

PAPER

# Decomposing the effect of anomalous diffusion enables direct calculation of the Hurst exponent and model classification for single random paths

To cite this article: Philipp G Meyer *et al* 2022 *J. Phys. A: Math. Theor.* **55** 274001

View the [article online](#) for updates and enhancements.

## You may also like

- [Discrete-time random walks and Lévy flights on arbitrary networks: when resetting becomes advantageous?](#)  
Alejandro P Riascos, Denis Boyer and José L Mateos
- [Halos and undecidability of tensor stable positive maps](#)  
Mirte van der Eyden, Tim Netzer and Gemma De las Cuevas
- [Classification of anomalous diffusion in animal movement data using power spectral analysis](#)  
Ohad Vilik, Erez Aghion, Ran Nathan et al.

# Decomposing the effect of anomalous diffusion enables direct calculation of the Hurst exponent and model classification for single random paths

Philipp G Meyer<sup>1,2,4</sup> , Erez Aghion<sup>2,3,4,\*</sup>  and Holger Kantz<sup>2</sup> 

<sup>1</sup> Institute of Physics & Astronomy Karl-Liebknecht-Str. 24/25, University of Potsdam, D-14476 Potsdam-Golm, Germany

<sup>2</sup> Max-Planck Institute for the Physics of Complex Systems, D-01187 Dresden, Germany

<sup>3</sup> Departments of Physics and Chemistry, University of Massachusetts Boston, MA 02125, United States of America

E-mail: [philipp.meyer@uni-potsdam.de](mailto:philipp.meyer@uni-potsdam.de) and [erezagh5@gmail.com](mailto:erezagh5@gmail.com)

Received 6 December 2021, revised 16 May 2022

Accepted for publication 24 May 2022

Published 13 June 2022



## Abstract

Recently, a large number of research teams from around the world collaborated in the so-called ‘anomalous diffusion challenge’. Its aim: to develop and compare new techniques for inferring stochastic models from given unknown time series, and estimate the anomalous diffusion exponent in data. We use various numerical methods to directly obtain this exponent using the path increments, and develop a questionnaire for model selection based on feature analysis of a set of known stochastic processes given as candidates. Here, we present the theoretical background of the automated algorithm which we put for these tasks in the diffusion challenge, as a counter to other pure data-driven approaches.

**Keywords:** time-series analysis, decomposing anomalous diffusion, anomalous diffusion exponent, process inference

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

\*Author to whom any correspondence should be addressed.

<sup>4</sup>Both authors contributed equally to this work.

## 1. Introduction

The phenomenon of anomalous diffusion was reported in data obtained from many different systems, e.g., microbiological and ecological experiments [1–3] (and see references within), financial data [4], and more e.g. [5, 6]. Processes that exhibit anomalous diffusion are characterized by the scaling of their mean square-displacement (MSD), with respect to time  $t$

$$\langle x^2(t) \rangle \sim t^{2H}, \quad (1)$$

where  $\sim$  is to be understood as *asymptotically proportional to*. Here the so-called *Hurst exponent*  $H$  is  $\neq 1/2$ . In normal diffusion  $H \equiv 1/2$ . The value  $2H$  is often called the anomalous diffusion exponent [7]. In most literature examples  $H$  is inferred from an ensemble of trajectories via equation (1), however many real world systems, where only one realization is measurable (for example, in geoscience [8, 9]), are suspected to exhibit anomalous diffusion too. This phenomenon can arise from different dynamical origins, and be described by various stochastic models. In addition, the effect can be very weak, with  $H$  being only slightly different from  $1/2$ , or much stronger, in which case  $|H - 1/2|$  is larger. So-called subdiffusion occurs when  $H < 1/2$ , and superdiffusion when  $H > 1/2$ .

The anomalous diffusion challenge (ANDI) [6, 7] is an initiative to find new algorithms for the task of inferring the anomalous diffusion exponent from single trajectories, in 1, 2 and 3 dimensions, classifying the type of dynamics which generated them and finding the models most suitable to describe it. The third part of the challenge was related to process segmentation [7], but this is beyond the scope of our current paper. Several teams have worked in parallel on these tasks and compared the performance of their algorithms on a uniform dataset prepared by the organizers. The best-performing methods were found to be based on deep learning algorithms [6, 10, 11], e.g., using recurrent neural networks with long short-term memory layers [12, 13] and huge training sets which are often specifically chosen for various path-lengths. Even though such algorithms are advantageous, especially when dealing with short time series, they also have some significant drawbacks. First, they require a long computation time, and second, they only characterize the models that they were specifically trained on and might give wrong suggestions for other dynamics. A third issue is their complexity, because the quality of the results may depend strongly on the training method, for example; if the network was trained on an ensemble of mixed-length paths, or a set of separate uniform-length ensembles [6]. Therefore, constructive methods based on theory remain highly important, and they can be used for example to enhance confidence in the results of a purely data-driven algorithm, or in cases when the use of the latter is less practical.

One standard way for calculating the anomalous diffusion exponent in data uses the ensemble-averaged mean squared displacement [14]. For single trajectory analysis, this is often replaced by the time-averaged square-displacement (TASD)

$$\overline{\delta^2(\Delta)} = \frac{1}{T - \Delta} \sum_{t=0}^{T-\Delta} [x(t + \Delta) - x(t)]^2, \quad (2)$$

where it is assumed that  $\overline{\delta^2(\Delta)} \propto \Delta^{2H}$ . This assumption, however, is wrong for many stochastic processes, as we will discuss in this manuscript. Our alternative, a straightforward calculation of the anomalous diffusion exponent presented below, at the same time offers a way of distinguishing between different processes based on their theoretical features. It utilizes a decomposition of the Hurst exponent into the Joseph, Noah, and Moses effects, proposed by Chen *et al* [15] based on Mandelbrot's work [16]. As shown in [6], our approach performs sig-

nificantly better than the TASD in inferring the anomalous diffusion exponent. However, just relying on this decomposition is not sufficient for a reliable classification of the anomalous diffusion model, thus we mixed it with a basic data-driven classification algorithm in order to close the performance gap with other methods.

The paper is organized as follows: in section 2, we introduce the underlying theoretical considerations behind the three-effect decomposition of anomalous diffusion. In section 3 we discuss the models considered in ANDI, and the calculation of the Moses, Noah, and Joseph effects in each of them. These are; fractional Brownian motion (FBM) [17], scaled Brownian motion (SBM) [18], continuous time random walk (CTRW) [19], Lévy walks (LWs) [20], and annealed transient time motion (ATTM) [21]. In sections 4 and 5 we show how the theory can be applied for the task of inferring the anomalous diffusion exponent in unknown data, and for process classification, respectively. The discussion appears in section 6.

## 2. Decomposing the Hurst exponent into basic elements

We define a diffusion process as ‘normal’ if it obeys the Gaussian central limit theorem [22]. The theorem addresses the statistical properties of any process which can be written as a sum of random increments, or  $v_i$ , that are *independent* (or with positive temporal correlations that decay at least exponentially fast with time), *identically distributed*, and with a *finite variance*. If the increment series has all these properties, the sum  $x = \sum_{i=1}^N v_i$  is Gaussian-distributed at large  $N$  and  $H = 1/2$ . If one of the premises of the theorem is violated, i.e., if (a) the increments are long-ranged or negatively correlated, (b) the process is non-stationary, or (c) the variance of the distribution is not finite, anomalous diffusion may occur due to (a) the Joseph effect, (b) the Moses effect, or (c) the Noah effect, which are quantified by the Joseph, Moses and Noah exponents  $J, M, L$  respectively [15]. This three-effect decomposition was demonstrated also for dynamical systems [23], LWs [24], citation-history data of scientific publications [25], and microbiological and ecological data [3]. The three exponents, together with  $H$ , fulfill the summation relation [15]

$$H = J + M + L - 1. \quad (3)$$

The summation relation in equation (3) holds generally with the only assumption of some scaling behavior of the MSD, see [24] for an analytic derivation, and [3, 25] for evidence from empirical data. It is important to note however, that as shown by the empirical studies, in data the scaling exponents might not always be constant over all timescales.

As opposed to the assumption below equation (2), in [23, 24] it was shown that instead of  $H$ , the scaling of the TASD equation (2) with respect to  $\Delta$  in fact only corresponds to the Joseph exponent  $J$

$$\overline{\delta^2(\Delta)} \propto \Delta^{2J}, \quad (4)$$

at large  $\Delta$ . This result was shown to follow directly from the relation between the scaling of the TASD and the velocity–velocity autocorrelation function [24, 26]. When there is no long-ranged or negative autocorrelation,  $J = 1/2$ . On the other hand, long-ranged positively-correlated processes are characterized by  $J > 1/2$ , and for anti-correlated processes  $J < 1/2$ . In any case, when there is a Moses or Noah effect, besides Joseph, as shown in equation (3);  $2H \neq 2J$ .

Another measure that can be used for obtaining  $J$ , besides equation (4), is the *fluctuation function* of detrended fluctuation analysis (DFA) [27]. In appendix B and reference [24] we explain this method in more detail, and compare it with the TASD. The Noah and

Moses effects are related to the increment probability density function (PDF). Their combined effect in the time series is observed by the scaling of the sum of squared increments;  $v(n) = x(n) - x(n-1)$ :

$$\sum_{n=1}^{t-1} v^2(n) \sim t^{2M+2L-1}. \quad (5)$$

If the PDF of  $v(n)$  has fat, power-law shaped tails that fall off as  $\sim |v|^{-\beta}$  at large values, with  $2 < \beta < 3$ , its second moment (namely the mean  $\langle v^2 \rangle$  over an infinite ensemble of time series) diverges, while the first moment is still finite. In this case, if  $M = 1/2$ , namely if the process is stationary; we conjecture that  $L = 2 - \beta/2$  (details in appendix A). In contrast to this, intrinsic non-stationarity adds a time dependence to both the first and the second moment. Therefore, the Moses exponent  $M$  can be defined from the scaling of the cumulative absolute increments

$$\sum_{n=1}^{t-1} |v(n)| \sim t^{M+1/2}. \quad (6)$$

If the tails of the velocity PDF decay even slower than  $|v|^{-2}$ , then we can get a combination of both Noah and Moses effects, which can be disentangled by combining equations (5) and (6).

