gaussian distribution with mean 0 and variance 1, the square value of which is thereby equivalent to twice the (short-time) diffusion constant. As trajectories will be normalised later on, only the ratio of the diffusion constants of the two modes is of importance here. Note that the same diffusion model and/or parameters may be chosen for both modes.

To generate the change-points between the modes we draw dwell times from an exponential distribution

$$\Psi(t) = \frac{1}{\beta} \times \exp\left(-\frac{t}{\beta}\right). \tag{7}$$

The scale parameter β is itself a random variable drawn for each trajectory from a reciprocal distribution $p(\beta) \propto 1/\beta$, but limited within $50 < \beta < 250$. This choice corresponds to the non-informative prior (Jeffrey's prior) for the exponential distribution [99, 100], the limits were chosen to generate trajectories that feature a few but not too many change-points within the total trajectory length $T = 500$.

All data are corrupted by white Gaussian noise with a signal to noise strength ratio randomly chosen from $\text{snr} \in \{1, 2, 10\}$. Given the trajectory $x(t)$, we obtain the noisy trajectory $\tilde{x}(t) = x(t) + \xi(t)$ with the superimposed noise

$$\xi(t) \sim \frac{\sigma_{\Delta x}}{\text{snr}} \mathcal{N}(0, 1), \tag{8}$$

where $\sigma_{\Delta x}$ is the standard deviation of the increment process $\Delta x(t) = x(t+1) - x(t)$. Note that if the diffusion constants of the two modes differ strongly, one will be much more affected by noise than the other, as would also be the case in real tracking experiments.

We generate a total of 4 data sets. A training data set consisting of $N = 5 \times 10^6$ trajectories used to train the neural network (NN), a validation data set of $N = 1 \times 10^5$ for fine-tuning of the machine learning hyper-parameters, as well as two test data sets consisting of $N = 5 \times 10^5$ trajectories each. The first is utilised to fine-tune the methods for change-point extraction, which are then applied to the second test data set to generate the results herein.

To allow for general applicability independent of scale as well as in order to stabilise the training process, trajectories undergo a minimal amount of preprocessing. Concretely, each trajectory is converted into its increment series and normalised to a uniform standard deviation before being fed into the NN.

2.2. Machine-learning procedure

In recent years many NN architectures for the analysis of anomalous diffusion trajectories have been proposed [66, 81]. These range from fully connected [67] to convolutional [68–71], recurrent [72–76] or even graph NNs [77]. After some initial probing we here decided on a recurrent NN as depicted in figure 1. The network consists of three stacked bi-LSTM [96] layers of size 128, which return a time series of 128-dimensional vectors. At each time step t, these are passed through a fully connected layer, outputting estimates for mean α_t and variance σ_t^2 of the anomalous diffusion exponent.

We choose the recurrent LSTM-architecture due to its design for time-series prediction as well as its immediately accessible output of a time-series of the same length as the input. Additionally, this architecture has been successfully applied to anomalous diffusion trajectories before [65, 74]. We added bidirectionality as the point-wise predictions pursued here should heavily depend on future data points, which is not the case in a unidirectional LSTM. The three layers of size 128 were chosen via trial-and-error, testing larger networks until no further improvement was perceived. Note that we also tested a multi-headed attention architecture [101] but found that this did not improve results noticeably for our setup, especially considering its significantly increased computational cost.

To find the optimal weights of the NN we minimise a loss function on the training data set. As the outputs include uncertainty estimates, the network cannot be trained by using an—otherwise standard for regression tasks—squared error loss. Instead we minimise the *Gaussian negative log-likelihood loss*

$$\mathcal{L}_{\text{gnll}} = \sum_{n,t} \frac{1}{2} \left(\log\left(\sigma_{n,t}^2\right) + \frac{|\alpha_{\text{pred},n,t} - \alpha_{\text{true},n,t}|^2}{\sigma_{n,t}^2} \right), \tag{9}$$

where $\alpha_{\text{pred},n,t}$ and $\sigma_{n,t}^2$ are the mean and variance outputs of the NN for the nth input trajectory at time t [102].

We train the network for 90 epochs, where each epoch constitutes a full parse through the training data set. For each epoch the training data set is randomly divided into batches of size 480, updating the network weights after each batch is passed through the network. Weight updates are typically achieved using some variant of a stochastic-gradient-descent algorithm [103]. We here use *Adam*-descent [104] with a cyclic learning rate between 1×10^{-3}–2×10^{-5}. During the last 10 training epochs, we build a so-called *SWAG*-estimate [105] of the neural-network weights. These assign a Gaussian probability distribution to the NN weights by tracking their change over the training process. The algorithm was implemented by means of *pytorch* [106] using the *NVIDIA T4* GPU with *CUDA 12.0*. A full training run took roughly 50–60 h. The run time depends on the load of the cluster, which consisted of four such GPUs and two *AMD EPYC 7453* CPUs, allowing for up to four simultaneous runs.

This training process was repeated multiple times resulting in a total of 6 such *SWAG*-estimates, which constitute a *Multi-SWAG* estimate, as proposed in [97]. The final predictions on a test data set are then obtained by sampling 10 networks from each *SWAG*-estimate for a total of 60 Monte-Carlo-samples, which are then combined into a single prediction. Note that the variance between Monte-Carlo-samples is used as an additional, so-called *epistemic*, contribution to the uncertainty estimate [107, 108]. For details on the *Multi-SWAG* procedure see [97, 105] as well as [65, 66, 78] for example applications to anomalous diffusion. A flowchart of the described procedure is depicted in figure 2.

2.3. Evaluation metrics

To evaluate the performance of the network on the prediction of the full time series of α, we define the *mean absolute error*

$$\texttt{MAE}_\alpha = \frac{1}{N} \sum_{n=1}^{N} \frac{1}{T} \sum_{t=0}^{T-1} |\alpha_{\text{pred},n,t} - \alpha_{\text{true},n,t}|, \tag{10}$$

as well as the *root mean squared error*

$$\texttt{RMSE}_\alpha = \sqrt{\frac{1}{N} \sum_{n=1}^{N} \frac{1}{T} \sum_{t=0}^{T-1} \left(\alpha_{\text{pred},n,t} - \alpha_{\text{true},n,t}\right)^2}. \tag{11}$$

To visualise the performance of the error prediction we use reliability diagrams [109, 110], which illustrate the observed errors as a function of the predicted uncertainties. To achieve this performance quantification, we sort the outputs of the NN into sets based on a binning of the predicted standard deviation into intervals $I_m = \left((m-1)\Delta_\sigma, m\Delta_\sigma\right]$. We define B_m as the set of output indices with a predicted standard deviation within the interval I_m,

$$B_m = \left\{(i,\tau) | \sigma_{i,\tau} \in I_m = \left((m-1)\Delta_\sigma, m\Delta_\sigma\right]\right\}, \tag{12}$$

where $\sigma_{i,\tau}$ is the output standard deviation for trajectory i at time index τ. For each interval I_m we calculate the (observed) *root mean squared error*

$$\texttt{RMSE}_\alpha\left(B_m\right) = \sqrt{\frac{1}{|B_m|} \sum_{(i,\tau)\in B_m} \left(\alpha_{\text{pred},i,\tau} - \alpha_{\text{true},i,\tau}\right)^2} \tag{13}$$

and compare it to the (predicted) *root mean variance*

$$\texttt{RMV}_\alpha\left(B_m\right) = \sqrt{\frac{1}{|B_m|} \sum_{(i,\tau)\in B_m} \left(\sigma_{i,\tau}\right)^2}, \tag{14}$$

