

Record ages of non-Markovian scale-invariant random walks

Received: 19 June 2023

Léo Régnier¹, Maxim Dolgushev¹ & Olivier Bénichou¹ ✉

Accepted: 25 September 2023

Published online: 09 October 2023

 Check for updates

How long is needed for an observable to exceed its previous highest value and establish a new record? This time, known as the age of a record plays a crucial role in quantifying record statistics. Until now, general methods for determining record age statistics have been limited to observations of either independent random variables or successive positions of a Markovian (memoryless) random walk. Here we develop a theoretical framework to determine record age statistics in the presence of memory effects for continuous non-smooth processes that are asymptotically scale-invariant. Our theoretical predictions are confirmed by numerical simulations and experimental realisations of diverse representative non-Markovian random walk models and real time series with memory effects, in fields as diverse as genomics, climatology, hydrology, geology and computer science. Our results reveal the crucial role of the number of records already achieved in time series and change our view on analysing record statistics.

The statistics of records in a discrete time series $(X_t)_{t=0,1,\dots}$ is one of the main topics of interest in the study of extreme events¹, with applications in an increasing number of fields. A record event occurs at time t if all prior observations $(X_{t'})_{t'=0,\dots,t-1}$ are smaller than the last value X_t . In this context, the inter record times τ_n , also called record ages^{2–9}, between the n^{th} and $(n+1)^{\text{st}}$ record, are pivotal, as they characterise the time of occurrence of the next record breaking event such as heatwaves¹⁰, earthquakes^{11,12} or record temperatures¹³.

The theory of records has been studied since the mid-20th century^{14,15}, and is well understood when the random variables $(X_t)_{t=0,1,\dots}$ are independent and identically distributed (i.i.d.)^{16–18}. An important step in the study of records was recently made when observations are the successive positions of a Markovian RW^{4,19–22}, $X_{t+1} = X_t + \eta_{t+1}$, where the steps $(\eta_t)_{t=0,1,\dots}$ are still i.i.d. and symmetric. In this situation, record ages are strictly given by the time T needed to reach a given value for the first time, regardless of the past. This time follows an algebraic tail distribution $\mathbb{P}(T \geq \tau) \propto \tau^{-\theta}$, where θ is the persistence exponent²³, provided by the celebrated Sparre-Andersen theorem²⁴, yielding $\theta = 1/2$. We emphasise that, despite the fact that this RW model accounts for correlations between the observations $(X_t)_{t=0,1,\dots}$, the steps $(\eta_t)_{t=0,1,\dots}$ themselves are independent. As a

result, this model cannot account for memory effects in the increments.

However, as a general rule, real time series are not only correlated but also exhibit such memory effects. When the evolution of an observable is influenced by interactions with hidden degrees of freedom, such as the previous steps of the RW or its interaction with the environment, it cannot be modeled as a Markov process.

This is typically the case for displacement data from various tracers (microspheres, polymers, cells, vacuoles...) in simple²⁵ and viscoelastic fluids^{26–28}, soil^{29,30} and air temperatures³¹, river flows^{32,33}, nucleotide sequence locations^{34,35} and Ethernet traffic^{36–38}. So far, as highlighted in the recent review Ref. 4, almost nothing is known about the record age statistics of non-Markovian processes. The only exceptions concern processes amenable to a Markovian process by adding an extra degree of freedom^{3,8,39}, and a numerical observation in the specific case of the fractional Brownian motion⁹. Here, we provide a general scaling theory which determines the time dependence of the record age statistics of non-Markovian RWs. We show that memory effects significantly alter these statistics. They are no longer solely governed by the persistence exponent θ , but also by another explicitly calculated exponent, which is the hallmark of non-Markovian dynamics.

¹Laboratoire de Physique Théorique de la Matière Condensée, CNRS/Sorbonne Université, 4 Place Jussieu, 75005 Paris, France.