### 3. Expectation values for some known anomalous diffusion processes

As mentioned in the introduction, in ANDI [6] five distinct anomalous diffusion models have been studied. Here we present the various values of  $J$ ,  $L$  and  $M$  for these processes.

#### 3.1. Fractional Brownian motion

FBM is the paradigmatic model for the Joseph effect [14, 15, 17]. In one dimension, it is characterized by the autocorrelation function  $\langle x(t_1)x(t_2) \rangle \propto t_1^{2H} + t_2^{2H} - |t_1 - t_2|^{2H}$ . The increment process is stationary and Gaussian, and therefore does not show a Moses nor a Noah effect. In this case, the time averaged mean squared displacement (or DFA) directly yields the correct anomalous diffusion exponent, since  $H = J$ . In two dimensions, the latter is valid for  $\langle |\vec{r}|^2 \rangle$ , where  $\vec{r}$  is the position vector  $\vec{r} = \vec{x} + \vec{y}$  (this extends similarly also to higher dimensions), and the autocorrelation function is presented in [6].

#### 3.2. Scaled Brownian motion

Uni- and multidimensional SBM is the paradigmatic model for the Moses effect [15]. This motion is similar to Brownian motion, but with a time-dependent diffusion coefficient (this is true in any number of dimensions). Here, both the absolute mean and the variance of the increments  $v$  are time-dependent [15]. More explicitly, in the model used in ANDI, SBM is generated by Brownian motion with a diffusion-coefficient whose time-dependence follows a power-law [7, 14]. This leads to a Moses effect via equation (6) [15], where  $M$  can be either larger or smaller than  $1/2$ , if the diffusion coefficient becomes larger or smaller with time, respectively. In this process, however, there is no effect from rare extremes, hence  $L = 1/2$ . Also, since the increments of the process are independent, the Joseph exponent is  $J = 1/2$  [15].

**Table 1.** Anomalous diffusion exponents  $J$ ,  $M$ , and  $L$  for the five processes. For FBM, LW, and SBM the Hurst exponent is equivalent to one of the three constituents. For CTRW  $M$ , and  $L$  both contribute to  $H$ . ATTM is more complex, as it exhibits three phase region with different scaling relations and with an additional parameter independent of  $H$  (here we chose  $L$ ).

Process	$J$	$M$	$L$
ATTM ( $\sigma < \gamma < \sigma + 1$ )	$1/2$	$2H - 1/2$	$1 - H$
ATTM ( $\sigma < \gamma - 1 < 2\sigma$ )	$H + L - 1/2$	$3/2 - 2L$	$L$
ATTM ( $2\sigma + 1 < \gamma$ )	$3/2 - L$	$H - 1/2$	$L$
CTRW	$1/2$	$2H - 1/2$	$1 - H$
FBM	$H$	$1/2$	$1/2$
LW	$H > 1/2$	$1/2$	$1/2$
SBM	$1/2$	$H$	$1/2$

### 3.3. Continuous-time random walk

In unidimensional CTRW, the process consists of a series of random sojourn times  $\{\tau\}$ , chosen from a power-law distribution with no typical scale, each followed by a random jump displacement (see, e.g. [28]). In ANDI, the sojourn time PDF  $g(\tau) \sim \tau^{-1-2H}$  at large  $\tau$ , where  $0 < H < 1/2$ , and the jumps PDF is non fat-tailed, e.g., Gaussian. In  $d = 2$  and  $d = 3$  dimensions, one can generate CTRW such that after each sojourn time, the particle makes an independent random jump along each axis, then a new waiting time is chosen. Another option is to choose a random direction first and then the jump occurs along the radial axis [29]. In any case, the process ends when  $\sum_{i=1}^N \tau_i \geq t$ , where  $t$  is the total measurement time and the displacement at every dimension is the sum of  $N$  random jumps. The three-effect decomposition gives similar results in any dimension, regardless of the method by which the path was generated.

Here, since the waiting times have no finite typical mean value, two things occur: first, the process slows down with time, since as  $t$  increases, longer and longer waiting times are selected. Second, almost every single jump event in this process constitutes a ‘rare extreme event’. This means that in this process, we have a Moses effect, which *leads to* a subsequent Noah effect. Here, by direct calculation, we can find that  $M = 2H - 1/2$ ,  $L = 1 - H$  and  $J = 1/2$ , see table 1. In addition, we note that  $M$  and  $L$  satisfy a particularly special relation

$$L = 3/4 - M/2. \quad (7)$$

which indicates that the first and second moments of the velocity are obtained from an infinite-invariant density function, see appendix F for the explanation. We show in section 5, that we can use this property to infer this process in a ‘blind’ test.

### 3.4. Lévy walk

In simple LW, see [30] for a review, the random-walker chooses a series of jump durations, statistically similar to the sojourn times of CTRW above, until their sum crosses the total measurement time, and travels in a series of steps whose sizes are coupled to the durations; each displacement is  $\pm \tau$  (in one dimension), proportional to the time. The  $\pm$  direction is chosen randomly with equal probability. As in CTRW, in high dimensions in ANDI the duration at each step is chosen once, then a random direction and random displacement along the radial axis [6]. The three exponents for unidimensional LW were analyzed in [24], and they are similar

also in  $d > 1$ , see table 1. Here, there are no Moses and Noah effects:  $M = L = 1/2$ . A time-averaged MSD immediately yields the correct anomalous diffusion exponent, since  $H = J$ , although note that other models of non-linearly coupled LW lead to richer dynamics.

### 3.5. Annealed transient time motion

In ATTM [21], the particle moves in a series of intervals of Brownian motion (in one dimension or above), with random diffusivity  $D_i$  at each interval, and a random duration  $\tau_i$ . The diffusivities are fat-tailed distributed, with PDF  $P_D(D) \sim D^{\sigma-1}$  at large  $D$ s, and correlated to the step durations via  $\tau_i = D_i^{-\gamma}$  (where  $\gamma$  and  $D$  are in the range  $\in (-\infty, \infty)$ ). Consequently, the PDF of the durations  $\psi(\tau) = \int_0^\infty \delta(\tau - D^{-\gamma}) P_D(D) dD$  (where  $\delta(\cdot)$  is the Dirac delta function) is therefore  $\sim \tau^{-1-\sigma/\gamma}$  at large  $\tau$ s. As shown in [21], the resulting MSD of this process, has three scaling regimes; two of them yield anomalous diffusion: (i) when  $\sigma < \gamma < \sigma + 1$ ;  $\langle x^2(t) \rangle \propto t^{\sigma/\gamma}$  and  $H = \sigma/(2\gamma)$ , and (ii) when  $\sigma + 1 < \gamma$ ;  $\langle x^2(t) \rangle \propto t^{1-1/\gamma}$  and  $H = 1/(2\gamma)$ . In the analysis of  $J, L$  and  $M$ , the last regime further divides into two subregimes as well, see table 1.

Since in Brownian motion, the displacement in an interval of motion of duration  $\tau$  is  $\chi \sim \sqrt{D\tau} \sim \tau^{1/2-1/(2\gamma)}$ , this motion can be mapped into a nonlinearly coupled LW-type process with coupled step-duration  $\tau$  and displacement  $\chi$ , as in [30–33], and then we can use existing results found in [21, 24] to obtain the values of  $M, L, J$  for this process as well. The different regimes of the Moses, Joseph, Noah, and Hurst exponents are displayed in table 1. We remark that for any process that can be mapped into a non-linearly coupled LW model, we can find phase diagrams for the exponents in a similar manner [24].

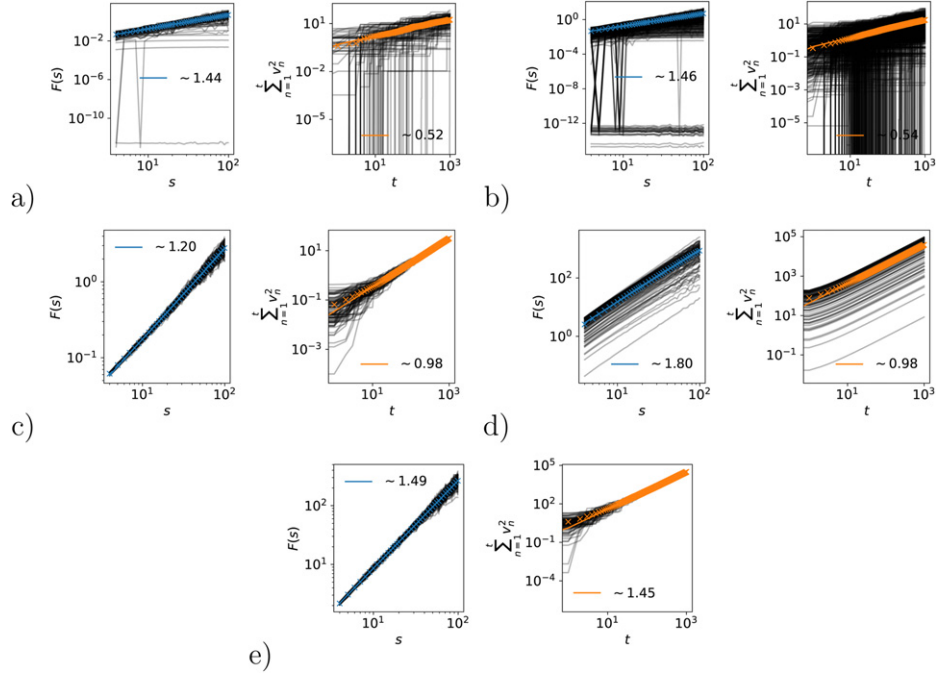
Note that in 2 and 3 dimensions, in ANDI, the waiting times and diffusion coefficient were sampled at each step like in one dimension, and then the path was generated by decomposing independent Brownian motion paths at each dimension.