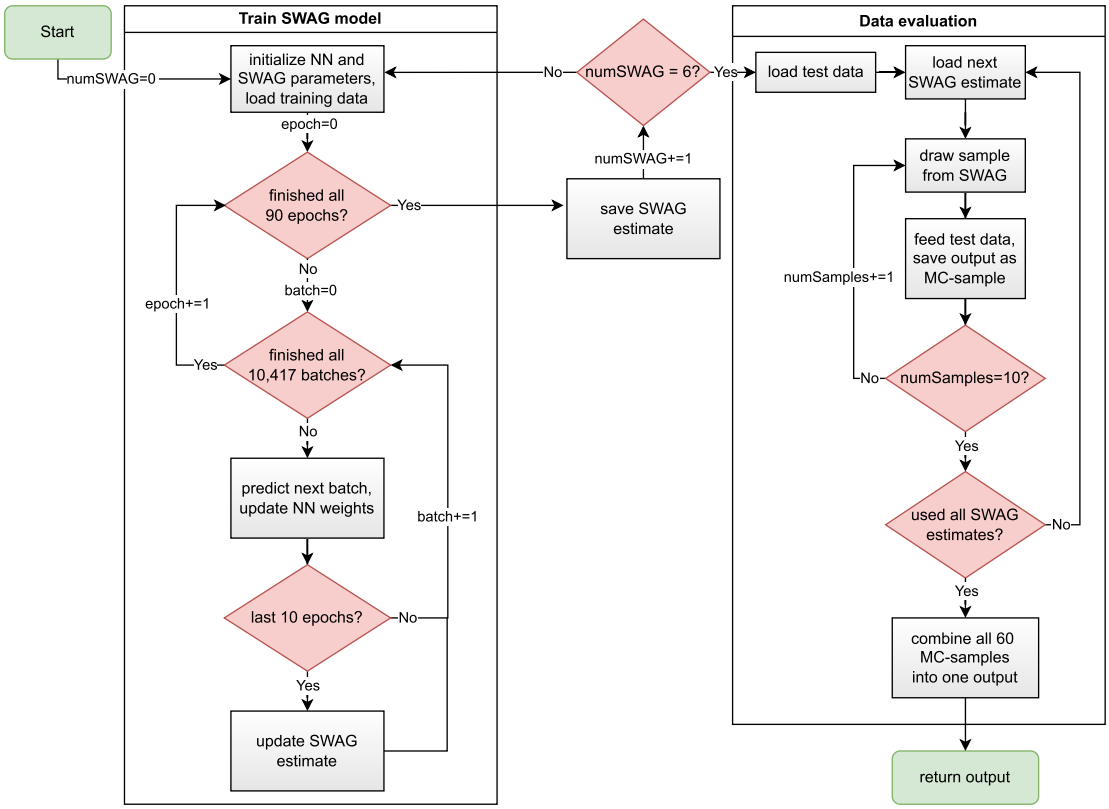

Figure 2. Flowchart of the *multi-SWAG* procedure used for training and evaluation. The *multi-SWAG* [97] model consist of 6 *SWAG* models, which estimate a Gaussian posterior on the neural network (NN) weights during the last 10 epochs of the training loop. During data evaluation we sample 10 networks from each *SWAG* estimate for a total of 60 Monte-Carlo-samples (MC-samples). These samples are combined into a single output prediction of means and uncertainties. Note that the final predicted uncertainties are the sum of the mean predicted variances and the variances of predicted means [65, 108].

where $|B_m|$ is the cardinality of B_m. Plotting RMSE in dependence of RMV for all m constitutes the reliability diagram. Coinciding RMSE and RMV in all bins represent a perfectly calibrated model. Deviations from the ideal error prediction can be summarised by the *expected normalised calibration error* (ENCE) [111]

$$\mathrm{ENCE} = \sum_{B_m} \frac{|B_m|}{NT} \frac{|\mathrm{RMV}(B_m) - \mathrm{RMSE}(B_m)|}{\mathrm{RMV}(B_m)}. \tag{15}$$

For change-point detection we will be given a set of predicted and true change-points for each trajectory. We define the gated absolute distance between the ith true change-point at time $t_{\mathrm{cp,true},i}$ and the jth predicted change-point at $t_{\mathrm{cp,pred},j}$ as [112]

$$d_{i,j} = \min\left(|t_{\mathrm{cp,true},i} - t_{\mathrm{cp,pred},j}|, d_{\max}\right), \tag{16}$$

where $d_{\max}$ is the maximum penalty for change-points too far apart. The choice of this threshold will significantly impact the performance metric. Here we choose $d_{\max} = 10$, aligning with the choice made in other works on this subject [81, 93–95] and thereby allowing for better comparability. To find the optimal pairing of true and predicted change-points, we need to minimise the sum of distances

$$d_{\mathrm{tot}} = \sum_{\text{pairs } i,j} d_{i,j}. \tag{17}$$

Such an assignment problem can easily be solved via, for example, the *Hungarian Algorithm* [113]. Utilising this optimal assignment we count the number of true positives (TP) as those pairs with a distance less than $d_{\max}$. Predicted change-points without any associated true change-point are counted as false positives (FP), while true change-points without any paired predicted change-point are false negatives (FN). Paired change-points with distance larger than $d_{\max}$ contribute to the count of both FP predicted and FN true change-points. We calculate the total number of TP, FP and FN over all N trajectories in the test data set.

From these we can calculate the *Jaccard Similarity Coefficient* [114]

$$J = \frac{\mathtt{TP}}{\mathtt{TP} + \mathtt{FP} + \mathtt{FN}}, \tag{18}$$

which corresponds to the intersection of the true and predicted set of change-points divided by their union. Additionally one may compute precision and recall as

$$\mathtt{PREC} = \frac{\mathtt{TP}}{\mathtt{TP} + \mathtt{FP}} \tag{19}$$

$$\mathtt{REC} = \frac{\mathtt{TP}}{\mathtt{TP} + \mathtt{FN}}. \tag{20}$$

Here the first indicates how many detected change-points were correct, while the latter measures the share of true change-points that were detected. The harmonic mean of precision and recall is called the F_1 score,

$$F_1 = \frac{2}{\mathtt{PREC}^{-1} + \mathtt{REC}^{-1}}. \tag{21}$$

Finally, for those predictions that were detected as TP, we also calculate the root mean squared error

$$\mathtt{RMSE}_{\mathrm{CP}} = \sqrt{\frac{1}{\mathtt{TP}} \sum_{\substack{\text{pairs } i,j \\ d_{i,j} < d_{\max}}} \left(t_{\mathrm{cp,true},i} - t_{\mathrm{cp,pred},j}\right)^2}. \tag{22}$$

3. Results

After training the *Multi-SWAG* model, we here report the obtained results. First we examine the performance of the network on the sequence prediction of the anomalous diffusion exponent and its uncertainty as a whole in section 3.1. This is followed by the change-points extraction in section 3.2, where we first analyse the performance when extracting change-points from only the series of anomalous diffusion exponent values *or* the predicted variances, and later we explore the possibility to improve the results by a combined change-point determination from both. This will include a comparison of different variants of extraction methods. Finally, we use the sequence-predictions, now segmented via the change-points, to determine single-valued predictions for the two motion modes and report on the obtained results in section 3.3.

Note that we use two test data sets for this purpose, both containing $N = 5 \times 10^5$ trajectories. The first is used to determine and fine tune the best method to extract change-points, which is then applied to a second test data set in order to avoid any possible bias.

3.1. Sequence performance

We start by examining the network performance for the sequence prediction as a whole. Figure 3 shows an example of an anomalous-diffusion trajectory and the predictions of anomalous diffusion exponent α and uncertainty σ^2 at each point in time. The panels on the right show the changes in each of these quantities. The figure shows that the anomalous diffusion exponent switches between two predictions with sharp changes at the change-points. Similarly the predicted variance appears to switch between two predictions, we here however see a peak-like behaviour at the change-points, which is mirrored in the change in the uncertainty (on the right of panel (c)) as a double-peak. This is caused by the uncertainty in the change-point location, which requires the predicted variance to account for the possibility of still being in either of the two modes. As such, the presence and severity of this peak will depend on the difference in α between the two motion modes in comparison to the base uncertainty levels in each mode. Nevertheless, the series of uncertainties may help detect change-points, where the predicted exponent α changes little while the variance σ^2 does.