✉ e-mail: benichou@lptmc.jussieu.fr

Results

Main results

We consider a general non-Markovian symmetric RW, whose successive positions form a time series $(X_t)_{t=0,1,\dots}$. These positions satisfy $X_{t+1} = X_t + \eta_{t+1}$, where now the statistics of the steps $(\eta_t)_{t=0,1,\dots}$ may exhibit (I) long-range correlations, (II) interactions with the environment (e.g. footprints left along the trajectory), or (III) explicit space-time dependence (see Fig. 1). Essentially all statistical mechanisms that lead to non-Markovian evolution are encompassed by these features of X_t ⁴⁰. In turn, they allow to account for a variety of real time series displaying memory effects^{41,42}. At large time, X_t is assumed to converge to a scale-invariant process that is continuous (i.e. excluding broadly distributed steps η_t) and non-smooth²³ (meaning that, as for the standard Brownian motion, the trajectory is irregular, having at each point an infinite derivative). Under these conditions, the process is characterised by a walk dimension⁴⁰ $d_w > 1$, such that $X_t \propto t^{1/d_w}$, and the random variable $X_t/t^{1/d_w}$ is asymptotically independent of t . To account for potential aging in the increments, X_t is more generally assumed to have scale-invariant increments, meaning that, for $1 \ll t \ll T$, $X_{t+T} - X_T \propto t^{1/d_w} T^{\alpha/2}$. This defines the aging exponent α ^{43,44} ($\alpha > 0$ corresponding qualitatively to accelerating processes and $\alpha < 0$ to slowing down processes) and an effective walk dimension at short times $d_w^0 \equiv (d_w^{-1} - \alpha/2)^{-1}$. We stress that the class of processes that we consider here covers a very broad range of examples of non-Markovian RWs, as detailed below, despite not covering the particular cases of Lévy flights¹⁹ (which are discontinuous) or of the Random Acceleration Process³ (smooth), which would require a different approach.

We report that the tail distribution $S(n, \tau) \equiv \mathbb{P}(\tau_n \geq \tau)$ of the record age τ_n asymptotically obeys a scaling behaviour $S(n, \tau) = n^{-1} \psi(\tau/n^{d_w})$, displaying two universal distinct algebraic regimes :

$$S(n, \tau) \propto \begin{cases} \frac{1}{n} \left(\frac{n^{d_w}}{\tau} \right)^{\frac{1}{d_w^0}} & \text{for } n^{d_w-d_w^0} \ll \tau \ll n^{d_w}, \\ \frac{1}{n} \left(\frac{n^{d_w}}{\tau} \right)^\theta & \text{for } 1 \ll n^{d_w} \ll \tau \end{cases} \quad (1)$$

where ψ is a process dependent scaling function and the persistence exponent θ has been defined above. Equation (1) explicitly determines the n and τ dependence of the record age statistics of non-Markovian RWs. Fundamental consequences of our results include: (i) In regime 1, defined by $n^{d_w-d_w^0} \ll \tau \ll n^{d_w}$, the record time's decay is governed by an exponent different from θ . While it is not unexpected that the memory of the past affects record age statistics for a non-Markovian process (in particular, it is known that it can change the persistence exponent^{45,46}), it is striking that the corresponding exponent is fully explicit and depends only on the effective walk dimension d_w^0 of the increments. Note that regime 1 can span several orders of magnitude as soon as sufficiently many records have been broken, and thus dominate the observations. (ii) In regime 2, defined by $\tau \gg n^{d_w}$, the decay in the record time can be very different from that of regime 1. This is particularly striking for processes with stationary increments for which the exponent involved in regime 2, $\theta = 1 - 1/d_w$ ⁴⁴, is markedly different from the exponent $1/d_w^0 = 1/d_w$ of regime 1 (with the exception of Markovian RWs for which the two exponents are both 1/2 and a single regime is recovered; note that this single regime of exponent 1/2 is also obtained in the case of Lévy flights, which are not covered by our approach). (iii) The record age distribution ages, in the sense that it depends on the number n of records already achieved. Consequently, the observations of early record ages are not representative of later records and call for a careful analysis of real data (note that the record distribution also ages in time series with i.i.d. observations X_t , which are thus not of the form $X_{t+1} = X_t + \eta_{t+1}$ considered here, but the dependence of this distribution on the number of records and the corresponding statistical mechanisms are very different⁴). Finally, note that despite the existence of two regimes for record ages, because of the explicit dependence of the prefactors of $S(n, \tau)$ on n , the number of records at time t displays a single time regime $n \propto t^{1/d_w}$ (see Supplementary Information, SI).