## 4. Estimating the exponents

Here, we show the numerical calculation of the exponents for one anomalous diffusion exponent of each stochastic process in figure 1. We see that the expectation values of  $M + L + J - 1$  are an unbiased measure for  $H$ . For FBM, SBM, and LW this measure works well for single trajectories, while for ATTM and CTRW the errors are larger. Nevertheless, it is a simple way to straightforwardly calculate the anomalous diffusion exponent. We simply combine the calculation of  $J$  (performed using the T ASD or DFA, see appendices B and D) with a measure of the squared increments, accounting for  $M$  and  $L$ . As a test of consistency, one might also calculate  $M$  from the absolute increments. This gives a lower bound to  $M + L$ , since  $L \geq 1/2$  always [24].

Our approach faces two challenges from the setup of the ANDI competition. First, in CTRW and ATTM, individual paths do not represent correctly the ensemble, due to weak ergodicity breaking [6]. This could be solved by changing the way how the time average is done (i.e. if we divide the path into time windows that begin at the ends of waiting periods, instead of using a uniform moving average), however, this is complicated to achieve and practically impossible for short trajectories. We, therefore, have to accept this limitation of our approach. Second, to simulate real-world measurement, the datasets are corrupted by external noise which is not inherent to any of the five stochastic models in the competition. This noise can act in an additive way as well as in a multiplicative way [6]. This issue affects all the trajectories to different degrees. While in superdiffusive time-series the scaling is only weakly affected by the noise





**Figure 1.** Calculating the exponents  $J + 1$  (right panels) and  $2M + 2L - 1$  (left panels) for the five stochastic processes. (a) ATTM with  $2H = 0.5$ , (b) CTRW with  $2H = 0.5$ , (c) FBM with  $2H = 0.5$ , (d) LW with  $2H = 1.5$ , and (e) SBM with  $2H = 1.5$ . The gray lines indicate the results of individual paths with length 1000, in an ensemble of 1000 trajectories. The colored lines represent the ensemble mean for each point.

for long lag times, for subdiffusive dynamics this is a bigger challenge due to the similarity between signal and noise.

#### 4.1. Estimating $J$

In order to filter out the additive noise we apply DFA [27] (appendix B) of order one (below we refer to it as DFA(1), for short) on the process itself rather than its increments. Typically, DFA is applied to non-diffusive time series, i.e., the increments [34], just like the TASD (2) and R/S statistics [15]. The algorithm is as follows. First, the cumulative sum is calculated  $y(t) = \sum_{n=1}^t x(n)$ . Then, the time-series  $y(t)$  is divided into  $N$  non-overlapping segments of length  $\Delta$ . From each of these segments a linear fit  $p_k(t)$  is subtracted. The fluctuations in the segment are quantified by the variance of the residuals. Finally, the fluctuation function  $F(s)$  is calculated as a function of segment length  $\Delta$  by averaging the  $t/\Delta$  (i.e. the largest integer value  $< t/\Delta$ ) segments with the same length. It scales with  $2J + 2$

$$F^2(\Delta) = \frac{1}{t/\Delta} \sum_{k=1}^{t/\Delta} \left[ \frac{1}{\Delta} \sum_{i=1}^{\Delta} |y(i + (k-1)\Delta) - p_k(i)|^2 \right] \sim \Delta^{2J+2}. \quad (8)$$

Note that, if  $x(t)$  has more than one dimension,  $y(t)$  is evaluated as the cumulative process in each dimension independently, and the linear fit  $p_k$  is produced in each dimension. The squared deviations are then added up, as the absolute value in the equation suggests. An alternative way



is to use a moving average of the increments  $v$  instead of using the trajectory. More details on both approaches and a numerical evaluation of the performance are given in appendices B and D.

#### 4.2. Estimating $M$ and $L$

For calculating  $M$  and  $L$ , it is not possible to get rid of the noise by using a cumulative time series. For this reason, the noise is fully visible. One way to filter the noise would be to apply a moving average to  $v^2$ . A second idea is to calculate squared deviations in windows, inspired by the DFA fluctuation function. It is explained in appendices B and C.

A third way is to subtract a constant from the squared increments, that corresponds to the estimated variance of the noise (note, that the noise variance is always constant in ANDI). Such an estimation could be the median of the squared increments or be inferred from the short-time behavior of the DFA fluctuation function of  $v$ . The corresponding scaling function is

$$G(t) = \sum_{n=1}^t (|v(n)|^2 - \text{const}) \sim t^{2M+2L-1}. \quad (9)$$

Analogously, we can define a scaling function for  $M$  without the square value on  $|v(n)|$ , as in equation (6). Note, that calculating  $M$  individually is not necessary for task 1 of ANDI, as only the sum of  $M$  and  $L$  is relevant for the anomalous diffusion exponent. However, knowing both exponents individually can be useful for consistency checks on the result and the process inference in task 2.

More details and a numerical evaluation of the performances of promising methods are given in appendix C. We found that a combination of two methods leads to the best performance.

### 5. Process inference

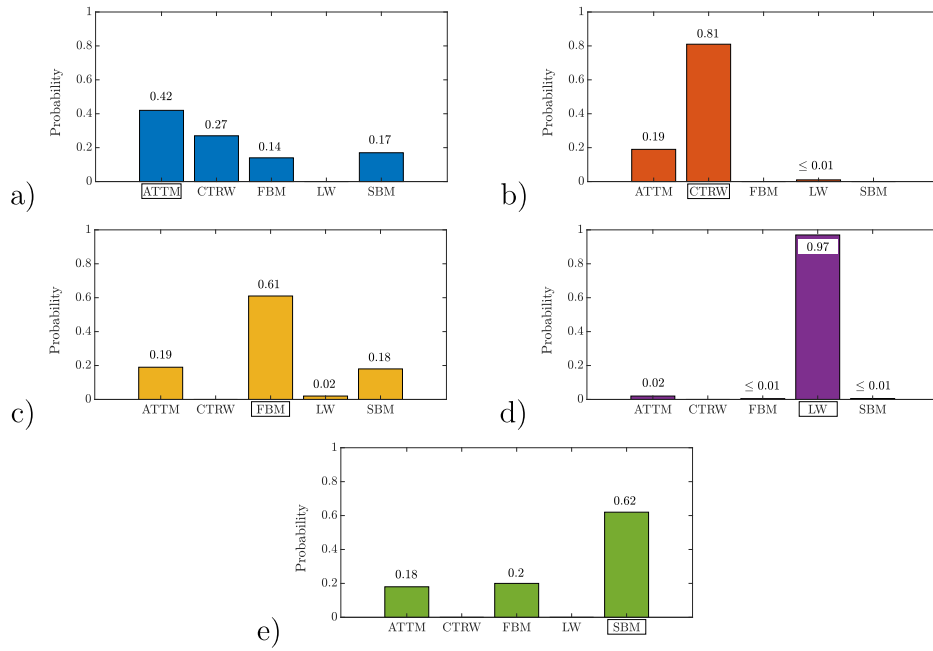
As mentioned in the introduction, one of the main tasks in ANDI was the following: given a time-series  $x(t)$ , with arbitrary length, can we develop an automated algorithm to provide the probability that this time-series was generated by FBM, SBM, CTRW, LW or ATTM? Here we present our idea for tackling this task, and the method by which we applied it (see also appendix E). Any educated (theory-based) approach that one could derive for this task, as opposed to data-driven methods such as machine learning, will be based on the formulation of a set of questions about various features of the process, to which numerical analysis of the time-series will try to provide quantitative or qualitative answers. Then, those answers are used to sort the properties of the path. The most immediate idea then might be to construct a sort of *decision tree* for this purpose, which will either lead to or exclude some of the models, until finally deciding on a single one. The problem with this approach is due to the inaccurate nature of the data. In practice, no single question/answer can be conclusive enough to fully exclude any model. Therefore, a better approach is the ‘parallel’ one, which only makes a conclusion after all the answers to all the proposed questions have been answered. The list of answers generated for an unknown time series can then be compared with the statistical probability to obtain the same set of answers in an equal-sized ensemble of simulated paths from all five processes, where the model is always known. This is the data-driven element of our method, however, it includes far fewer parameters to be ‘trained’ than other data-driven methods. The technical details about the use of the simulated paths in order to get probabilities are found in appendix E.