In general, when evaluating the sequence performance on the second test data set, we were able to achieve a mean absolute error of $\mathtt{MAE}_\alpha = 0.178$ and a root mean squared error of $\mathtt{RMSE}_\alpha = 0.259$. The magnitude of this error fits well within a previous study, in which the trajectories consisted of only one motion mode, where we were able to achieve a mean absolute error between 0.21–0.12 for trajectories consisting of 100–500 data points [65]. The large difference between MAE and RMSE is due to the increased importance placed on larger deviations in the RMSE. The Gaussian negative log likelihood loss, on which the model was trained, is $\mathcal{L}_{\mathrm{gnll}} = -1.13$. As this number is hard to interpret, we rely on a more explicit measure, gained by examining the predicted root mean variance. Figure 4 shows the observed RMSE as a function of the predicted RMV in a reliability diagram. This is achieved by introducing a binning of the RMV as described in equations (14) and (13). In the figure, we observe a slightly overestimated error for high predicted variances. This is likely

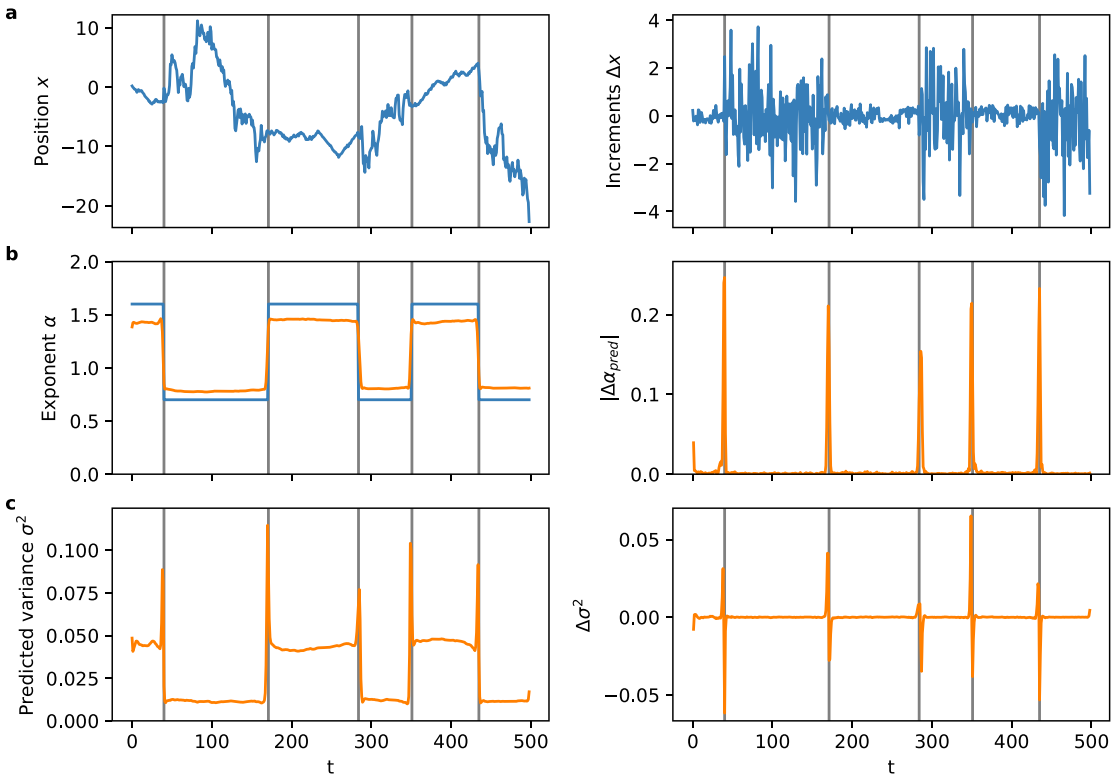

Figure 3. Sample trajectory (a) and prediction of the anomalous diffusion exponent (b) and uncertainty (c) given by the neural network. In this sample the trajectory changes between an LW with $\alpha = 1.6$ and an SBM with $\alpha = 0.7$. The changes were generated with mean dwell time $\beta \approx 184$. The panels on the right show the changes for each of the time series (a)–(c). In (b) the blue line corresponds to the true and the orange line to the predicted anomalous diffusion exponent. The change-points, indicated by gray lines, appear to coincide with maximum changes in the predicted exponent α or uncertainty σ^2. Note that the change in uncertainty in (c) features a double-peak-like behaviour; this is caused by a lack of confidence in the position of the change-point leading to a high predicted uncertainty. As such the presence and severity of the double peak mainly depends on the corresponding change in α compared to the base uncertainty levels outside of the change-point region.

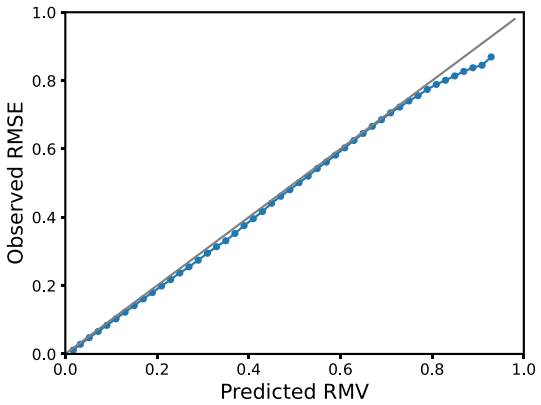

Figure 4. Reliability diagram for predictions of the whole sequence. The diagram shows the observed root mean squared error (RMSE) for given predicted root mean variance (RMV). Ideally predicted and observed errors would coincide as indicated by the grey line. As we see here the network slightly overestimates the error, this is especially prevalent for the upper end of the predicted RMV and likely caused by the high errors typically assigned to the change-point regions.

caused by the high error predictions associated with the change-point regions. The mean deviation between true and predicted errors, as defined by the expected normalised calibration error in equation (15), was determined as $\mathrm{ENCE} = 5.3\%$. This slight overestimation of the uncertainty is also mirrored when comparing the mean predicted variance on all data samples without binning, which yields $\mathrm{RMV}_\alpha = 0.269$, to the observed $\mathrm{RMSE}_\alpha = 0.259$.

Algorithm 1. Change-point extraction from $\alpha(t)$.

Input: series of predicted anomalous diffusion exponents $\alpha_0, \dots \alpha_{T-1}$, relative threshold r_α, minimum threshold $r_{\min,\alpha}$, minimum distance between change-point regions $l_{\min}$

Output: vector of predicted change-points $t_{\mathrm{cp},0}, \dots, t_{\mathrm{cp},K-2}$

1: Let $\Delta\alpha_0, \dots, \Delta\alpha_{T-2}$ be the series of changes in α:
$\Delta\alpha_t = \alpha_{t+1} - \alpha_t$.

2: Let Δ_{thres} be the threshold:
$\Delta_{\mathrm{thres}} = \max(r_\alpha \times \max_t(|\Delta\alpha_t|), r_{\min,\alpha})$.

3: Let $t_{\mathrm{exc},0}, \dots, t_{\mathrm{exc},S-1}$ be the time-indices of $\Delta\alpha_t$ exceeding the threshold: $|\Delta\alpha_{t_{\mathrm{exc},s}}| \geqslant \Delta_{\mathrm{thres}}$. Additionally add $t_{\mathrm{exc},0} = 0$ and $t_{\mathrm{exc},S-1} = T-1$ if not already included, as otherwise the first/last change-point will be missed.

4: Let $s_0, \dots, s_{K-1}$ be the indices of $t_{\mathrm{exc},s}$ with minimum distance to the previous: $t_{\mathrm{exc},s_k} - t_{\mathrm{exc},s_k-1} \geqslant l_{\min}$, thereby marking the beginning of change-point regions.

5: Initialise the set of change-points $t_{\mathrm{cp},0}, \dots, t_{\mathrm{cp},K-2}$.

6: **for** $k = 0$ to $K-2$ **do**

7: $\quad t_{\mathrm{cp},k} = \mathrm{argmax}_{t, t_{\mathrm{exc},s_k} \leqslant t < t_{\mathrm{exc},s_{k+1}-1}}(|\Delta\alpha_t|)$

8: **return** $t_{\mathrm{cp},0}, \dots, t_{\mathrm{cp},K-2}$

3.2. Change-point extraction

3.2.1. Extracted from $\alpha(t)$ and $\sigma^2(t)$ independently

In the previous section, we determined that the change-points can be extracted from the sequence predictions by detecting peaks in the changes of α or σ^2. As an example, the procedure used for change-point extraction from α is shown in algorithm 1. In short, we start by finding all changes that exceed a certain threshold, which will result in regions of data points surpassing this threshold. We consider regions as separate if there is a minimum number of non-exceeding data points between them. Here we found a minimum of $l_{\min} = 6$ data points between change-point regions to yield the best results. From each exceeding region we extract a single change-point. For α this is the point of the maximum (absolute) change in the region. To accommodate the often-occurring double-peak behaviour in the uncertainty sequence σ^2, we differentiate between two cases. If there is a positive and negative peak exceeding the threshold in a region, we take the midpoint between the two as change-point, otherwise we proceed as for α.