Derivation of the results

The following outlines the derivation of these results (see SI Sec. S1 for details):

The first step consists in noting that, due to the scale-invariance of the process X_t , the time T_n to reach the n^{th} record, $T_n \equiv \sum_{k=0}^{n-1} \tau_k$, satisfies $T_n \propto n^{d_w}$ and its increments obey $T_{m+n} - T_m \propto m^{d_w-d_w^0} n^{d_w^0}$ (see SI Sec. S1.B). In other words, $\mathbb{P}(T_{m+n} - T_m \leq T)$ is a function of a single variable $T/(m^{d_w-d_w^0} n^{d_w^0})$. Then, $T_{m+n} - T_m = \sum_{k=m}^{m+n-1} \tau_k$ is dominated by the largest record age^{40,47} under the self-consistent assumption that $S(n, \tau) \propto n^{-1+\epsilon_1} \tau^{-y_1}$ for $\tau \ll n^{d_w}$ (regime 1) and $S(n, \tau) \propto n^{-1+\epsilon_2} \tau^{-y_2}$ for $\tau \gg n^{d_w}$ (regime 2) with y_i between 0 and 1. This results in the equation

$$\mathbb{P}(T_{m+n} - T_m \leq T) \simeq \mathbb{P}(\max(\tau_m, \dots, \tau_{m+n-1}) \leq T). \quad (2)$$

Adapting the argument of Ref. 48, we show for continuous scale-invariant non-smooth processes analytically (see Sec. S1.D of SI) and verify numerically (see Sec. S2.C of SI) that, in Eq. (2), the record ages τ_k

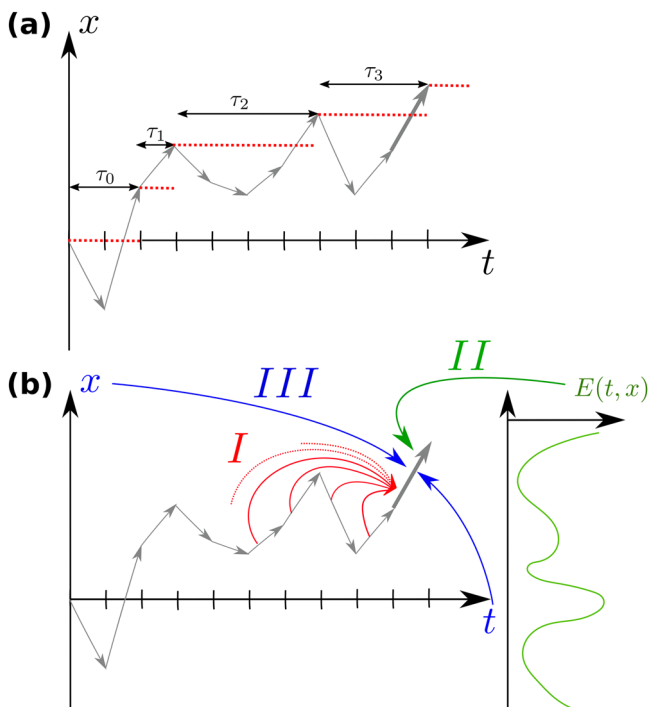


Fig. 1 | Record ages for non-Markovian random walks (RWs). **a** Sketch of a space time trajectory of the RW represented by successive discrete steps η_t (grey arrows). The records in the trajectory are identified by red dotted lines. The record age τ_n of the RW is defined as the time between the n^{th} and $(n+1)^{\text{st}}$ records. **b** Different statistical mechanisms giving rise to a non-Markovian evolution: The statistics of the RW steps η_t may depend on (I) the previous steps of the walk (red arrow), (II) the environment with which the RW interacts (green arrow, schematically represented by the function $E(t, x)$), or (III) the current time or position (blue arrows). In this article, we show that these memory effects strongly modify the record age statistics, which are no longer simply given by the usual persistence exponent θ , but also by a distinct exponent that we determine explicitly.