Based on our knowledge of the five processes in ANDI, we aimed the small set of questions towards the statistics of the increments of the time series, in order to harness the decomposition of anomalous diffusion described above. This choice allows us to utilize an array of characteristic properties from each process, and at the same time construct questions that are comparative in nature, namely which do not rely on arbitrary decision-thresholds or exact numerical values. By definition, the answer to each of the questions below is either ‘yes’ ( $= 1$ ), ‘no’ ( $= 0$ ), or ‘maybe’ ( $= 2$ ). The latter occurs when the data obtained from the path is insufficient to give a strict 0/1 answer, which itself is still a valid result since sometimes ‘maybe’ can be the most probable answer for certain questions, for certain specific models, see table E1 in appendix E. Here are few examples for questions we used, which are also explained in detail in the appendix:

- Which effect is most prominent in the process? For example, is  $|J - 1/2| > |L - 1/2|$  and  $|J - 1/2| > |M - 1/2|$ ? Namely, is Joseph the most prominent effect in this process? This question screens for FBM and LW, where the answer is ‘yes’, as opposed to e.g., SBM. Alternatively, this question can be reformulated to point to the dominance of the Moses or Noah effects, in order to screen for other processes. Here, in CTRW, for example, the most probable answer may be ‘maybe’, because the nature of the process makes exponent estimation unreliable.
- Is  $J > 1/2$ ? Namely, is the process temporally long-ranged correlated?  
A ‘no’ answer, meaning that  $J < 1/2$ , for processes where  $J$  is the dominant exponent, is a clear indicator of subdiffusive FBM and excludes the superdiffusive LW.
- Is the relation between  $M$  and  $L$  consistent with an infinite-density function, like in CTRW, where  $M, L$  are related via equation (7) or is it closer to ‘trivial scaling’ expected when all the moments of the velocity scale in the same way as in SBM? (see explanation in appendix E.)

Fully distinguishing between all models is not feasible solely with the three-effect decomposition. The above questions are not able to distinguish between superdiffusive FBM and LW, and also cannot single out ATTM, since the latter displays a large range of values at different parameter regimes. We can push the exponents-based analysis a little bit more, by performing simple manipulations on the paths. For example, it turns out that LW can be inferred by adding a small trick. Since in the simple LW model used in ANDI all the velocities are only either  $= 1$  or  $-1$ : if we take the absolute value of the increments and subtract their new mean value, we are able to tap only to the external noise element ‘dressed’ upon the process (see the introduction and [6]). This procedure does not work for FBM because the size of the velocities is random. Since, as mentioned, the external noise is known to be uncorrelated, finding the combination of  $J > 1/2$ , when this exponent is evaluated from the original increments process, and  $J \approx 1/2$  for the transformed time series indicates LW and excludes FBM. Beyond ‘tricks’ like this, in order not to rely on the three-effect decomposition alone, besides the questions above we added two more questions about other properties of the increments of the paths (details in appendix E): (i) a ‘fat-tailness’ test, and (ii) a ‘plateau detection’ test, designed to distinguish ATTM from CTRW with waiting times. Note that the model inference method described above is similar in one, and higher dimensions.

In figure 2, we provide ‘confusion charts’, demonstrating the average probabilities of various trajectories from different stochastic models to be correctly or incorrectly be classified using our method. For this example, we used a set of answers generated for 4000 trajectories from each stochastic model; ATTM, CTRW, FBM, LW and SBM (in total 20 000 paths), of 1000 time steps, out of which 40 paths (1% of the total) of each kind were arbitrarily extracted.



**Figure 2.** Confusion charts demonstrating our method for process inference. To produce the bar-plots, in our simulations we generated 4000 trajectories from each model; ATTM, CTRW, FBM, LW, and SBM, in one dimension, of an equal number of 1000 steps. We used our code [35] to generate answers to each of our classifying questions, as explained in section 5 and appendix E, for all the trajectories. The resulting list of answers (also found online [35]) is our ‘training set’. We arbitrarily chose 40 trajectories (1%) of each model, took them out of the ensemble, and calculated back the probability that they were generated from each model, based on their relative occurrence in the training set. The numerics and the noise in the paths make the classification results probabilistic. Panel (a) shows the averaged results of 40 ATTM paths; on average these paths, for example, were correctly identified with probability = 0.42, and misclassified as CTRW, with probability 0.27, FBM; 0.14 and SBM; 0.17. (b) Mean results from CTRW paths, (c) FBM, (d) LW and (e) SBM. The results demonstrate good classification ability, with the best results obtained for LW paths, and the worst for ATTM. The quality of the classification improves with better numerical results for longer paths. These sample probabilities may also vary a little between realizations of similar sized ensemble. The paths used in this analysis were generated using the algorithms provided by the ANDI challenge [6] and were subjected to the same types of additional noise.

We applied our classification algorithm to these paths and obtained the probabilities that they were generated from each of the models. Diagram (a) shows that ATTM paths were correctly classified with a probability of 0.42, and wrongly classified as CTRW with probability 0.27, FBM: 0.14, and SBM: 0.17. This was the hardest model to classify. In (b) we see the results for CTRW paths. These were correctly detected with probability 0.81, and wrongly reported as ATTM with probability 0.19. Almost none of the paths were wrongly attributed to the other models. (c) FBM is correctly classified with probability 0.61, as ATTM with 0.19, LW with 0.02 and SBM with 0.18. (d) LW was excellently detected, with probability 0.97, while (e) SBM was detected with 0.62. The data and codes we used for this analysis are found in [6, 35].

## 6. Discussion

The ANDI clearly showed the superiority of machine learning methods over methods that require less intensive training for model selection and parameter estimation. Due to the black box character of most machine learning methods, however, this result does not undermine the usefulness of other, more ‘classical’ methods. These are still required for understanding the most important features of the data, and for verification of the trained models which always suffer from sample errors, namely its performance is impaired if the unknown test data does not resemble the data featured in the training set. Particularly, the three-effect decomposition of anomalous diffusion, which is at the center of our method, has been shown to lead to significant new insights in empirical data from a variety of sources [3, 15, 25], which cannot be reviled using an algorithm focused on finding value for a single anomalous diffusion exponent.

Here, we proposed several different methods for calculating both  $J$  and  $M + L$  for noisy time series. These methods are best applied together since their errors are not strongly correlated. The best single method for inferring  $J$  is a DFA(1) of the cumulative process, with estimated noise subtracted. For  $M + L$ , it is the direct calculation of the scaling of the moments of the increments of the process (with time), after subtracting an estimated noise level. For a quick estimation of the exponent  $H$  for noisy time series, we recommend these approaches.

For the task of process classification, we found that a pure decision tree is not feasible due to the inaccuracy of our estimators for process properties. Instead, we combine a set of selected questions with the training of the respective probabilities for all models. Using seven questions that are mostly inspired by the decomposition of the anomalous diffusion into the Joseph, Moses, and Noah effects, we were able to classify the correct model (from the five options in the competition) for unknown processes, with accuracy close to 50% according to the evaluation in ANDI. In contrast, a purely random selection would yield an accuracy of only 20%. We remark that for this task, and for practical applications beyond it, improving the selectivity of the set of questions, adding new tests, or finding better ones, will improve the quality of the final results. Similarly, the quality of the results depends on the time-series analysis techniques used to obtain e.g., the Joseph exponent (see appendices D and C). We tried to avoid defining parameters for hard thresholds which are difficult to estimate due to the noisy signals. However, we were not able to define the classifiers solely without them. These parameters are based on educated guesses which could be improved with big data and a sophisticated fitting algorithm.

The performance of all methods participating in the ANDI is summarized in the supplementary material of [6]. For the task of inferring  $H$  our method performs well for time series of length  $> 250$  (time points), but less so for very short time series. The average absolute error in our estimation of  $H$  was around 10%, independent of the number of dimensions. Our additional analysis in appendices D and C shows that the accuracy of  $J$  is similar for all processes, while  $M + L$  is most accurate for FBM, LW, and SBM. It also shows the benefit of using a combination of calculation techniques for each exponent, rather than relying on the results of a single method. For the task of process inference, we achieved good hit rates for the same three processes; FBM, LW, and SBM. The process which is most difficult to detect using our method is ATTM.

Our struggle with the method shows that a straightforward calculation of the anomalous diffusion exponent and a reliable process inference is difficult, especially for short time series, and that external noise is a challenge for methods based on theoretical considerations. Nevertheless, we provide several ideas on how to build a theory-based algorithm for this task, which far exceeds the performance of the classical T ASD calculation. We tried to keep the

balance between performance and an algorithm that is still understandable and does not consume too long computation time. We are confident that such a direct calculation of relevant path properties could still be improved significantly, however, it would probably never beat the performance of the machine learning approach. The latter has two advantages: one, is that machine learning algorithms have more free parameters; the second is that these parameters are all trained automatically without educated guesses which are prone to errors.

## Acknowledgments

PGM and EA wish to thank Dr Alex Zamakhovski for the useful discussions.

## Data availability statement

All data that support the findings of this study are included within the article (and any supplementary files). See [35].

## Appendix A. On the relation between $L$ and the power-law shape of the increment PDF

We conjecture the relation between the exponent  $\beta$ , describing the power-law shape of the increment PDF  $\propto |v|^{-\beta}$  at large  $v$ , where  $2 < \beta < 3$  and the Noah exponent  $L$ , by analogy with results obtained for unidimensional LW (defined in see section 3 in the main text), presented in [36, 37]. These papers discuss the relation between the power-law shape of the *position* PDF of the walker and the temporal scaling of the mean squared displacement  $\langle x^2(t) \rangle$ . They show, that if the PDF has a power-law shape  $\propto |x|^{-\beta}$  at large  $x$  but also  $x < O(t)$  (smaller than the order of magnitude  $O(t)$  proportional to  $t$ ), with some time dependent pre-factor [37], then  $\langle x^2(t) \rangle \propto t^{3-\beta}$ . In our case, the Noah effect creates a situation where as time increases, increasingly larger increments are detected, due to appearance of increasingly large extreme events, hence  $\langle v^2(t) \rangle$  grows quickly with time. Replacing  $x$  with  $v$ , similar arguments as in [36, 37] suggest that the  $\langle v(t) \rangle \propto t^{3-\beta}$ . Comparing this with equation (5), we find that  $t^{3-\beta} = t^{2M+2L-2}$ , which for  $M = 1/2$  implies  $L = 2 - \beta/2$ . To support this result with a ‘sanity check’, note that by definition [15, 38];  $1/2 \leq L \leq 1$ . This is indeed consistent with  $2 < \beta < 3$  (the bounds should be treated separately, we leave this out of the scope of this paper). A more rigorous examination of this conjecture will be beneficial in future work.