The choice of the threshold heavily depends on the desired metric. A high threshold will usually result in a high precision but at the cost of recall (see equations (19) and (20)), while a lower threshold will increase recall at the cost of precision. We chose a threshold such that the best Jaccard similarity (see equation (18)) is achieved. Figure 5 shows the change in the Jaccard index depending on the chosen thresholds. Here we find best results when choosing a threshold relative to the maximum (absolute) change in the trajectory $r_{\alpha/\sigma}$, which is however limited by a minimum cutoff threshold $r_{\min,\alpha/\sigma^2}$. The optimal thresholds, with respect to the Jaccard index, were determined as $r_{\min,\alpha} = 0.015, r_\alpha = 24\%$ and $r_{\min,\sigma^2} = 0.0025, r_{\sigma^2} = 28\%$. Note that we also tried using the sequence of predicted standard deviations σ instead of variances σ^2, but found this to diminish performance.

We then apply these thresholds, determined utilising the first test data set, to extract change-points on the second test data set. For the α-sequence we thus achieved $J = 58.4\%$, $\mathtt{RMSE}_{\mathrm{CP}} = 2.57$, $\mathtt{PREC} = 81.1\%$, $\mathtt{REC} = 67.7\%$, while for the uncertainty sequence we obtained $J = 56.6\%$, $\mathtt{RMSE}_{\mathrm{CP}} = 2.63$, $\mathtt{PREC} = 75.2\%$ and $\mathtt{REC} = 69.6\%$. While extracting change-points from the sequence of means α does achieve higher similarity than from the variances σ^2, we do see an improved recall when using the variance sequence. This indicates that while more of the change-points are found, this does however appear to come with a significant cost of predicted non-existent change-points, reducing precision and ultimately resulting in a lower Jaccard index J.

3.2.2. Combined performance

In this section we determine whether a combined change-point prediction from both series can improve the performance. To achieve this we tested a multitude of methods. We found that approaches to extract change-points from a combined series, for instance via defining exceeding data points as those for which the change in α and/or the change in σ^2 exceed a threshold, were inferior to an approach for which one first extracts change-points from the two series individually and then combines the two change-point sets. The combination of the two change-point sets is achieved using the Hungarian algorithm with gated absolute distance as in equation (16), equivalent to the one used for the calculation for the Jaccard index. Paired change-points with distance less than the maximum are combined into their mean, while for others both are added into the combined set of predicted change-points.

Note that we cannot use the same thresholds as in the previous section, as this results in a Jaccard index between those obtained for the two individual series. Instead we again need to determine optimal thresholds, this time by varying all four $(r_\alpha, r_{\sigma^2}, r_{\min,\alpha}, r_{\min,\sigma^2})$ at the same time. We show a projection of this parameter

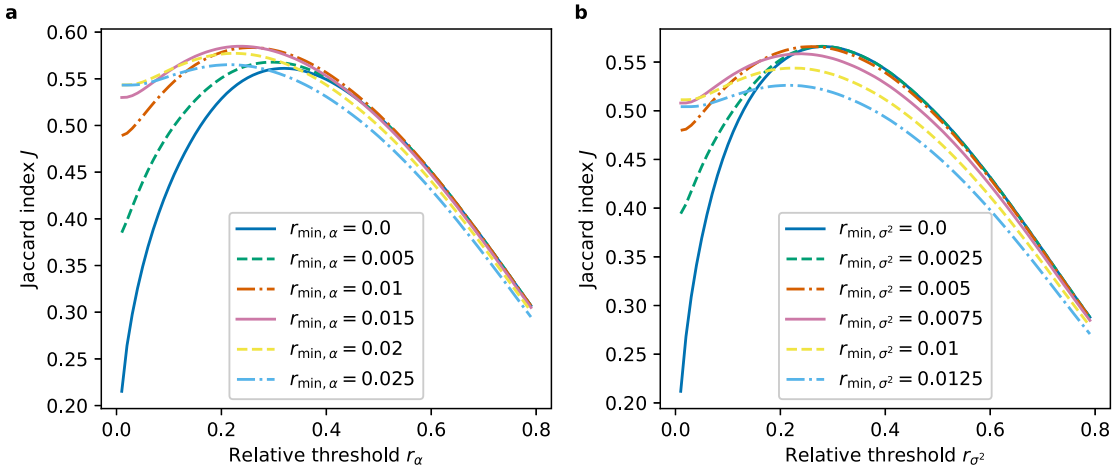

Figure 5. Individual change-point performance as characterised by the Jaccard Index in dependence of relative and minimum thresholds. Change-points are extracted from only the series of anomalous diffusion exponents α (a) or the series of predicted uncertainties σ^2 (b). Change-points are determined via peak-detection in the increment series, the optimal thresholds for this detection may be inferred here. As is evident from the figure, the best performance is achieved when using both a threshold relative to the maximum observed change as well as a minimum cutoff threshold (r_{α/σ^2} and $r_{\min,\alpha/\sigma^2}$). For α we find the best score of $J = 0.585$ at a minimum threshold of $r_{\min,\alpha} = 0.015$ and relative threshold $r_\alpha = 0.24$. For the uncertainty the change-point detection appears to be more difficult, here we find a best score of $J = 0.566$ at minimum threshold $r_{\min,\sigma^2} = 0.0025$ and relative threshold $r_{\sigma^2} = 0.28$.

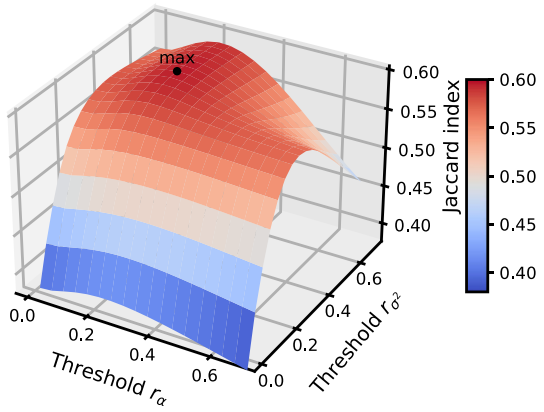

Figure 6. Combined change-point performance as characterised by the Jaccard index for different relative thresholds. Using the test data set we extracted change-points for different values of thresholds for both the series of anomalous diffusion exponents α and the series of predicted errors σ^2. The sets of change-points were subsequently combined into a single set, which was used for the calculation of the Jaccard index. Note that we here only show a projection from the true parameter space with set minimum thresholds of $r_{\min,\alpha} = 0.02$ and $r_{\min,\sigma^2} = 0.0015$. The best performance of $J = 0.595$, indicated by the black dot in the figure, was found for relative thresholds at $r_\alpha, r_{\sigma^2} = 0.21, 0.45$ (and $r_{\min,\alpha}, r_{\min,\sigma^2} = 0.02, 0.0015$).

space in figure 6. The maximum Jaccard similarity is achieved when using thresholds similar to the previous ones for the α-series ($r_\alpha = 21\%, r_{\min,\alpha} = 0.02$), while using a high relative threshold for the series of uncertainties σ^2 ($r_{\sigma^2} = 45\%, r_{\min,\sigma^2} = 0.0015$), indicating that only the more certain change-points of the latter are used to supplement the—mostly similar to the independently extracted—set of change-points from the former.