are asymptotically ($n \gg 1$) effectively independent, which leads to

$$\mathbb{P}(T_{m+n} - T_m \leq T) \simeq \prod_{k=m}^{n+m-1} (1 - S(k, T)). \quad (3)$$

First, for time scales T much smaller than the typical time $T_m \propto m^{d_w}$ required to break m records and for $n \ll m$ (regime 1), Eq. (3) becomes

$$\mathbb{P}(T_{m+n} - T_m \leq T)_{T, n^{d_w} \ll m^{d_w}} \propto \exp\left[-\frac{\text{const.}n}{m^{1-\epsilon_1} T^{\epsilon_1}}\right]. \quad (4)$$

Using $T_{m+n} - T_m \propto m^{d_w - d_w^0} n^{d_w^0}$ gives the exponents of regime 1 as $y_1 = 1/d_w^0$ and $\epsilon_1 = d_w/d_w^0$.

Second, for $\tau \gg n^{d_w}$ (regime 2), the memory of the n broken records no longer affects the algebraic time decay of $S(n, \tau)$, which is thus given by the persistence exponent $\theta = y_2$. Taking $m = 0$ in Eq. (3), we get

$$\mathbb{P}(T_n \leq T) \propto \exp\left[-\text{const.}n^{\epsilon_2}/T^\theta\right]. \quad (5)$$

Using $T_n \propto n^{d_w}$ leads to the exponent $\epsilon_2 = d_w \theta$.

Comparison with numerical simulations of non-Markovian models

We confirm the validity of our analytical results in Fig. 2 by comparing them to numerical simulations of a broad range of representative RW examples, which illustrate the classes (I), (II), and (III) of non-

Markovianity discussed above. Specifically, we consider (see SI for precise definitions and Supplementary Table 1 for a summary of their characteristics): (I) (a) the fractional Brownian motion (fBm), a non-Markovian Gaussian process, with stationary increments given by $\langle (X_t - X_0)^2 \rangle = t^{2H}$, where H is the Hurst exponent; this paradigmatic model has been used repeatedly to account for anomalous diffusion induced by long-range correlations in viscoelastic fluids²⁶ as well as temporal series displaying memory effects^{41,42}; (b) its extension to quenched initial conditions (qfBm), for which the statistics of increments is not stationary anymore, and which describes for instance the height fluctuations under Gaussian noise of an initially flat interface^{44,45}; (c) the elephant RW (eRW)⁴⁹, for which the current step is drawn uniformly from all of the previous steps performed by the RW, and then reversed with probability β ; (II) (d) The Self-Attractive Walk (SATW), (e) Sub-Exponential Self-Repelling Walk (SESRW) and (f) True Self-Avoiding Walk (TSAW) are prototypical examples of self-interacting RWs⁵⁰⁻⁵³, for which the RW deposits a signal at each lattice site it visits and then has a transition probability depending on the number of visits to its neighbouring sites (see SI for precise rules), so that memory emerges from the interaction of the walker with the territory already visited; these RWs have been shown to be relevant in the case of living cells, where it was demonstrated experimentally that various cell types can chemically modify the extracellular matrix, which in turn deeply impact their motility⁵⁴; (III) Two models involving an explicit spatial or temporal dependence of the steps: (g) the subdiffusive (resp. (h) the superdiffusive) Average Lévy Lorentz model (subALL and supALL, respectively)⁵⁵⁻⁵⁷ for which the transmission (resp. reflection) probability at every site decays algebraically with the