## Appendix B. Fluctuation functions

The ensemble averaged TASD carries information of all three exponents  $M, L$  and  $J$

$$\langle \overline{\delta^2} \rangle = \left\langle \frac{1}{t - \Delta} \sum_{t_0=1}^{t-\Delta} [x_n(t_0 + \Delta) - x_n(t_0)]^2 \right\rangle_n \sim t^{2L+2M-2} \Delta^{2J}, \quad (\text{B.1})$$

where the subscript  $n$  denotes different realizations of the process and the brackets  $\langle \rangle_n$  the average over these realizations  $1/n \sum_n$ . The above scaling can be derived mathematically [24]. Usually,  $t$  is fixed as the length of the time series and thus, the TASD is used for calculating  $J$ . Instead of the moving window average, the TASD for one realization (without the ensemble

average) could also be defined for non-overlapping windows

$$\overline{\delta^2} = \frac{1}{t/\Delta} \sum_{k=1}^{t/\Delta} [x(1+k\Delta) - x(1+(k-1)\Delta)]^2 \sim t^{2L+2M-2} \Delta^{2J}, \quad (\text{B.2})$$

where the scaling can be assumed to be unchanged (at least for the expectation values).

An analogous measure can be defined if we use variances of segments with length  $\Delta$  instead of squared displacements of segments of length  $\Delta$  in equation (B.2) (namely, if we also subtract the squared mean of the increments segment, inside  $t$ , the square brackets []). The resulting method for measuring  $J$  is called fluctuation analysis or DFA of order zero

$$F_0^2 = \frac{1}{t/\Delta} \sum_{k=1}^{t/\Delta} \left[ \frac{1}{\Delta} \sum_{i=1}^{\Delta} x(i + (k-1)\Delta) - \frac{1}{\Delta} \sum_{j=1}^{\Delta} x(j + (k-1)\Delta) \right]^2 \sim t^{2L+2M-2} \Delta^{2J}. \quad (\text{B.3})$$

The scaling is plausible due to the analogy to the TASD. The general form of DFA uses variances of residuals of polynomial fits to segments of length  $\Delta$  instead of the pure variance of  $x$ . The squared fluctuation function of DFA of order  $q$  reads

$$F_q^2 = \frac{1}{t/\Delta} \sum_{k=1}^{t/\Delta} \left[ \frac{1}{\Delta} \sum_{i=1}^{\Delta} |x(i + (k-1)\Delta) - p_{q,k}(i)|^2 \right] \sim t^{2L+2M-2} \Delta^{2J}. \quad (\text{B.4})$$

Here,  $p_{q,k}$  is a polynomial fit of order  $q$  to  $x$  in the segment  $k$ . For order  $q = 0$  it is the mean of the segment and we obtain equation (B.3). The method DFA [27] is usually used in order to infer  $J$ , so  $t$  is fixed as the length of the time series. Also, instead of the squared fluctuation function,  $F(\Delta) \sim \Delta^J$  is more common. For a theoretical discussion of the method see [39]. For two or three dimensional vectors  $x = (x_1, x_2, \dots)$ , the equation can still be used. In this case, the polynomial fit is performed in all dimensions ( $p_{q,k} = (p_{q,k,1}, p_{q,k,2}, \dots)$ ), and the variances of the residuals are added up over the dimensions  $(x_1(i) - p_{q,k,1}(i))^2 + (x_2(i) - p_{q,k,2}(i))^2 + \dots$ , i.e. the absolute value  $\|^2$  is interpreted as a scalar product. This is not a standard procedure in the DFA literature, however, it is a straight-forward extension of the method.

DFA has one advantage over the TASD. Looking at the cumulative time series

$$y(t) = \sum_{i=1}^t x(i), \quad (\text{B.5})$$

with superballistic behavior ( $H > 1$ ), the TASD (just like DFA of order zero) will always yield a scaling  $\overline{\delta^2}(\Delta) \sim \Delta^2$ , which is the maximal scaling of the function, because a higher scaling would induce a statistical difference (a Moses effect) among the different windows and therefore lead to a Moses effect ( $t$ -dependence) instead of a Joseph effect ( $\Delta$ -dependence) [40]. In contrast to this, DFA of order one or higher will yield the scaling  $F_1^2(\Delta) \sim \Delta^{2+2J}$ , so it is still useful for inferring  $J$ . This is an advantage, because the cumulative time series will always have a reduced level of noise, so noise can be filtered in an easy way.

The effect of noise is similar for both DFA and TASD. The noise  $x^{(s)}$  dominates the variance of small increments  $x(t_0 + \Delta) - x(t_0)$ , while the signal  $x^{(s)}$  dominates the variance of the full time series. Accordingly, the noise dominates  $F(\Delta)$  for small  $\Delta$  and the signal dominates for large  $\Delta$ . More precisely, for sums  $x^{(s)} + x^{(n)}$  the squared fluctuation function  $F^2(\Delta)$  follows a superposition principle  $F^2(\Delta) = F^{(s)2}(\Delta) + F^{(n)2}(\Delta)$ . This can make fitting the scaling very

hard because the fluctuation function is a superposition of two functions with different scaling exponents. The exact shape depends on the amplitudes (variances) of both noise and signal. This enables direct estimation of the noise level by fitting the short- $\Delta$  behavior with the respective noise scaling (see [41]), or estimating the variance of the pure signal by fitting the long- $\Delta$  behavior. Such piecewise DFA fits were performed e.g. in [42].

### Appendix C. Comparison of methods for inferring $M + L$

As shown in the main text, the previously introduced technique of calculating the Noah and Moses exponents is unbiased and reliable for long time series without noise. However, ANDI included noisy time series of varying lengths. Therefore, the estimator has to be improved. Note that  $M$  and  $L$  are two different exponents with different interpretations, however, for the scaling relation  $H = J + L + M - 1$ , we only need their sum, which can be calculated directly from equation (5). Here we propose two alternative algorithms to deal with noise. One is based on subtracting the effect of the noise, and one is based on a modified fluctuation function. In both algorithms, we employ some exception handling, discussed below, for values outside the range of possible results or if the fit fails. The numerical implementation in python is available on GitHub [35] ([https://github.com/ANDICChallenge/ANDI2020\\_TeamK\\_TSA](https://github.com/ANDICChallenge/ANDI2020_TeamK_TSA)) for the most promising algorithms.

The two approaches are given by the following equations: the first one yields  $M_{dif}$  and  $L_{dif}$ , with subtracted noise level, which can replace the standard  $M, L$  in the analysis above;

$$\sum_{n=1}^t |v(n)|^2 - c^2 \sim t^{2M_{dif}+2L_{dif}-1}, \quad (C.1)$$

where the overline indicates the average of the value over the full time series and  $c$  is the mean noise amplitude estimated from small- $\Delta$  fit to the fluctuation function as described in section appendix B (equivalently, the signal variance can be estimated from the large- $\Delta$  fit and subtracted from the total variance).

The second approach is based on a modified DFA fluctuation function applied to the time series  $x(t)$

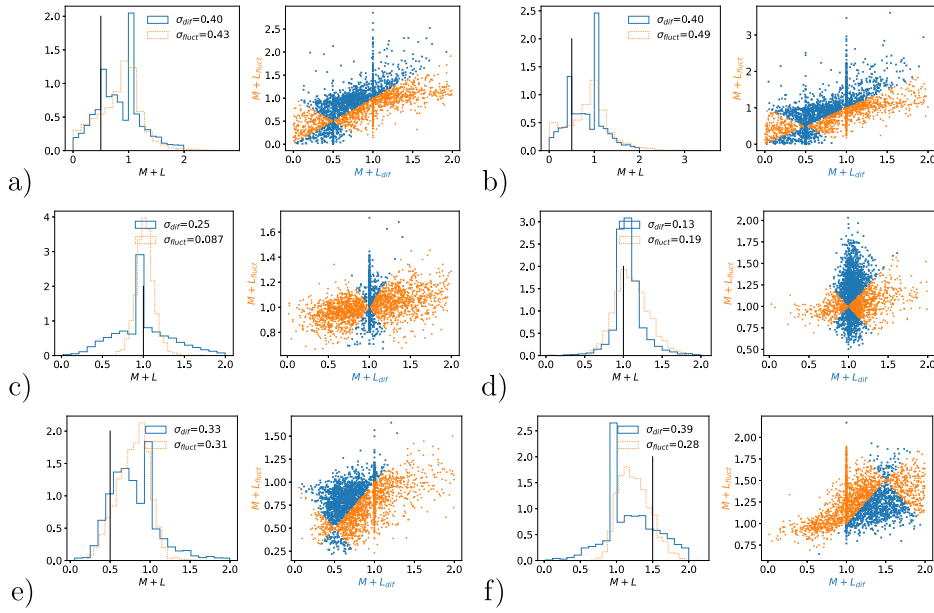
$$\sum_{k=1}^{t/\Delta} \left[ \frac{1}{\Delta} \sum_{i=1}^{\Delta} \left| x(i + (k-1)\Delta) - \frac{1}{\Delta} \sum_{j=1}^{\Delta} x(j + (k-1)\Delta) \right|^2 \right] \sim t^{2L_{fluct}+2M_{fluct}-1}. \quad (C.2)$$

This measure is inspired from equation (B.4) and uses the scaling with  $t$  instead of  $\Delta$ . The window length  $\Delta$  can be chosen freely. We used a window length of  $s = 20$ , which gives reasonably robust results. However, this means that for shorter time series than 20 steps, the measure cannot be used.