Applying the same thresholds to the second test data set, we were able to improve the Jaccard index to $J = 59.5\%$, 1.1 and 2.9 percentage points higher than the ones obtained for the individual series of anomalous diffusion exponent and uncertainty respectively. Additionally we find $\mathtt{RMSE}_{\mathrm{CP}} = 2.56$, $\mathtt{PREC} = 78.8\%$ and $\mathtt{REC} = 70.8\%$. Similar to the difference between change-points from only the α- and σ^2-series, we observe an increase in recall at the cost of a decrease in precision—here however the decrease in precision is significantly lower and the increase in recall higher, ultimately leading to an increased Jaccard similarity. A similar effect may also be found in the harmonic mean of precision and recall (F_1 score), which increases to $F_1 = 74.5\%$ for the combined prediction compared to $F_1 = 73.8\%$ for change-points from the α-series ($F_1 = 72.3\%$ for σ^2-series). While the increase in J or F_1 is small, we see a substantial increase in recall by 3.1 percentage

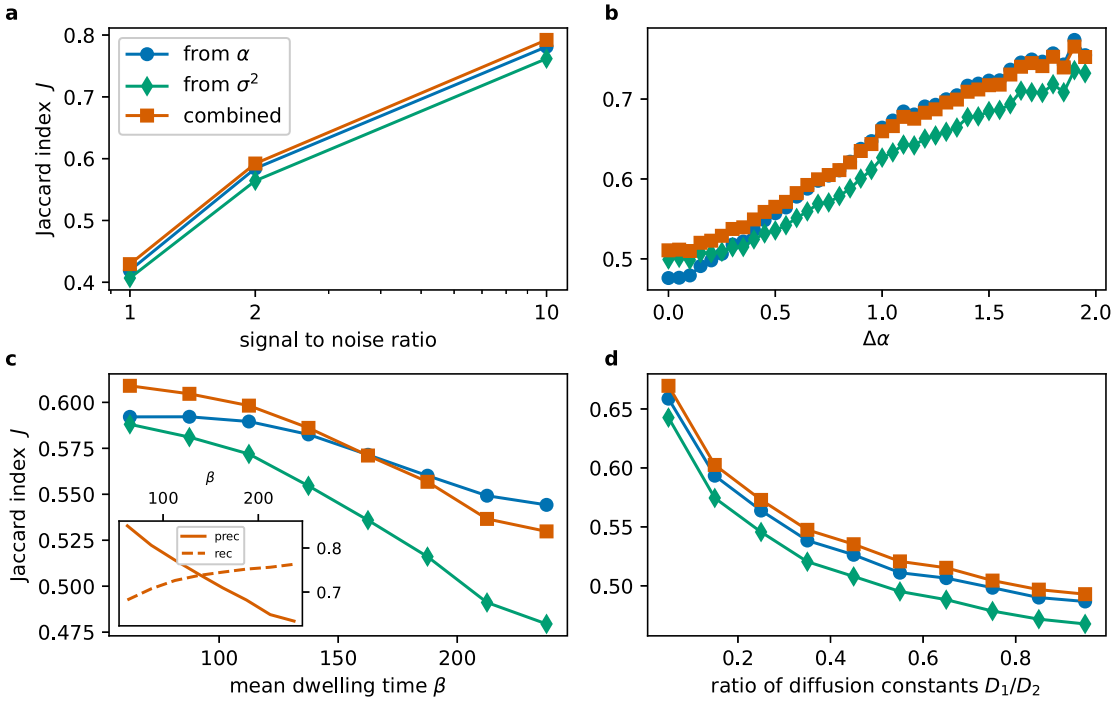

Figure 7. Performance as characterised by the Jaccard index in dependence of the data set generating parameters, consisting of (a) the signal to noise strength ratios, (b) the absolute difference in anomalous diffusion exponent $\Delta\alpha$ between the two diffusive modes, (c) the parameter β (equal to mean and inverse of rate) used for the generation of dwell times (see equation (7)) and (d) the ratio of diffusion constants between the two modes. The figure shows the performance for extracted change-points from the predicted series of anomalous diffusion exponents (blue circles), variances (green diamonds) and the combined approach (red squares). The inset in panel (c) shows precision (solid line) and recall (dashed line) for the combined approach. The change-points extracted from the α-series consistently outperform the ones extracted from σ^2-series, with the exception of small $\Delta\alpha$ in (b). In most cases the combined approach improves on both other variants, with the exception of high dwell times between change-points in (c).

points as compared to the extraction from only α. This can be significant, as often ensuring that change-points are found (recall) is more import than not detecting false change-points (precision).

Figure 7 shows the obtained Jaccard score for all three extraction variants (α, σ^2 and combined) in dependence of the parameters, that were randomly chosen for the data set generation. When varying the noise strength in panel (a) the score of the combined approach shows a strong drop to $J = 59\%, 42\%$ for high noise strengths of snr $= 2, 1$, with a significantly higher score of $J = 79\%$ for the (typically more realistic) snr $= 10$. The drop is similar, regardless of which series was used, though the combined approach outperforms the change-points extracted from α and σ^2 individually.

In panel (b) we see the only point where the series of predicted variances can outperform the change-points extracted from the series of predicted means of α, which occurs when the change in α between the two motion modes is small. This is the area where we also expected the inclusion of uncertainties to yield improvements. In the combined approach this improvement is even stronger, increasing J by 3.5 percentage points when $\Delta\alpha = 0$ and also persisting for longer, up to $\Delta\alpha = 0.5$, closely following the score of the α-series afterwards.

Interestingly in panel (c), wee see that the score generally drops with lower change-rates (higher mean dwell time β). This is likely caused by the high cost in accuracy that occurs when a single wrong change-point is predicted for low change-rates, which will have a significantly less impact for high rates. This can be further understood by examining the inset of panel (c), which shows a decreasing precision (ratio of retrieved change-points that were correct) and an increasing recall (ratio of true change-points that were retrieved) for the combined extraction. The increase in J for low change-rates (high β) is especially severe for the series of uncertainties, which also appears to drag the performance of the combined approach below the the one using only the α-series for high β. As this is the only case in which the combined approach shows a significantly reduced performance, it might indicate a possible improvement avenue for future research.

The final panel (d) reveals a to-be-expected decreased performance when the two diffusion coefficients of the diffusive modes are closer to each other, dropping from $J = 67\%$ to $J = 49\%$ in the combined case. Here again we see the typical pattern of the combined approach outperforming the change-points extracted from the α-series, which again outperforms the σ^2-series.

As a point of reference, during the second *AnDi-Challenge* the top participants reached similarity scores from $J = 65\%$ up to $J = 70\%$ (and errors from $\texttt{RMSE}_{\text{CP}} = 1.63$ up to $\texttt{RMSE}_{\text{CP}} = 1.7$) when identifying change-points in a more phenomenological trajectory setup [93]. Requena *et al* were able to obtain a similarity of $J = 51.5\%$ when detecting change-points using the anomalous diffusion exponent in FBM trajectories of length 200 [94]. In a similar setup, Qu *et al* achieved an F_1 score of $F_1 = 88\%$ for FBM trajectories of length 500, here however all diffusive modes featured the same diffusion coefficient $K_\alpha = 1$ and noise standard deviation $\sigma_n = 0.125$ and were limited to a maximum of five change-points per trajectory [95]. While none of these can be compared directly to the results obtained in this work, they may be used as an indication of the magnitude to be expected for J or F_1.

3.3. Parameter prediction for the two modes

We conclude by extracting the parameter values of the two motion modes and evaluating the results. Note that simply training a NN to directly output single-valued predictions for the two modes is likely to yield better results, such a method however would not allow for change-point extraction and would not be easily transferable to an arbitrary number of motion modes. Instead we will here extract single-valued predictions $(\alpha_1, \sigma_1^2, \alpha_2, \sigma_2^2)$ for the two modes from the predicted time series of $\alpha(t)$ and $\sigma^2(t)$.

Each trajectory was divided into sequences based on the predicted change-points, and the means of α and σ^2 for each sequence were calculated. Subsequently these values were sorted into two sets based on proximity, and the means of each were used as final predictions of $\alpha_{1/2}$ and $\sigma^2_{1/2}$ for the first and second model. This results in an $\texttt{MAE}_\alpha = 0.190$ and $\texttt{RMSE}_\alpha = 0.268$, slightly worse than the values obtained for the whole sequence in section 3.1. This is not surprising, as the calculations of the error here place equal importance on both motion modes, whereas for the whole sequence the mode maintained for a longer time (and thus easier to predict) has a larger impact on the error. Likely due to the additional uncertainty connected to the position of the change, the predicted error extracted in this manner slightly overestimates the true observed error, resulting in a predicted $\texttt{RMV} = 0.298$ and mirrored in the calibration error of $\texttt{ENCE} = 7.4\%$.

