Journal of Physics A: Mathematical and Theoretical

PREFACE

Preface: characterisation of physical processes from anomalous diffusion data

To cite this article: Carlo Manzo *et al* 2023 *J. Phys. A: Math. Theor.* **56** 010401

IOP Publishing Journal of Physics A: Mathematical and Theoretical

J. Phys. A: Math. Theor. **56** (2023) 010401 (6pp) https://doi.org/10.1088/1751-8121/acb1e1

Preface

Preface: characterisation of physical processes from anomalous diffusion data

Carlo Manzo[1,*], Gorka Muñoz-Gil[2], Giovanni Volpe[3], Miguel Angel Garcia-March[4], Maciej Lewenstein[5,6] and Ralf Metzler[7,8,*]

[1] Facultat de Ciències, Tecnologia i Enginyeries, Universitat de Vic—Universitat Central de Catalunya (UVic-UCC), C. de la Laura, 13, 08500 Vic, Spain

[2] Institute for Theoretical Physics, University of Innsbruck, Technikerstr. 21a, A-6020 Innsbruck, Austria

[3] Department of Physics, University of Gothenburg, Origovägen 6B, SE-41296 Gothenburg, Sweden

[4] Instituto Universitario de Matemática Pura y Aplicada, Universitat Politècnica de València, València, Spain

[5] ICFO—Institut de Ciències Fotòniques, The Barcelona Institute of Science and Technology, Av. Carl Friedrich Gauss 3, 08860 Castelldefels (Barcelona), Spain

[6] ICREA, Pg. Lluís Companys 23, 08010 Barcelona, Spain

[7] Institute of Physics & Astronomy, University of Potsdam, Karl-Liebknecht-St. 24/25, D-14476 Potsdam-Golm, Germany

[8] Asia Pacific Centre for Theoretical Physics, Pohang 37673, Republic of Korea

E-mail: carlo.manzo@uvic.cat and rmetzler@uni-potsdam.de

Received 30 December 2022
Accepted for publication 10 January 2023
Published 19 January 2023

CrossMark

Keywords: anomalous diffusion, AnDi challenge, machine learning

1. Introduction

Since Albert Einstein provided a theoretical foundation [1] for Robert Brown's observation of the movement of microscopic granules contained in pollen grains [2], significant deviations from the laws of Brownian motion have been uncovered in an impressively wide variety of animate and inanimate systems, from biology to the stock market. Anomalous diffusion, as it has come to be called, extends the concept of Brownian motion and is connected to disordered systems, non-equilibrium phenomena, flows of energy and information, and transport in living

[*] Authors to whom any correspondence should be addressed.

systems [3]. Anomalous diffusion is 'non-universal' in the sense that physically very different systems share the same power-law form of the mean squared displacement $\langle x^2(t)\rangle \sim t^\alpha$. To properly understand a system exhibiting anomalous diffusion, it is therefore important to have reliable analysis methods to unveil the exact physical mechanisms effecting the observed anomalous diffusion dynamics.

Several methods for detecting the occurrence of and the mechanisms behind anomalous diffusion have been developed using classical statistics [4–10]. However, in the last years, the booming of machine learning has boosted the development of data-driven methods to characterise anomalous diffusion from single trajectories, providing more refined tools for this problem [11–15].

In 2020, we launched the Anomalous Diffusion (AnDi) challenge to provide the first assessment of classical and novel methods for quantifying anomalous diffusion in various realistic conditions through a community-based effort [16]. The challenge consisted of an open competition to benchmark existing methods and spur the invention of new approaches. The AnDi challenge brought together a vibrating and multidisciplinary community of scientists working on this problem, involving more than 30 participants from 22 institutions and 11 countries. Ultimately, the analysis of the results obtained on a reference dataset [17] provided an objective assessment of the performance of methods to characterise anomalous diffusion from single trajectories for three specific tasks, including anomalous diffusion exponent inference, model classification, and trajectory segmentation. The study, published in Nature Communications, analyses the results of the community effort and determines that machine learning greatly improves the estimation of the properties of diffusing particles [18].

This special issue includes the details of several of the methods that participated in the AnDi challenge. Some of the articles describe updated versions of the software originally used for the challenge, showing improved performance. Most of these methods rely on state-of-the-art machine learning approaches. For instance, Gentili and Volpe [19] combine feature engineering based on classical statistics with feed-forward neural networks. Interestingly, this work shows how to create an adapted pipeline specific to the dataset of the challenge to reach some of the best performance of the competition across all tasks. Similarly, Kowalek *et al* [20] also uses a set of statistical features as the input to an extreme gradient boosting model, which then takes care of classifying trajectories among diffusion models. The authors further show that the proper choice of features heavily affects classification accuracy.