We apply the following consistency checks and exception handling: analogously to the calculation of  $M_{dif} + L_{dif}$  in equation (C.3), we can independently calculate  $M_{dif}$  by

$$\sum_{n=1}^t |v(n)| - c \sim t^{M_{dif}+1/2}, \quad (C.3)$$





**Figure C1.** Estimating  $M + L$  in an ensemble of 3000 trajectories of each process with random lengths between 100 and 1000 and noise of random amplitudes: left panels show histograms of estimated exponents; the black line indicates the true value and the labels show the mean absolute error; right panels show results of different algorithms ‘dif’ or ‘fluc’ for the same trajectories; the colors indicate the best performing method; the analysis yields no strong correlation between errors of the methods; (a) ATTM (in the regime  $\sigma < \gamma < \sigma + 1$ ) with  $2H = 0.5$ , (b) CTRW with  $2H = 0.5$ , (c) FBM with  $2H = 0.5$ , (d) LW with  $2H = 1.5$ , (e) SBM with  $2H = 0.5$ , and (f) SBM with  $2H = 1.5$ .

- Since  $L > 1/2$ , always (see table 1), the calculation of  $M$  from equation (6) yields a lower bound  $2M < L + M$ .
- $M + L$  for none of the processes can be  $> 1$  (see table 1), implying superdiffusion, if not  $M > 1/2$  (see table 1). If  $M + L - 1$  and  $M - 1/2$  have opposite signs, we treat the result like a failed fit and use the default value  $M + L = 1$ .
- we also decided (after some trial and error), to use an average of the estimation of  $M + L$  and the estimation of  $M + 1/2$ ; for ATTM and CTRW, this introduces a bias, however, due to the large errors for these processes, this bias had little impact on the performance and the advantage of the average seemed to be more important.
- If the fit does not produce a valid number, i.e. ‘nan’, the result is set to  $M + L = 1$ .
- If  $M + L > 2$ , we set the result to the default  $M + L = 1$ .
- If  $M + L < 0$ , we set the result to  $M + L = 0.5$ , as a negative exponent is more likely to be produced by subdiffusive dynamics in contrast to superdiffusive dynamics.

In figure C1, we show a numerical assessment of the two approaches for all processes with trajectories of random length between 100 steps and 1000 steps (shorter trajectories are neglected). The errors (see figure labels) are large, especially for ATTM, CTRW, and SBM. The estimated values tend to be too close to 0.5 due to our exception handling. The best correlation between true value and estimation is obtained from the first method of subtracting the estimated noise level. The methods have advantages and disadvantages for different processes and the errors are not strongly correlated (i.e. if one method overestimates the exponent, the

other methods might as well underestimate it). Therefore, an average of the two estimators for  $M + L$  seems to be the most promising way of calculating the exponent with a reasonable performance.

## Appendix D. Comparison of methods for inferring $J$

Similar to the calculation of  $M + L$ , the previously introduced method for calculating the parameter  $J$  (equation (2)) also performs less well for noisy data if the path length is short. We present two algorithms based on two strategies for noise reduction: cumulative and moving averages. In both algorithms, we employ the same exception handling for values outside the range of possible results  $0 < J < 1$  or if the fit fails. The numerical implementation in python is available on GitHub ([https://github.com/ANDICChallenge/ANDI2020\\_TeamK\\_TSA](https://github.com/ANDICChallenge/ANDI2020_TeamK_TSA)) [35].

Both approaches are based on the DFA fluctuation function, which is a measure of  $J$  and can be used as an alternative to the TASD [24]. The first algorithm uses the cumulative trajectory  $y(t) = \sum_{n=1}^t x(n)$ , in order to be less affected by noise. Here,

$$F_{\text{tra}}^2(\Delta) = \frac{1}{t/\Delta} \sum_{k=1}^{t/\Delta} \left[ \frac{1}{\Delta} \sum_{i=1}^{\Delta} |y(i + (k-1)\Delta) - p_{1,k}(i)|^2 \right] \sim \Delta^{2J_{\text{tra}}+2}, \quad (\text{D.1})$$

where  $p_{1,k}(t)$  are linear fits to the respective segments. This linear detrending enables measurement of exponents  $> 1$ . Analogous to equation (C.3), when calculating the scaling behavior, an estimated noise level can be subtracted (estimated from the short  $\Delta$  behavior).

The second algorithm uses the cumulative 20-steps moving average  $U(t) = \sum_{n=1}^t u(n)$  of the increments  $v$ . Here we use DFA of order 0, without detrending, and with subtracting the window average instead of a linear fit (detrending order 0). The equations reads

$$F_{\text{ma}}^2(\Delta) = \frac{1}{t/\Delta} \sum_{k=1}^{t/\Delta} \left[ \frac{1}{\Delta} \sum_{i=1}^{\Delta} \left| U(i + (k-1)\Delta) - \frac{1}{\Delta} \sum_{i=1}^{\Delta} U(i + (k-1)\Delta) \right|^2 \right] \sim \Delta^{2J_{\text{ma}}}. \quad (\text{D.2})$$

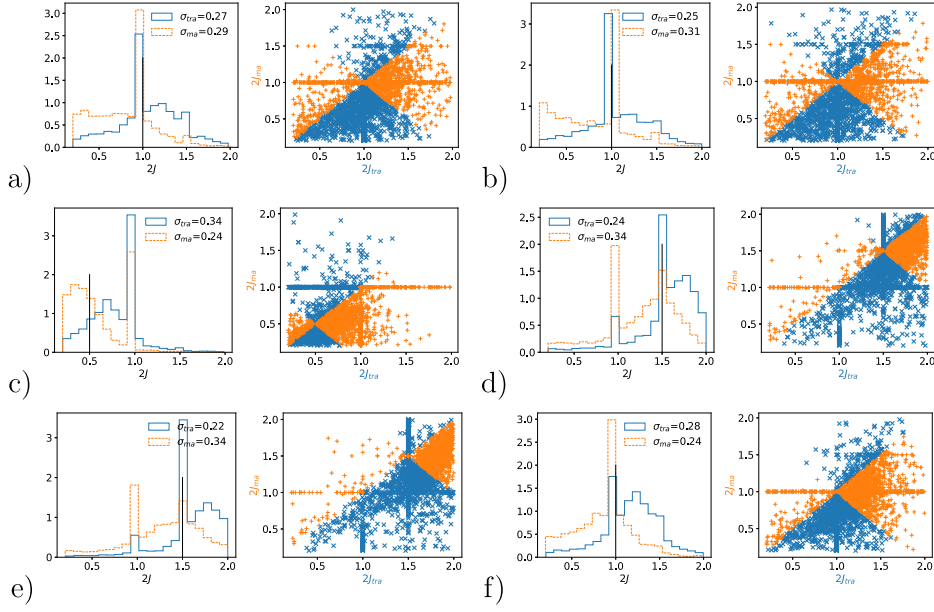
We apply the following exception handling:

- If the result of the fit is invalid or yields an exponent of zero or negative,  $J$  is set to 0.5
- If the fit yields values of  $J \geq 1$ , the exponent is set to  $J = 0.75$ , i.e. the average exponent for superdiffusive processes, which are likely to produce large  $J$

Figure D1 shows the performance of both algorithms for different processes. The errors are smaller compared to  $M + L$  as the left panels show. The first algorithm using the cumulative signal performs slightly better than the moving averages. The right panels indicate, that the errors are not correlated. For this reason, averaging both estimators  $J_{\text{tra}}$  and  $J_{\text{ma}}$  leads to the best performance.

## Appendix E. Technical details of the process inference method

In this section, we provide the technical details of the process inference method we used for ANDI. For the general idea, see section 5 in the main text. Recall that the answer to each of the questions below can be either yes/no, or maybe, to which we assign the values 1, 0 and 2, respectively.



**Figure D1.** Estimating  $J$  in an ensemble of 3000 trajectories of each process with random lengths between 100 and 1000 and random noise amplitudes: left panels show histograms of estimated exponents; the black line indicates the true value and the labels show the mean absolute error; right panels show results of different algorithms ‘tra’ or ‘ma’ for the same trajectories; the colors indicate the best performing method; the analysis yields no strong correlation between errors of the methods; (a) ATTM (in the regime  $\sigma < \gamma < \sigma + 1$ ) with  $2H = 0.5$ , (b) CTRW with  $2H = 0.5$ , (c) FBM with  $2H = 0.5$ , (d) FBM with  $2H = 1.5$ , (e) LW with  $2H = 1.5$ , and (f) SBM with  $2H = 1.5$ .