4. Discussion

Recent years have seen an upsurge in machine learning solutions for the analysis of anomalous diffusion [65–81, 115, 116]. The potential of these methods was displayed during the *AnDi-Challenge* [80, 81], which already featured a segmentation task. This task however remained with a simple problem of a single change-point and did not see a very high participation. Since then applications of machine learning to segmentation tasks in diffusion trajectories have developed further.

While in some cases change-points may be determined without resorting to machine learning [88–91], this can prove difficult for more subtle changes. In such cases, recent works indicate that a point-wise prediction of anomalous diffusion parameters using machine learning may be a promising solution [94, 95]. Detecting and specifying change-points of diffusion trajectories is also the topic of the just-completed *2nd AnDi-Challenge* [93].

In a previous paper [65] we discussed the potential of an added uncertainty prediction when determining parameters of a single-particle anomalous-diffusion trajectory using a NN. This is expanded upon in the current work, where we explored the applicability of these uncertainty outputs for change-point detection. For this purpose, we generated data sets of trajectories switching between two diffusive modes. Through *Bayesian deep learning* [97, 105], we were able to obtain a point-wise prediction not only of the anomalous diffusion exponent α but also of the corresponding uncertainty, as measured by the variance σ^2. In figure 3 we saw that this variance shows an interesting behaviour at the change-points, often resulting in distinct spikes at the change-point position. We hypothesised, that the uncertainty series may help detect change-points where the predicted exponent changes little, as those would be hidden in the α-series.

While in the present study we found that extracting the change-points from these uncertainties does indeed lead to more true positive change-point determinations, improving recall, it struggles in precision, adding many false positive change-points as well. Ultimately change-points extracted from only the uncertainty-series performed worse than those extracted from the series of anomalous diffusion exponents, when measured by the Jaccard index. However, when utilising a combined change-point extraction from both series we were able to improve the results, significantly increasing recall with a smaller loss in precision, which resulted in an improvement of the Jaccard index from $J = 58.4\%$ to $J = 59.5\%$. Additionally we showed that the parameters of the two motion modes can be recovered from the predicted sequence and change-points, albeit with a small loss in accuracy.

An extension of the discussed methods to trajectories switching between more than two modes may be of interest for the future, though the methods discussed here were intentionally chosen to be easily transferable to an arbitrary number of modes. Similarly an application to different anomalous diffusion models, like, for

instance, the more phenomenological models used in the *2nd AnDi-Challenge* [93], could yield notable results. The challenge considered predictions of not only the anomalous diffusion exponent but also the diffusion coefficient and model. Therefore examining the impact of uncertainty predictions for change-point detection in such cases, when more than one parameter is to be predicted, may be of interest. Similarly the use of uncertainty predictions in a classification task could be considered. Here however the difference to standard classification approaches would be small, as these already include class probabilities, even if those are often not very well calibrated due to overfitting [105, 109, 117]. Another challenging point to include is dynamical error [66, 118].

The simple peak-detection method used here for change-point detection could also be reconsidered. A more involved method like, for example, Bayesian inference may be able to achieve better results. While Bayesian inference has already been applied directly to diffusion trajectories [62–64, 81] and an application to change-point detection is being worked on [119], it is known to struggle when competing with machine learning approaches due to its high computational cost. An application to the point-wise predictions obtained from a machine learning algorithm however should incur a significantly smaller cost.

Finally one may also wish to test the limitations of the algorithms discussed here. Previous works have shown that a blind application of machine learning will often lead to erroneous predictions [66]. Observing if change-points could still be extracted, even if the predicted values in the sequence are sub-optimal, will be interesting.

Such a consideration would also shed light on the to-be-expected behaviour for experimental data. The prediction of change-points can there be utilised to segment trajectories showing changing behaviour, such as observed, for example, in the movement patterns of animals during foraging, a cell during feeding, as an effective behaviour in an heterogeneous environment or due to transient interactions with other objects [46, 82–87, 120–122]. A properly adjusted algorithm may even be applied to find change-points in financial time series, which often show changing behaviour due to corresponding real world decisions [48, 49]. Once segmented the predictions of the machine can be easily double-checked using the more traditional methods, such as the mean squared displacement, which can otherwise be hard to apply to heterogeneous trajectories.

Data availability statement

No new data were created or analysed in this study.

Acknowledgments

We thank the German Ministry for Education and Research (NSF-BMBF Project STAXS) and the German Research Foundation (DFG, Grant No. ME 1535/12-1).

ORCID iDs

Henrik Seckler https://orcid.org/0000-0002-7729-433X
Ralf Metzler https://orcid.org/0000-0002-6013-7020

References

[1] Carus T L 1975 *De Rerum Natura (On the Nature of Things)* (Harvard University Press) translated by Rouse W H D, revised by Smith M F
[2] Brown R 1828 XXVII. A 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 *Phil. Mag.* **4** 161
[3] Einstein A 1905 Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen *Ann. Phys., Lpz.* **4** 549
[4] Pearson K 1905 The problem of the random walk *Nature* **72** 294
[5] Weiss G H 1994 *Aspects and Applications of the Random Walk* (North-Holland)
[6] Mises R V 1919 Fundamentalsätze der Wahrscheinlichkeitsrechnung *Math. Z.* **4** 1
[7] von Smoluchowski M 1906 Zur kinetischen Theorie der Brownschen Molekularbewegung und der Suspensionen *Ann. Phys., Lpz.* **326** 756
[8] Sutherland W 1905 LXXV. A dynamical theory of diffusion for non-electrolytes and the molecular mass of albumin *Phil. Mag.* **9** 781
[9] Langevin P 1908 Sur la théorie du mouvement Brownien *C. R. Acad. Sci., Paris* **146** 530
[10] van Kampen N G 1981 *Stochastic Processes in Chemistry and Physics* (North Holland)
[11] Hughes B D 1995 *Random Walks and Random Environments* vol I (Oxford University Press)
[12] Bouchaud J-P and Georges A 1990 Anomalous diffusion in disordered media: statistical mechanisms, models and physical applications *Phys. Rep.* **195** 127
[13] Metzler R and Klafter J 2000 The random walk's guide to anomalous diffusion: a fractional dynamics approach *Phys. Rep.* **339** 1