Among the different machine learning approaches, recurrent neural networks (RNNs) have attracted a lot of interest due to their suitability when dealing with data with temporal information and long-range correlations. Three works show different implementations based on RNN for the challenge tasks [21–23]. Garibo-i-Orts *et al* [21] combines a bidirectional long short-term memory (LSTM), a state-of-the-art RNN, with a convolutional neural network (CNN). The CNN is used as a feature extractor before a stack of LSTM layers to boost the RNN performance, achieving outstanding results for the regression of the anomalous exponent. Argun *et al* [22] proposes an architecture where trajectories are directly fed to LSTM layers, an implementation that enables the analysis of time traces of arbitrary size, without the need for any padding or preprocessing. This method shows top performance across several tasks of the AnDi challenge, demonstrating that similar architectures can be successfully used for different purposes. Last, Li *et al* [23] describes the use of one of the most promising RNN architectures, the WaveNet. This work further shows how the size of the training dataset (one order of magnitude larger than the rest of the models used in the challenge) can be key to enhancing the method's accuracy.

Other machine learning approaches have also shown remarkable performance in dealing with stochastic diffusion. Manzo [24] attempts to establish a baseline for machine learning approaches using an extreme learning machine, a fast-converging training algorithm for single hidden layer feedforward neural network applied over a set of statistical features. Al-hada *et al* [25] uses a pretrained CNN, the ResNet-50, to classify different stochastic processes and compare it with other CNNs. Conejero *et al* [26] proposes a new architecture, the Convolutional Transformer, to extract features from trajectories and feed them to two transformer encoding blocks that perform either regression or classification. Verdier *et al* [27] presents a method based on graph neural networks (GNNs) where a vector of features is associated with each trajectory position and a sparse graph structure with each trajectory. Similar to [23], the authors use representation learning techniques to study the latent space features of their model and propose a visual exploratory method to analyse trajectories from walks never seen by the GNN. In fact, the unsupervised analysis of diffusion models is a promising tool to characterise unknown datasets that could even lead to the identification of new mechanisms. Along this line, Muñoz-Gil *et al* [28] studies the suitability of auto-encoders as feature extractors for anomalous diffusion trajectories and proposes a method to characterise them using anomaly detection.

Besides machine learning approaches, theory-based methods were also proposed for characterising anomalous diffusion and tested in the AnDi challenge. Meyer *et al* [29] discusses numerical methods to obtain the anomalous diffusion exponent and proposed a questionnaire for model selection based on feature analysis. Bayesian inference was instead used in [30] to distinguish between scaled and fractional Brownian motion and in [31] that presents an approach to deal with Lévy walk trajectories. Bayesian methods are particularly effective when enough information is known about the trajectories and specific priors associated with the type of walk one aims to characterise can be constructed.

The special issue also hosts several theoretical contributions pushing forward the field of stochastic processes and/or discussing applications to time series. Thus, Vitali *et al* [32] discusses emerging transient anomalous diffusion in Markovian hopping-trap scenarios. Transient anomalous diffusion is also obtained in a tempered fractionally integrated process [33]. Maraj *et al* [34] introduces the empirical anomaly measure as a means to measure the distance between the anomalous diffusion process and normal diffusion. Limit properties of Lévy walks are shown to be useful in the recognition and verification of Lévy walk-type motion, as well as the parameter estimation in maximum likelihood methods [35]. Wang *et al* [36] studies the emerging residual nonergodicity in fractional Brownian motion with random diffusivity that may help distinguish and categorise certain nonergodic and non-Gaussian features of particle displacements. Applications of single-trajectory power spectral methods to movement data of kites and storks are discussed in [37], demonstrating how stochastic models can be extracted with this method.

Quantum walks are considered in [38], showing how the interplay between quantum coherence and the mean squared displacement of the walker can provide information on the process. Ablowitz *et al* [39] studies applications of the inverse scattering transform to fractional versions of non-linear equations of, e.g. the Korteweg–deVries equation, that provide a framework for solitonic solutions with power-law dispersion relations. Initial strong non-Gaussianity concurrent with Brownian scaling of the mean squared displacement is reported for self-avoiding random walks in [40].

The research reported in this special issue provides a major contribution toward the understanding of anomalous diffusion processes and their analysis. In particular, a palette of tools is introduced, which are poised to become standard methods for the analysis of trajectories generated from various experiments, from atomic physics to ecology. Moreover, the outcome of these studies reinforces the importance of community-based efforts in the search for the advancement of science. The success of this initiative triggered us to organise the 2nd AnDi challenge around the problem of detecting changes in transport properties and interactions between moving objects from single trajectories.

Data availability statement

No new data were created or analysed in this study.