- Regarding the first and second questions, in section 5. We found that to tackle errors in the numerical estimations of the exponents from the time series, it might be beneficial to ask some questions with similar nature, in a different variation, in order to re-enforce the distinction pointed out by the first. For example, in addition to the  $J$ -centered questions presented in the main text, we also ask about the dominance of the Moses versus the Joseph effect and Noah effect, which is designed to distinguish SBM from the other models.
- Regarding the third question, in the main text. ‘Is the relation between  $M$  and  $L$  consistent with an infinite-density function, or is it closer to trivial scaling expected when all the moments of the velocity scale in the same way—like in SBM for example?’ Explanation: when there is no Noah effect,  $\sum_{n=1}^t |v_n| / (\sqrt{t} \sum_{n=1}^t v_n^2) = \text{constant}$ . The Noah effect breaks this stationarity. More specifically, in CTRW the Noah and Moses effects are coupled, and  $L$  and  $M$  satisfy equation (7) because the velocity PDF approaches an infinite-invariant density (see section 3.3). In this case, from equations (5)–(7), we find that  $\sum_{n=1}^t |v_n| / (\sum_{n=1}^t v_n^2) = \text{constant}$ . In order to test which of the above scalings holds, we compare the variance over time of the (square root of the) first expression, and the second expression over time, with some pre-factor (chosen to be 0.6). If the dynamics are well described by an infinite density, the second expression has a smaller variance, because its expectation value is constant, while the first expression has a larger variance, because it is not bounded. For processes, where  $L$  is close to  $1/2$ , it is the other way round.

- Regarding the fourth question. ‘only in LW with noise, when you take the absolute value of the process and subtract its mean—the velocity correlations vanish entirely’. Here we therefore ask (in one dimension): converting  $x(t)$  to  $y(t) = |x(t)| - \langle |x(t)| \rangle$ ; is  $J = 1/2$  in the new process  $y(t)$ , whereas for  $x(t)$  this exponent was larger than  $1/2$ ? (And in  $d > 1$  dimensions we use the increments of the process in the radial direction). Explanation: first, note that in the superdiffusive LW model from ANDI;  $J > 1/2$ , see table 1. As explained in the introduction, in ANDI all the paths are superimposed with uncorrelated random fluctuations, which are selected from a narrow (non-fat-tailed) probability distribution. Since in the process  $y(t)$  we measure only the correlations of the noise itself; here  $J = 1/2$ . This question distinguishes LW from all the other processes, and as a side-note, according to our numerical analysis, it was the most successful question of all (see table E1, under the title ‘ $|v|$  correlated?’).
- Regarding the ‘fat-tailness’ question, and ‘plateau detection’. In the first test, we build upon the knowledge that if the increments of the process are taken from a fat-tailed PDF with no typical scale, no matter how long the process is—it will almost always be characterised by a single big fluctuation which may be comparable to the combined size of all the others. The path, in this case, resembles the ‘devil’s staircase’ [22], which characterizes e.g., trajectories generated by Lévy flight. In ANDI, none of the processes is expected to show this extreme shape, however, ATTM and LW, which have a Noah effect, are expected still to exhibit very few, but extremely large fluctuations which therefore means that the median of the increments is expected to be far from the mean of the largest and smallest value. We thus define a small threshold value  $a = 0.2$ , and ask: ‘is the median( $|v|$ )/ ( $[\max(|v|) + \min(|v|)] / 2$ )  $< a$ ’? Finally, the ‘flatness’ test is designed to detect CTRW: defining another small threshold value, to account for fluctuations generated due to the external noise superimposed with the paths; we construct a loop which seeks for the biggest jumps in the data, and if all other jumps between them are small, we consider it as a ‘flat’ interval, namely a plateau. If e.g., over 90% of the time-series is flat, the test will give an answer 1, hinting towards CTRW. To strengthen the results of this question, we constructed a second flatness test using an upper threshold on the variance of the jumps during the intervals. In questions vi and vii, the answer is ‘maybe’ ( $= 2$ ), if the thresholds for a strict answer were not crossed, but the results were sufficiently close. This means that this case requires a second set of arbitrary intermediate threshold values.

*Using the questions to construct the detection algorithm.* Using numerical simulations, we generated 50 000 trajectories, for each of the processes above, with 10 different length categories. For each trajectory in the simulated data set, we generated a set of answers for the questions above, for example: [0, 1, 0, 2, . . .], if the process returned ‘no’, for the first question, ‘yes’ for the second, ‘no’ for the third followed by ‘maybe’ and so on. Upon receiving a new *unknown* time-series to analyse, from the ANDI database, we generated a set of answers for this path in the same way as for each of the simulated paths, and then, using our pre-generated archive we obtained the relative occurrence of this vector of answers in the various models. This generates an array of probabilities for this new process to be either ATTM, CTRW, FBM, LW, or SBM. The larger the simulated ‘training’ set, the more accurate the evaluation of the probabilities is. If the specific set of answers for the unfamiliar time series was not found in the training file, we select a subset of the original questions again and define the probabilities from them only. This makes the choice less selective. If again the set of answers is not found, we return a default value: probability = 0.2 that this process is FBM, 0.3 for CTRW, since the properties of this process are most difficult to obtain numerically, 0.1 for LW since we had a

**Table E1.** Average probability of obtaining each of the answers ‘0’, ‘1’ and ‘2’, for each question in the process inference, calculated from the simulation results in the ‘training set’ [35] used also in figure 2. This table demonstrates the selectivity of the various questions, for different models, and the motivation for choosing the options ‘yes/no’ and ‘maybe’ as possible answers, see section 5 and appendix E. Note, for example, that CTRW paths always returned the answer 1 (=‘yes’) on the question of flatness, whereas LW always returned 0 (=‘no’), and ATTM returned both with roughly equal probability. Hence, through the yes/no answers this question is highly efficient for distinguishing between the first two. On the other hand, on the issue of  $|v|$  correlations; ATTM usually replies 0 (with probability  $\approx 0.9$ ), and LW always returns 1. Due to its design, the two ‘plateau’ questions mostly returned ‘maybe’ (answer cannot be determined) for LW, whereas this answer was more common, e.g., in the question ‘ $J$  strongest?’ in ATTM and  $J > 1/2$  in ATTM and CTRW, which call for reliable numerical exponent measurements for a yes/no answer. Hence here, ‘maybe’ may be used to distinguish these processes from each other. Our numerical algorithm performs this analysis automatically, in a way that is reminiscent of data-driven algorithms, though the questions were specifically designed for their respective purposes based on theory. Other distinguishable features may be observed which are not currently explicitly pointed out in the example.

Process	‘Fat-tailness’	Plateaus, method1	Plateaus, method2	$J$ strongest effect?	$J > 1/2$	Infinite-density?	$ v $ correlated?	$M$ dominant effect?
ATTM,	0 0.42	0.64	0.45	0.37	0.26	0.8	0.91	0.83
	1 0.50	0.29	0.47	0.25	0.21	0.2	0.03	0.17
	2 0.08	0.07	0.07	0.38	0.53	0	0.07	0
CTRW,	0 0	0.30	0.34	0.80	0.21	0.31	0.83	0.94
	1 1	0.54	0.49	0.07	0.21	0.52	0	0.06
	2 0	0.17	0.17	0.13	0.58	0.17	0.17	0
FBM,	0 1	0.71	0.32	0.04	0.31	1	0.57	0.96
	1 0	0.03	0.42	0.72	0.40	0	0.20	0.04
	2 0	0.26	0.26	0.24	0.28	0	0.22	0
LW,	0 1	0	0	0.02	0.01	0.10	0	1
	1 0	0	0	0.88	0.81	0.90	1	0
	2 0	1	1	0.10	0.18	0	0	0
SBM,	0 0.89	0.80	0.48	0.37	0.19	0.64	0.91	0.60
	1 0.05	0.07	0.39	0.19	0.22	0.36	0.02	0.40
	2 0.06	0.13	0.13	0.44	0.59	0	0.07	0

question that could detect it very well, and 0.2 for SBM and ATTM (other options are also possible, e.g., simply choosing equal probability for all the models).

Table E1 shows an example containing the probabilities for all the different answers (0, 1 and 2) generated for the five models as part of the ‘training set’ [35], for paths of = 1000 steps. This example demonstrates the ability of the three-answers method to generate distinguishable patterns that can be used for the process classification. One way by which this method can be improved in future work can be by using designated statistical tools to more accurately determine the statistical significance of the results.

## Appendix F. On the link between CTRW and the infinite-density function

In the unidimensional process, the total displacement  $x(t) = \sum_{i=1}^N \xi_i$ , where the  $\xi_i$ s are the random step displacements, chosen in ANDI from a Gaussian distribution. In  $d = 2$  (and  $d = 3$ ) dimensions [6], after each  $\tau_i$  the particle chooses three random step-lengths from an identical

Gaussian distribution, representing the current displacement along the  $x$  and  $y$  (and  $z$ ) axes, and its total displacement along the radial line is  $\Delta r_i = \sqrt{\xi_{i,x}^2 + \xi_{i,y}^2 + \xi_{i,z}^2}$ . If we divide the particles path  $x(t)$  (or  $r(t)$ ) into increments of some small arbitrary duration  $\Delta$ , we find that the mean increment-velocity  $v$  (the displacement during the increment, over the duration  $\Delta$ ) has a fat-tailed distribution, which is a particular case of the generalized model studied in [24, 32, 33].