[14] Golding I and Cox E C 2006 Physical nature of bacterial cytoplasm *Phys. Rev. Lett.* **96** 098102
[15] Manzo C, Torreno-Pina J A, Massignan P, Lapeyre G J Jr, Lewenstein M and Parajo M F G 2015 Weak ergodicity breaking of receptor motion in living cells stemming from random diffusivity *Phys. Rev.* X **5** 011021
[16] Krapf D, Lukat N, Marinari E, Metzler R, Oshanin G, Selhuber-Unkel C, Squarcini A, Stadler L, Weiss M and Xu X 2019 Spectral content of a single non-Brownian trajectory *Phys. Rev.* X **9** 011019
[17] 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
[18] 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
[19] Sokolov I M 2012 Models of anomalous diffusion in crowded environments *Soft Matter* **8** 9043
[20] Höfling F and Franosch T 2013 Anomalous transport in the crowded world of biological cells *Rep. Prog. Phys.* **76** 046602
[21] Horton M R, Höfling F, Rädler J O and Franosch T 2010 Development of anomalous diffusion among crowding proteins *Soft Matter* **6** 2648
[22] Tolić-Nørrelykke I M, Munteanu E-L, Thon G, Oddershede L and Berg-Sørensen K 2004 Anomalous diffusion in living yeast cells *Phys. Rev. Lett.* **93** 078102
[23] Jeon J-H, Leijnse N, Oddershede L B and Metzler R 2013 Anomalous diffusion and power-law relaxation of the time averaged mean squared displacement in worm- like micellar solutions *New J. Phys.* **15** 045011
[24] Saxton M J 1994 Anomalous diffusion due to obstacles: a Monte Carlo study *Biophys. J.* **66** 394
[25] Saxton M J 2001 Anomalous subdiffusion in fluorescence photobleaching recovery: a Monte Carlo study *Biophys. J.* **81** 2226
[26] Barkai E, Garini Y and Metzler R 2012 Strange kinetics of single molecules in living cells *Phys. Today* **65** 29
[27] Burov S, Jeon J-H, Metzler R and Barkai E 2011 Single particle tracking in systems showing anomalous diffusion: the role of weak ergodicity breaking *Phys. Chem. Chem. Phys.* **13** 1800
[28] Ernst D, Köhler J and Weiss M 2014 Probing the type of anomalous diffusion with single-particle tracking *Phys. Chem. Chem. Phys.* **16** 7686–91
[29] Großmann R, Bort L S, Moldenhawer T, Stange M, Panah S S, Metzler R and Beta C 2024 Non-Gaussian displacements in active transport on a carpet of motile cells *Phys. Rev. Lett.* **132** 088301
[30] 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
[31] Mandelbrot B B and van Ness J W 1968 Fractional Brownian motions, fractional noises and applications *SIAM Rev.* **10** 422
[32] Jeon J-H, Chechkin A V and Metzler R 2014 Scaled Brownian motion: a paradoxical process with a time dependent diffusivity for the description of anomalous diffusion *Phys. Chem. Chem. Phys.* **16** 15811
[33] Lim S C and Muniandy S V 2002 Self-similar Gaussian processes for modeling anomalous diffusion *Phys. Rev.* E **66** 021114
[34] Chechkin A V, Metzler R, Klafter J and Gonchar V Y 2008 Introduction to the theory of Lévy flights *Anomalous Transport: Foundations and Applications* vol 129 (Springer) (https://doi.org/10.1002/9783527622979)
[35] Lévy P 1937 *Théorie de l'Addition des Variables aléatoires* (Gauthier-Villars)
[36] Montroll E W and Weiss G H 1965 Random walks on lattices. II *J. Math. Phys.* **6** 167
[37] Hughes B D, Shlesinger M F and Montroll E W 1981 Random walks with self-similar clusters *Proc. Natl Acad. Sci. USA* **78** 3287
[38] Weissman H, Weiss G H and Havlin S 1989 Transport properties of the continuous-time random walk with a long-tailed waiting-time density *J. Stat. Phys.* **57** 301
[39] Shlesinger M F and Klafter J 1986 Lévy walks versus Lévy flights *On Growth and Form* ed N Ostrowsky and H E Stanley (Springer)
[40] Zaburdaev V, Denisov S and Klafter J 2015 Lévy walks *Rev. Mod. Phys.* **87** 483
[41] Elf J and Barkefors I 2019 Single-molecule kinetics in living cells *Ann. Rev. Biochem.* **88** 635
[42] Cherstvy A G, Thapa S, Wagner C E and Metzler R 2019 Non-Gaussian, non-ergodic and non-Fickian diffusion of tracers in mucin hydrogels *Soft Matter* **15** 2526
[43] Leijnse N, Jeon J H, Loft S, Metzler R and Oddershede L B 2012 Diffusion inside living human cells *Eur. Phys. J. Spec. Top.* **204** 377a
[44] Codling E A, Plank M J and Benhamou S 2008 Random walk models in biology *J. R. Soc. Interface* **5** 813
[45] Okubo A 1986 Dynamical aspects of animal grouping: swarms, schools, flocks and herds *Adv. Biophys.* **22** 1
[46] Vilk O *et al* 2022 Unravelling the origins of anomalous diffusion: from molecules to migrating storks *Phys. Rev. Res.* **4** 033055
[47] Bartumeus F, da Luz M G E, Viswanathan G M and Catalan J 2005 Animal search strategies: a quantitative random-walk analysis *Ecology* **86** 3078
[48] Malkiel B G 1999 *A Random Walk Down Wall Street: Including A Life-Cycle Guide to Personal Investing* (W. Norton & Co)
[49] Plerou V, Gopikrishnan P, Amaral L A N, Gabaix X and Stanley H E 2000 Economic fluctuations and anomalous diffusion *Phys. Rev.* E **62** R3023
[50] Metzler R, Tejedor V, Jeon J H, He Y, Deng W H, Burov S and Barkai E 2009 Analysis of single particle trajectories: from normal to anomalous diffusion *Acta Phys. Pol.* B **40** 1315
[51] Magdziarz M, Weron A, Burnecki K and Klafter J 2009 Fractional Brownian motion versus the continuous-time random walk: a simple test for subdiffusive dynamics *Phys. Rev. Lett.* **103** 180602
[52] Metzler R 2019 Brownian motion and beyond: first-passage, power spectrum, non-Gaussianity and anomalous diffusion *J. Stat. Mech.* 114003
[53] Vilk O, Aghion E, Nathan R, Toledo S, Metzler R and Assaf M 2022 Classification of anomalous diffusion in animal movement data using power spectral analysis *J. Phys. A: Math. Theor.* **55** 334004
[54] Condamin S, Bénichou O, Tejedor V, Voituriez R and Klafter J 2007 First-passage times in complex scale-invariant media *Nature* **450** 77
[55] Slezak J, Metzler R and Magdziarz M 2019 Codifference can detect ergodicity breaking and non-Gaussianity *New J. Phys.* **21** 053008
[56] Aghion E, Meyer P G, Adlakha V, Kantz H and Bassler K E 2021 Moses, Noah and Joseph effects in Lévy walks *New J. Phys.* **23** 023002
[57] Meyer P G, Aghion E and Kantz H 2022 Decomposing the effect of anomalous diffusion enables direct calculation of the Hurst exponent and model classification for single random paths *J. Phys. A: Math. Theor.* **55** 274001
[58] Sposini V, Grebenkov D S, Metzler R, Oshanin G and Seno F 2020 universal spectral features of different classes of random-diffusivity processes *New J. Phys.* **22** 063056
[59] Sposini V *et al* 2022 Towards a robust criterion of anomalous diffusion *Commun. Phys.* **5** 305