Acknowledgments

We thank all the authors for their contributions, the editorial staff of the journal for their support, and the participants of the AnDi challenge for their involvement.

C M acknowledges support through the Grant RYC-2015-17896 funded by MCIN/AEI/10.13039/501100011033 and 'ESF Investing in your future', Grants BFU2017-85693-R and PID2021-125386NB-I00 funded by MCIN/AEI/10.13039/501100011033/ and 'ERDF A way of making Europe'. G M G acknowledges support from the European Union (ERC, QuantAI, Project No. 10105529) and the Austrian Science Fund (FWF) through the SFB BeyondC F7102. G V acknowledges funding from ERC StG ComplexSwimmers (Grant No. 677511) and from the Knut and Alice Wallenberg Foundation. M A G M acknowledges funding from the Spanish Ministry of Education and Professional Training (MEFP) through the Beatriz Galindo program 2018 (BEAGAL18/00203), Spanish Ministry MINECO (FIDEUA PID2019-106901GBI00/10.13039/501100011033), QuantERA II Cofund 2021 PCI2022-133004, Project of MCIN with funding from European Union NextGenerationEU (PRTR-C17.I1) and by Generalitat Valenciana, with reference 20220883 (PerovsQuTe). M L acknowledge support from: ERC AdG NOQIA; Ministerio de Ciencia y Innovación Agencia Estatal de Investigaciones (PGC2018-097027-B-I00/10.13039/501100011033, CEX2019-000910-S/10.13039/501100011033, Plan National FIDEUA PID2019-106901GB-I00, FPI, QUANTERA MAQS PCI2019-111828-2, QUANTERA DYNAMITE PCI2022-132919, Proyectos de I+D+I 'Retos Colaboración' QUSPIN RTC2019-007196-7); MICIIN with funding from European Union NextGenerationEU (PRTR-C17.I1) and by Generalitat de Catalunya; Fundació Cellex; Fundació Mir-Puig; Generalitat de Catalunya (European Social Fund FEDER and CERCA program, AGAUR Grant No. 2017 SGR 134, QuantumCAT 16-011424, co-funded by ERDF Operational Program of Catalonia 2014-2020); Barcelona Supercomputing Center MareNostrum (FI-2022-1-0042); EU Horizon 2020 FET-OPEN OPTOlogic (Grant No. 899794); EU Horizon Europe Program (Grant Agreement 101080086—NeQST), National Science Centre, Poland (Symfonia Grant No. 2016/20/W/ST4/00314); ICFO Internal 'QuantumGaudi' project; European Union's Horizon 2020 research and innovation program under the Marie-Skłodowska-Curie Grant Agreement Nos. 101029393 (STREDCH) and 847648 ('La Caixa' Junior Leaders fellowships ID100010434: LCF/BQ/PI19/11690013, LCF/BQ/PI20/11760031, LCF/BQ/PR20/11770012, LCF/BQ/PR21/11840013). Views and opinions expressed in this work are, however, those of the author(s) only and do not necessarily reflect those of the European Union, European Climate, Infrastructure and Environment Executive Agency (CINEA), nor any other granting authority. Neither the European Union nor any granting authority can be held responsible for them. R M acknowledges the German Science Foundation (DFG Grant ME 1535/12-1) for funding.

ORCID iDs

Carlo Manzo https://orcid.org/0000-0002-8625-0996
Gorka Muñoz-Gil https://orcid.org/0000-0001-9223-0660
Giovanni Volpe https://orcid.org/0000-0001-5057-1846
Maciej Lewenstein https://orcid.org/0000-0002-0210-7800
Ralf Metzler https://orcid.org/0000-0002-6013-7020