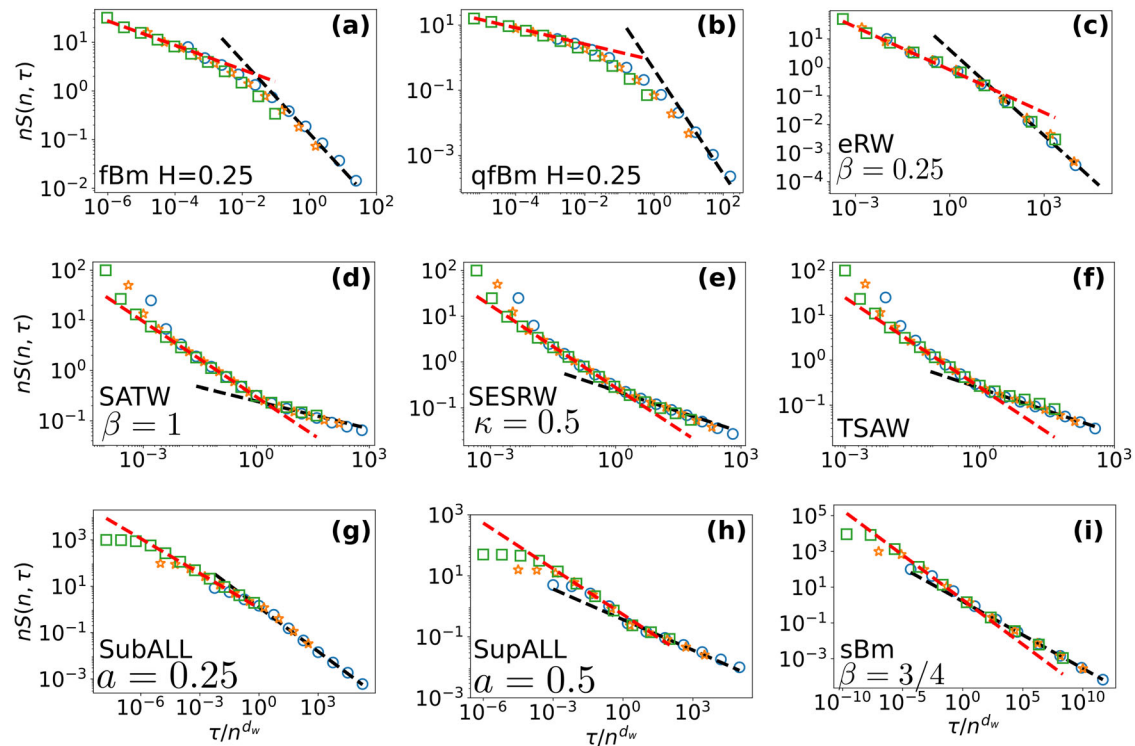


Fig. 2 | Universal record age distributions for non-Markovian RWs: theoretical predictions (lines) vs numerical simulations (symbols). Simulated rescaled tail distribution of record ages τ_n for different values of record number n displayed for various representative RW models: (a) fractional Brownian motion (fBm) of Hurst exponent $H = 0.25 = 1/d_w = 1 - \theta$ for $n = 8, 16$ and 32 (b) quenched fBm (qfBm) of Hurst exponent $H = 0.25 = 1/d_w$ and $\theta \approx 1.55$ for $n = 5, 10$ and 20 (c) elephant RW (eRW) with $\beta = 0.25$ such that $d_w = 2$ and $\theta = 1$, for $n = 10, 25$ and 50 (d) Self-Attractive Walk (SATW) with $\beta = 1$, such that $d_w = 2$ and $\theta = e^{-1/2}$ for $n = 25, 50$ and 100 (e) Sub-Exponential Self-Repelling Walk (SESRW) with $\beta = 1$ and $\kappa = 0.5$ such that $d_w = 5/3$

and $\theta \approx 0.3$ for $n = 25, 50$ and 100 (f) True Self-Avoiding Walk (TSAW) with $\beta = 1$ such that $d_w = 3/2$ and $\theta = 1/3$ for $n = 25, 50$ and 100 (g) subdiffusive Average Lévy Lorentz (subALL) with $a = 0.25$ such that $d_w = 2.75, d_w^0 = 2$ and $\theta = 7/11$ for $n = 10, 100$ and 1000 (h) superdiffusive ALL (supALL) with $a = 0.5$ such that $d_w = 3/2, d_w^0 = 2$ and $\theta = 1/3$ for $n = 10, 100$ and 1000 (i) exact rescaled tail distribution (see SI) for scaled Brownian motion (sBm) with $\beta = 0.75$ such that $d_w = 8/3, d_w^0 = 2$ and $\theta = 3/8$ for $n = 100, 1000$ and $10,000$. Increasing values of n are represented respectively by blue circles, orange stars and green squares. The black dashed line represents the algebraic decay $\tau^{-\theta}$ while the red dashed line stands for the algebraic decay τ^{-1/d_w^0} .