[60] 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
[61] Wyłomańska A, Chechkin A, Gajda J and Sokolov I M 2015 Codifference as a practical tool to measure interdependence *Physica* A **421** 412
[62] Thapa S, Lomholt M A, Krog J, Cherstvy A G and Metzler R 2018 Bayesian analysis of single-particle tracking data using the nested-sampling algorithm: maximum-likelihood model selection applied to stochastic-diffusivity data *Phys. Chem. Chem. Phys.* **20** 29018
[63] Thapa S, Park S, Kim Y, Jeon J-H, Metzler R and Lomholt M A 2022 Bayesian inference of scaled versus fractional Brownian motion *J. Phys. A: Math. Theor.* **55** 194003
[64] Park S, Thapa S, Kim Y, Lomholt M A and Jeon J-H 2021 Bayesian inference of Lévy walks via hidden Markov models *J. Phys. A: Math. Theor.* **54** 484001
[65] Seckler H and Metzler R 2022 Bayesian deep learning for error estimation in the analysis of anomalous diffusion *Nat. Commun.* **13** 6717
[66] Seckler H, Szwabiński J and Metzler R 2023 Machine-learning solutions for the analysis of single-particle diffusion trajectories *J. Phys. Chem. Lett.* **14** 7910
[67] Gentili A and Volpe G 2021 Characterization of anomalous diffusion classical statistics powered by deep learning (CONDOR) *J. Phys. A: Math. Theor.* **54** 314003
[68] 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
[69] 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
[70] Firbas N, Garibo-i-Orts Ò, Garcia-March M Á and Conejero J A 2023 Characterization of anomalous diffusion through convolutional transformers *J. Phys. A: Math. Theor.* **56** 014001
[71] AL-hada E A, Tang X and Deng W 2022 Classification of stochastic processes by convolutional neural networks *J. Phys. A: Math. Theor.* **55** 274006
[72] Bo S, Schmidt F, Eichhorn R and Volpe G 2019 Measurement of anomalous diffusion using recurrent neural networks *Phys. Rev.* E **100** 010102
[73] Gajowczyk M and Szwabiński J 2021 Detection of anomalous diffusion with deep residual networks *Entropy* **23** 649
[74] 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
[75] Garibo-i-Orts Ò, Baeza-Bosca A, Garcia-March M A and Conejero J A 2021 Efficient recurrent neural network methods for anomalously diffusing single particle short and noisy trajectories *J. Phys. A: Math. Theor.* **54** 504002
[76] Li D, Yao Q and Huang Z 2021 WaveNet-based deep neural networks for the characterization of anomalous diffusion (WADNet) *J. Phys. A: Math. Theor.* **54** 404003
[77] 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
[78] Seckler H, Metzler R, Kelty-Stephen D G and Mangalam M 2024 Multifractal spectral features enhance classification of anomalous diffusion *Phys. Rev.* E **109** 044133
[79] 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
[80] Muñoz-Gil G, Volpe G, García-March M A, Metzler R, Lewenstein M and Manzo C 2020 The anomalous diffusion challenge: single trajectory characterisation as a competition *Proc. SPIE* **11469** 42
[81] Muñoz-Gil G *et al* 2021 Objective comparison of methods to decode anomalous diffusion *Nat. Commun.* **12** 6253
[82] Yin S, Song N and Yang H 2018 Detection of velocity and diffusion coefficient change points in single-particle trajectories *Biophys. J.* **115** 217
[83] Saha S, Lee I-H, Polley A, Groves J T, Rao M and Mayor S 2015 Diffusion of GPI-anchored proteins is influenced by the activity of dynamic cortical actin *Mol. Biol. Cell.* **26** 4033
[84] Bag N, Huang S and Wohland T 2015 Plasma membrane organization of epidermal growth factor receptor in resting and ligand-bound states *Biophy. J.* **109** 1925
[85] Low-Nam S T, Lidke K A, Cutler P J, Roovers R C, van Bergen en Henegouwen P M P, Wilson B S and Lidke D S 2011 ErbB1 dimerization is promoted by domain co-confinement and stabilized by ligand binding *Nat. Struct. Mol. Biol.* **18** 1244
[86] Moldenhawer T, Moreno E, Schindler D, Flemming S, Holschneider M, Huisinga W, Alonso S and Beta C 2022 Spontaneous transitions between amoeboid and keratocyte-like modes of migration *Front. Cell Dev. Biol.* **10** 898351
[87] Monasson R and Rosay S 2014 Crosstalk and transitions between multiple spatial maps in an attractor neural network model of the hippocampus: collective motion of the activity *Phys. Rev.* E **89** 032803
[88] Wagner T, Kroll A, Haramagatti C R, Lipinski H-G and Wiemann M 2017 Classification and segmentation of nanoparticle diffusion trajectories in cellular micro environments *PLoS One* **12** e0170165
[89] Persson F, Lindén M, Unoson C and Elf J 2013 Extracting intracellular diffusive states and transition rates from single-molecule tracking data *Nat. Methods* **10** 265
[90] Stanislavsky A A and Weron A 2024 Fractional Lévy stable motion from a segmentation perspective *Fractal Fract.* **8** 336
[91] 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
[92] Arcizet D, Meier B, Sackmann E, Rädler J O and Heinrich D 2008 Temporal analysis of active and passive transport in living cells *Phys. Rev. Lett.* **101** 248103
[93] Muñoz-Gil G, Bachimanchi H, Pineda J, Midtvedt B, Lewenstein M, Metzler R, Krapf D, Volpe G and Manzo C 2024 Quantitative evaluation of methods to analyze motion changes in single-particle experiments *Nat. Commun.* Registered Report Stage 1 Protocol
[94] Requena B, Masó-Orriols S, Bertran J, Lewen- stein M, Manzo C and Muñoz-Gil G 2023 Inferring pointwise diffusion properties of single trajectories with deep learning *Biophys. J.* **122** 4360
[95] Qu X, Hu Y, Cai W, Xu Y, Ke H, Zhu G and Huang Z 2024 Semantic segmentation of anomalous diffusion using deep convolutional networks *Phys. Rev. Res.* **6** 013054
[96] Hochreiter S and Schmidhuber J 1997 Long short-term memory *Neural Comput.* **9** 1735

[97] Wilson A G and Izmailov P 2020 Bayesian deep learning and a probabilistic perspective of generalization *Advances in Neural Information Processing Systems* vol 33 p 4697
[98] Muñoz-Gil G, Manzo C, Volpe G, Garcia-March M A, Metzler R and Lewenstein M 2020 The anomalous diffusion challenge dataset (Zenodo) (https://doi.org/10.5281/zenodo.3707702) (available at: https://github.com/AnDiChallenge/ANDI_datasets)
[99] Jeffreys H 1946 An invariant form for the prior probability in estimation problems *Proc. R. Soc.* A **186** 453
[100] Jeffreys H 1998 *The Theory of Probability* (Oxford University Press)
[101] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A N, Kaiser Ł and Polosukhin I 2017 Attention is all you need *Advances in Neural Information Processing Systems* vol 30 (https://doi.org/10.48550/arXiv.1706.03762)
[102] Nix D A and Weigend A S 1994 Estimating the mean and variance of the target probability distribution *Proc. 1994 IEEE Int. Conf. Neural Networks (ICNN'94)* vol 1 (https://doi.org/10.1109/ICNN.1994.374138)
[103] Bottou L 2010 Large-scale machine learning with stochastic gradient descent *Proc. COMPSTAT'2010* (https://doi.org/10.1007/978-3-7908-2604-3_16)
[104] Kingma D P and Ba J L 2014 Adam: a method for stochastic optimization *3rd Int. Conf. for Learning Representations* (https://doi.org/10.48550/arXiv.1412.6980)
[105] Maddox W J, Izmailov P, Garipov T, Vetrov D P and Wilson A G 2019 A simple baseline for Bayesian uncertainty in deep learning *Advances in Neural Information Processing Systems* vol 32 (https://doi.org/10.48550/arXiv.1902.02476)
[106] Paszke A *et al* 2019 PyTorch: an imperative style, high-performance deep learning library *Advances in Neural Information Processing Systems* vol 32 (https://doi.org/10.48550/arXiv.1912.01703)
[107] Gal Y and Ghahramani Z 2016 Dropout as a Bayesian approximation: representing model uncertainty in deep learning *Int. Conf. on Machine Learning* (https://doi.org/10.48550/arXiv.1506.02142) (PMLR)
[108] Kendall A and Gal Y 2017 What uncertainties do we need in Bayesian deep learning for computer vision? *Advances in Neural Information Processing Systems* vol 30 (https://doi.org/10.48550/arXiv.1703.04977)
[109] Guo C, Pleiss G, Sun Y and Weinberger K Q 2017 On calibration of modern neural networks *Int. Conf. on Machine Learning* (https://doi.org/10.48550/arXiv.1706.04599) (PMLR)
[110] DeGroot M H and Fienberg S E 1983 The comparison and evaluation of forecasters *Statistician* **32** 12–22
[111] Levi D, Gispan L, Giladi N and Fetaya E 2022 Evaluating and calibrating uncertainty prediction in regression tasks *Sensors* **22** 5540
[112] Chenouard N *et al* 2014 Objective comparison of particle tracking methods *Nat. Methods* **11** 281
[113] Kuhn H W 1955 The Hungarian method for the assignment problem *Nav. Res. Logist. Q.* **2** 83
[114] Jaccard P 1901 Étude comparative de la distribution florale dans une portion des Alpes et des Jura *Bull. Soc. Vaudoise Sci. Nat.* **37** 547
[115] Kowalek P, Loch-Olszewska H, Łaszczuk Ł, Opała J and Szwabiński J 2022 Boosting the performance of anomalous diffusion classifiers with the proper choice of features *J. Phys. A: Math. Theor.* **55** 244005
[116] Loch-Olszewska H and Szwabiński J 2020 Impact of feature choice on machine learning classification of fractional anomalous diffusion *Entropy* **22** 1436
[117] Naeini M P, Cooper G and Hauskrecht M 2015 Obtaining well calibrated probabilities using Bayesian binning *29th AAAI Conf. on Artificial Intelligence* (https://doi.org/10.1609/aaai.v29i1.9602)
[118] Meyer P G and Metzler R 2023 Stochastic processes in a confining harmonic potential in the presence of static and dynamic measurement noise *New J. Phys.* **25** 063003
[119] Thapa S, Kantz H, Metzler R and Lomholt M A 2024 Bayesian inference of change-points for fractional Brownian motion (unpublished)
[120] Meyer P G, Cherstvy A G, Seckler H, Hering R, Blaum N, Jeltsch F and Metzler R 2023 Directedeness, correlations and daily cycles in springbok motion: from data over stochastic models to movement prediction *Phys. Rev. Res.* **5** 043129
[121] 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
[122] 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