References

[1] Einstein A 1905 *Ann. Phys., Lpz.* **322** 549–60
[2] Brown R 1828 *Phil. Mag.* **4** 161–73
[3] Metzler R, Jeon J H, Cherstvy A G and Barkai E 2014 *Phys. Chem. Chem. Phys.* **16** 24128–64
[4] Thapa S, Lomholt M A, Krog J, Cherstvy A G and Metzler R 2018 *Phys. Chem. Chem. Phys.* **20** 29018–37
[5] Burnecki K, Kepten E, Garini Y, Sikora G and Weron A 2015 *Sci. Rep.* **5** 11306
[6] Weron A, Burnecki K, Akin E J, Solé L, Balcerek M, Tamkun M M and Krapf D 2017 *Sci. Rep.* **7** 5404
[7] Sikora G, Wyłomańska A, Gajda J, Solé L, Akin E J, Tamkun M M and Krapf D 2017 *Phys. Rev.* E **96** 062404
[8] Kepten E, Weron A, Sikora G, Burnecki K and Garini Y 2015 *PLoS One* **10** e0117722
[9] Krapf D, Lukat N, Marinari E, Metzler R, Oshanin G, Selhuber-Unkel C, Squarcini A, Stadler L, Weiss M and Xu X 2019 *Phys. Rev.* X **9** 011019
[10] Thapa S, Wylomanska A, Sikora G, Wagner C, Krapf D, Kantz H, Chechkin A and Metzler R 2020 *New J. Phys.* **23** 013008
[11] Muñoz-Gil G, Garcia-March M A, Manzo C, Martín-Guerrero J D and Lewenstein M 2020 *New J. Phys.* **22** 013010
[12] Granik N, Weiss L E, Nehme E, Levin M, Chein M, Perlson E, Roichman Y and Shechtman Y 2019 *Biophys. J.* **117** 185–92
[13] Bo S, Schmidt F, Eichhorn R and Volpe G 2019 *Phys. Rev.* E **100** 010102
[14] Kowalek P, Loch-Olszewska H and Szwabiński J 2019 *Phys. Rev.* E **100** 032410
[15] Seckler H and Metzler R 2022 *Nat. Commun.* **13** 6717
[16] 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** 114691C
[17] Muñoz-Gil G, Manzo C, Volpe G, Garcia-March M A, Metzler R and Lewenstein M 2020 The anomalous diffusion challenge dataset (available at: https://github.com/AnDiChallenge/ANDI_datasets)
[18] Muñoz-Gil G *et al* 2021 *Nat. Commun.* **12** 6253
[19] Gentili A and Volpe G 2021 *J. Phys. A: Math. Theor.* **54** 314003
[20] Kowalek P, Loch-Olszewska H, Łaszczuk Ł, Opała J and Szwabiński J 2022 *J. Phys. A: Math. Theor.* **55** 244005
[21] Garibo-i-Orts Ò, Baeza-Bosca A, Garcia-March M A and Conejero J A 2021 *J. Phys. A: Math. Theor.* **54** 504002
[22] Argun A, Volpe G and Bo S 2021 *J. Phys. A: Math. Theor.* **54** 294003
[23] Li D, Yao Q and Huang Z 2021 *J. Phys. A: Math. Theor.* **54** 404003
[24] Manzo C 2021 *J. Phys. A: Math. Theor.* **54** 334002
[25] Al-hada E A, Tang X and Deng W 2022 *J. Phys. A: Math. Theor.* **55** 274006
[26] Conejero J A, Firbas N, Garibo-i-Orts Ò and Garcia-March M A 2023 *J. Phys. A: Math. Theor.* accepted (https://doi.org/10.1088/1751-8121/acafb3)

[27] Verdier H, Duval M, Laurent F, Cassé A, Vestergaard C L and Masson J B 2021 *J. Phys. A: Math. Theor.* **54** 234001
[28] Muñoz-Gil G, i Corominas G G and Lewenstein M 2021 *J. Phys. A: Math. Theor.* **54** 504001
[29] Meyer P G, Aghion E and Kantz H 2022 *J. Phys. A: Math. Theor.* **55** 274001
[30] Thapa S, Park S, Kim Y, Jeon J H, Metzler R and Lomholt M A 2022 *J. Phys. A: Math. Theor.* **55** 194003
[31] Park S, Thapa S, Kim Y, Lomholt M A and Jeon J H 2021 *J. Phys. A: Math. Theor.* **54** 484001
[32] Vitali S, Paradisi P and Pagnini G 2022 *J. Phys. A: Math. Theor.* **55** 224012
[33] Sabzikar F, Kabala J and Burnecki K 2022 *J. Phys. A: Math. Theor.* **55** 174002
[34] Maraj K, Szarek D, Sikora G and Wyłomańska A 2020 *J. Phys. A: Math. Theor.* **54** 024001
[35] Magdziarz M and Zorawik T 2020 *J. Phys. A: Math. Theor.* **53** 504001
[36] Wang W, Cherstvy A G, Chechkin A V, Thapa S, Seno F, Liu X and Metzler R 2020 *J. Phys. A: Math. Theor.* **53** 474001
[37] Vilk O, Aghion E, Nathan R, Toledo S, Metzler R and Assaf M 2022 *J. Phys. A: Math. Theor.* **55** 334004
[38] Hegde A S and Chandrashekar C 2022 *J. Phys. A: Math. Theor.* **55** 234006
[39] Ablowitz M J, Been J B and Carr L D 2022 *J. Phys. A: Math. Theor.* **55** 384010
[40] Marcone B, Nampoothiri S, Orlandini E, Seno F and Baldovin F 2022 *J. Phys. A: Math. Theor.* **55** 354003