distance to the origin, and (i) the scaled Brownian motion (sBm)⁵⁸ for which the jumping rate is an algebraic function of time, and which is a paradigmatic model of subdiffusion⁵⁹.

Figure 2 reveals excellent quantitative agreement between numerical simulations and our analytical results. The data collapse of the properly rescaled record ages tail distribution and the confirmation of the two successive algebraic decays τ^{-1/d_w^0} and $\tau^{-\theta}$ show that Eq. (1) unambiguously captures the dependence on both the number of records n and the time τ (further confirmed by the analytical determination of the full tail distribution in the solvable case of the sBm, see SI). We emphasise that the very different nature of these examples (subdiffusive and superdiffusive, aging and non-aging, covering all classes of non-Markovian RWs) shows the broad applicability of our approach.

Discussion

We demonstrate the relevance of our results by showing that they apply even when the hidden degrees of freedom responsible for the non-Markovianity of the dynamics are unknown, as is the rule in real observations.

This is illustrated by considering both trajectories involving a variety of tracers in complex fluids (see Fig. 3c–e, which provide experimental realisations²⁶ of several non-Markovian RW models

discussed above) and real time series in diverse fields displaying memory effects, for which record ages are crucial as they characterise the occurrence of extreme events (see Fig. 3a, b and f–h).

Specifically, we consider the following data: (a) river flows³² ($1/d_w \approx 0.14$), (b) volcanic soil temperatures^{29,30} ($1/d_w \approx 0.42$), (c) trajectories of microspheres in gels²⁶ ($1/d_w \approx 0.43$) (d) trajectories of vacuoles inside an amoeba²⁶ ($1/d_w \approx 0.67$), (e) trajectories of telomeres in a nucleus^{26,60} ($1/d_w \approx 0.25$), (f) pyrimidines/purines DNA RW where a step value is given by the nucleotide type, +1 for adenine/thymine, –1 for cytosine/guanine^{34,35} ($1/d_w \approx 0.67$), (g) cumulative air temperatures³¹ ($1/d_w \approx 0.8$), (h) cumulative Ethernet traffic^{36–38} ($1/d_w \approx 0.8$). The walk dimension d_w was estimated by applying the Detrending Moving Average (DMA) method^{61,62} to these data, which removed the deterministic behaviours (see SI for details on the datasets' analysis). Indeed, the characterisation of extreme events, and thus records, requires the meticulous examination of fluctuations around the trend, as underlined in Refs. 31,63.

We stress that we do not require any knowledge on the microscopic details of the process to obtain the record age statistics provided by Eq. (1). In particular, the processes are not necessarily Gaussian and can exhibit various distributions of the increments $x_t \equiv X_{T+t} - X_T$ (see Fig. 3), as long as they are asymptotically scale-invariant (the sampling time of the data is much longer than the