By direct calculation, following the references above, we find that;  $\sum_{n=1}^t |v(n)| \sim t^{M+1/2}$ , where  $M = 2H - 1/2$ ,  $\sum_{n=1}^t |v(n)| \sim t^{2M+2L-1}$ , with  $L = 1 - H$ . Both the absolute and the second moments of the velocity can be obtained from a time-invariant function [24]

$$\lim_{t \rightarrow \infty} t^\alpha P_t(v) \rightarrow \mathcal{I}(v), \quad \text{and} \quad L = \alpha/2 + 1/2, \quad M = -\alpha + 1/2. \quad (\text{F.1})$$

The limit function  $W(v)$  leads to the means of  $|v|$  and  $v^2$  via

$$\langle |v|^q \rangle = 2 \int_0^\infty dv |v|^q P_t(v) \underset{t \gg 1}{\approx} 2t^{-\alpha} \int_0^\infty dv |v|^q \mathcal{I}(v), \quad (\text{F.2})$$

with  $q = 1, 2$ , and since by definition  $1/2 < L < 1$  [24], we have  $0 < \alpha < 1$ . In this case, first notice that both the first and the second moments scale similarly in time  $\propto t^{-\alpha}$ , whereas in normal diffusion processes  $\langle v^2(t) \rangle \propto (\langle |v| \rangle)^2$ . Second, since  $0 < \alpha < 1$ , even though  $\mathcal{I}(v)$  is an invariant function to which the properly normalized velocity PDF converges in the long times limit; this function itself is not normalized, since  $\int_{-\infty}^\infty \mathcal{I}(v) dv = \lim_{t \rightarrow \infty} t^\alpha \rightarrow \infty$ . For this reason,  $\mathcal{I}(v)$  is called an *infinite invariant density* function, which has been well established as a mathematical theory for several decades, and in recent years discovered also in physical systems, e.g. [33, 43–49]. Finding the relation equation (7) in data is an indication that the increment process might be described by an infinite-invariant density, and hence might be related to CTRW (or other models that have similar statistical properties), as opposed to, e.g., FBM or SBM.

## ORCID iDs

Philipp G Meyer  <https://orcid.org/0000-0002-8341-2419>

Erez Aghion  <https://orcid.org/0000-0002-5591-8084>

Holger Kantz  <https://orcid.org/0000-0001-6921-6094>

## References

- [1] Metzler R, Jeon J-H, Cherstvy A G and Barkai E 2014 Anomalous diffusion models and their properties: non-stationarity, non-ergodicity, and ageing at the centenary of single particle tracking *Phys. Chem. Chem. Phys.* **16** 24128–64
- [2] Méndez V, Campos D and Bartumeus F 2016 *Stochastic Foundations in Movement Ecology* (Berlin: Springer)
- [3] Vilk O et al 2021 Unravelling the origins of anomalous diffusion: from molecules to migrating storks (arXiv:2109.04309)
- [4] Stojkoski V, Sandev T, Basnarkov L, Kocarev L and Metzler R 2020 Generalised geometric Brownian motion: theory and applications to option pricing *Entropy* **22** 1432
- [5] Mejía-Monasterio C, Metzler R and Vollmer J 2020 Editorial: anomalous transport: applications, mathematical perspectives, and big data *Front. Phys.* **8** 622417



- [6] Muñoz-Gil G *et al* 2021 Objective comparison of methods to decode anomalous diffusion *Nat. Commun.* **12** 6253
- [7] Muñoz-Gil G, Volpe G, Garcia-March M A, Metzler R, Lewenstein M and Manzo C 2020 AnDi: the anomalous diffusion challenge (arXiv:2003.12036)
- [8] Huybers P and Curry W 2006 Links between annual, Milankovitch and continuum temperature variability *Nature* **441** 329–32
- [9] Polotzek K and Kantz H 2020 An ARFIMA-based model for daily precipitation amounts with direct access to fluctuations *Stoch. Environ. Res. Risk Assess.* **34** 1487–505
- [10] Manzo C 2021 Extreme learning machine for the characterization of anomalous diffusion from single trajectories (AnDi-ELM) *J. Phys. A: Math. Theor.* **54** 334002
- [11] Gentili A and Volpe G 2021 Characterization of anomalous diffusion classical statistics powered by deep learning (CONDOR) *J. Phys. A: Math. Theor.* **54** 314003
- [12] 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
- [13] 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
- [14] Metzler R 2019 Brownian motion and beyond: first-passage, power spectrum, non-Gaussianity, and anomalous diffusion *J. Stat. Mech.* **114003**
- [15] Chen L, Bassler K E, McCauley J L and Gunaratne G H 2017 Anomalous scaling of stochastic processes and the Moses effect *Phys. Rev. E* **95** 042141
- [16] Mandelbrot B 1963 The variation of certain speculative prices *J. Bus.* **36** 394–419
- [17] Mandelbrot B B and Van Ness J W 1968 Fractional Brownian motions, fractional noises and applications *SIAM Rev.* **10** 422–37
- [18] Lim S C and Muniandy S V 2002 Self-similar Gaussian processes for modeling anomalous diffusion *Phys. Rev. E* **66** 021114
- [19] Harvey S and Montroll E W 1975 Anomalous transit-time dispersion in amorphous solids *Phys. Rev. B* **12** 2455
- [20] Klafter J and Zumofen G 1994 Lévy statistics in a Hamiltonian system *Phys. Rev. E* **49** 4873
- [21] Massignan P, Manzo C, Torreno-Pina J A, García-Parajo M F, Lewenstein M and Lapeyre G J Jr 2014 Nonergodic subdiffusion from Brownian motion in an inhomogeneous medium *Phys. Rev. Lett.* **112** 150603
- [22] Klafter J and Sokolov I M 2011 *First Steps in Random Walks: From Tools to Applications* (Oxford: Oxford University Press)
- [23] Meyer P G, Adlakha V, Kantz H and Bassler K E 2018 Anomalous diffusion and the Moses effect in an aging deterministic model *New J. Phys.* **20** 113033
- [24] 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
- [25] Zamani M, Aghion E, Pollner P, Vicsek T and Kantz H 2021 Anomalous diffusion in the citation time series of scientific publications *J. Phys. Complex.* **2** 035024
- [26] Meyer P, Barkai E and Kantz H 2017 Scale-invariant Green–Kubo relation for time-averaged diffusivity *Phys. Rev. E* **96** 062122
- [27] Peng C-K, Buldyrev S V, Havlin S, Simons M, Stanley H E and Goldberger A L 1994 Mosaic organization of DNA nucleotides *Phys. Rev. E* **49** 1685
- [28] Fromberg D, Schmiedeberg M, Barkai E and Zaboradaev V 2015 Asymptotic densities of ballistic Lévy walks *Phys. Rev. E* **91** 022131
- [29] Muñoz-Gil G, Garcia-March M A, Manzo C, Martín-Guerrero J D and Lewenstein M 2019 Machine learning method for single trajectory characterization (arXiv:1903.02850)
- [30] Zaboradaev V, Denisov S and Klafter J 2015 Lévy walks *Rev. Mod. Phys.* **87** 483–530
- [31] Aghion E, Kessler D A and Barkai E 2018 Asymptotic densities from the modified Montroll–Weiss equation for coupled CTRWS *Eur. Phys. J. B* **91** 17
- [32] Albers T and Radons G 2018 Exact results for the nonergodicity of  $d$ -dimensional generalized Lévy walks *Phys. Rev. Lett.* **120** 104501
- [33] Akimoto T, Barkai E and Radons G 2020 Infinite invariant density in a semi-Markov process with continuous state variables *Phys. Rev. E* **101** 052112
- [34] Kantelhardt J W, Koscielny-Bunde E, Rego H H A, Havlin S and Bunde A 2001 Detecting long-range correlations with detrended fluctuation analysis *Physica A* **295** 441–54
- [35] Aghion E 2021 [https://github.com/ANDICChallenge/ANDI2020\\_TeamK\\_TSA](https://github.com/ANDICChallenge/ANDI2020_TeamK_TSA)



- [36] Rebenshtok A, Denisov S, Hänggi P and Barkai E 2014 Non-normalizable densities in strong anomalous diffusion: beyond the central limit theorem *Phys. Rev. Lett.* **112** 110601
- [37] Rebenshtok A, Denisov S, Hänggi P and Barkai E 2014 Infinite densities for Lévy walks *Phys. Rev. E* **90** 062135
- [38] Mandelbrot B B and Wallis J R 1968 Noah, Joseph, and operational hydrology *Water Resour. Res.* **4** 909–18
- [39] Höll M, Kiyono K and Kantz H 2019 Theoretical foundation of detrending methods for fluctuation analysis such as detrended fluctuation analysis and detrending moving average *Phys. Rev. E* **99** 033305
- [40] Höll M, Kantz H and Zhou Y 2016 Detrended fluctuation analysis and the difference between external drifts and intrinsic diffusionlike nonstationarity *Phys. Rev. E* **94** 042201
- [41] Meyer P G and Kantz H 2019 Inferring characteristic timescales from the effect of autoregressive dynamics on detrended fluctuation analysis *New J. Phys.* **21** 033022
- [42] Meyer P G and Kantz H 2019 A simple decomposition of European temperature variability capturing the variance from days to a decade *Clim. Dyn.* **53** 6909–17
- [43] Aaronson J 1997 *An Introduction to Infinite Ergodic Theory* (Providence, RI: American Mathematical Society)
- [44] Akimoto T and Miyaguchi T 2010 Role of infinite invariant measure in deterministic subdiffusion *Phys. Rev. E* **82** 030102
- [45] Korabel N and Barkai E 2012 Infinite invariant density determines statistics of time averages for weak chaos *Phys. Rev. Lett.* **108** 060604
- [46] Meyer P and Kantz H 2017 Infinite invariant densities due to intermittency in a nonlinear oscillator *Phys. Rev. E* **96** 022217
- [47] Zhou T, Xu P and Deng W 2019 Continuous time random walks and Lévy walks with stochastic resetting (arXiv:1909.07213)
- [48] Aghion E, Kessler D A and Barkai E 2019 From non-normalizable Boltzmann–Gibbs statistics to infinite-ergodic theory *Phys. Rev. Lett.* **122** 010601
- [49] Barkai E, Radons G and Akimoto T 2021 Transitions in the ergodicity of subrecoil-laser-cooled gases (arXiv:2104.03816)