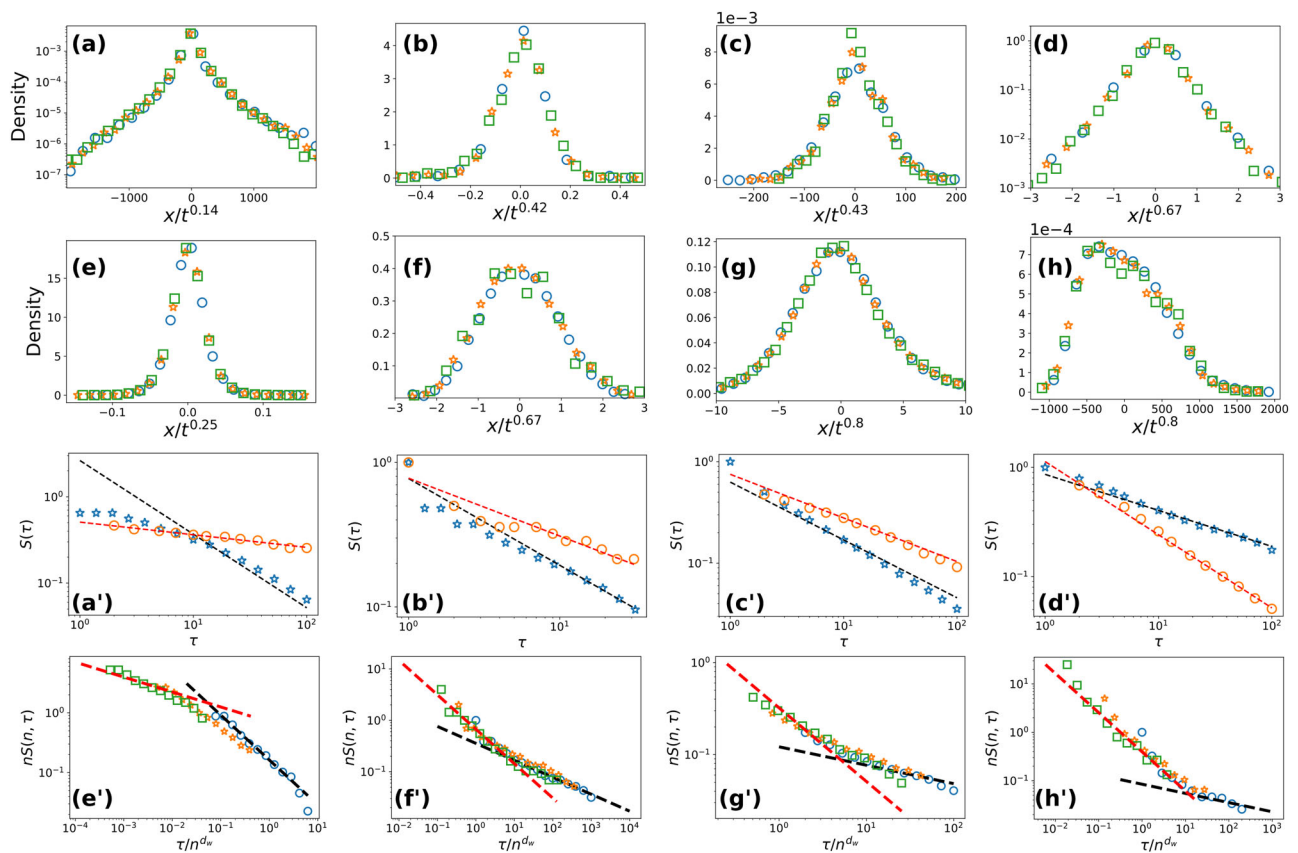


Fig. 3 | Universal record age distributions for non-Markovian RWs: theoretical predictions (lines) vs experimental RW realisations and real time observations (symbols). a–h Distribution of the increment $x_t = X_{t+\tau} - X_t$ at different times t normalised by t^{1/d_w} for: (a) river discharge ($t = 10, 20$, and 40), (b) volcanic soil temperature ($t = 5, 10$, and 20), (c) motion of microspheres in a gel ($t = 2, 4$, and 8), (d) motion of vacuoles inside an amoeba ($t = 10, 20$, and 40), (e) motion of telomeres ($t = 20, 40$, and 80), (f) DNA RW ($t = 20, 40$, and 80), (g) cumulative air temperature ($t = 5, 10$, and 20), and (h) Ethernet cumulative requests ($t = 500, 1000$, and 2000). Increasing values of times are represented successively by blue circles, orange stars and green squares. a'–d' Statistics of the time to first reach the initial

value in the sub interval (blue stars) and the statistics of the records (regardless of the number n of records, orange circles) for (a') river discharge, (b') volcanic soil temperature, (c') motion of microspheres in a gel, and (d') motion of vacuoles inside an amoeba. The black dashed line represents the algebraic decay τ^{-1+1/d_w} while the red dashed line stands for the algebraic decay τ^{-1/d_w} . (e'–h') Rescaled tail distribution of record ages τ_n for different values of the number of records n for (e') motion of telomeres ($n = 1, 3$, and 6), (f') DNA RW ($n = 1, 2$, and 4), (g') cumulative air temperatures ($n = 1, 2$, and 3), and (h') Ethernet cumulative requests ($n = 1, 5$, and 25). Increasing values of n are represented successively by blue circles, orange stars, and green squares. The lines represent the algebraic decays as for (a'–d').

microscopic time scales involved in the process to avoid effects similar to those observed in Ref. 64, as it is checked in Sec. S3 of SI).

Figure 3 demonstrates the quantitative agreement between various real data (see SI Supplementary Fig. 8 for additional datasets, including examples displaying aging of the increments x_i) and our analytical predictions given by Eq. (1). The strong dependence of record ages on the number n of records already achieved, predicted by our analytical approach and confirmed by both numerical simulations and real observations, is a direct manifestation of the non-Markovian feature of the underlying RWs. These results quantitatively demonstrate the significance of memory effects in the record ages of non-Markovian RWs, providing the tools to better predict record-breaking events.

Methods

Numerical simulations of non-Markovian RWs

In this section, we present briefly the models and the numerical methods used to generate the data in Fig. 2.

(a) *Fractional Brownian motion (fBm)*. The fBm is a non-Markovian Gaussian process, with stationary increments. Thus, an fBm X_t of Hurst index H is defined by its covariance

$$\text{Cov}(X_t, X_{t'}) = \frac{1}{2} (t^{2H} + t'^{2H} - |t - t'|^{2H}). \quad (6)$$

The steps $\eta_t = X_t - X_{t-1}$ are called fractional Gaussian noise (fGn). Nowadays, the fBm is broadly spread and its implementations could be found in standard packages of python or Wolfram Mathematica.

(b) *Quenched fBm (qfBm)*. This process is an extension of fBm to quenched initial conditions, which results in non-stationary increment statistics. In particular, it describes the height fluctuations under Gaussian noise of an initially flat interface. Then X_t corresponds to the height of the interface at position $x = 0$, $X_t = h(0, t)$, $h(x, t)$ following the Stochastic Differential Equation (SDE)

$$\partial_t h(x, t) = -(\Delta)^{z/2} h(x, t) + \eta(x, t). \quad (7)$$

Here $\eta(x, t)$ is a Gaussian noise with possible spatial correlations. We solve numerically this SDE with a spatial discretization $\Delta x = 1$ and a time discretization $\Delta t = 0.1$. The system is initially flat, $h(x, t = 0) = 0$.

(c) *Elephant RW (eRW)*. This process is representative of interactions with its own trajectory. At time t , the step η_t is drawn uniformly among all the previous steps η_i ($i < t$) and is reversed with probability β .

(d) *Self-attractive walk (SATW)*. This model is a prototypical example of self-interacting RWs. In the SATW model^{50–53}, the RW at position i jumps to a neighbouring site $j = i \pm 1$ with probability depending on the number of times n_j it has visited site j ,

$$p(i \rightarrow j) = \frac{\exp[-\beta H(n_j)]}{\exp[-\beta H(n_{i-1})] + \exp[-\beta H(n_{i+1})]}, \quad (8)$$

where $H(0) = 0$, $H(n > 0) = 1$ and $\beta > 0$.

(e-f) *Exponential self-repelling RW*. This is another example of self-interacting RW. In this model, the RW at position i jumps to a neighbouring site $j = i \pm 1$ with probability depending on the number of times n_j it has visited site j ,

$$p(i \rightarrow j) = \frac{\exp[-\beta n_j^\kappa]}{\exp[-\beta n_{i-1}^\kappa] + \exp[-\beta n_{i+1}^\kappa]} \quad (9)$$

where κ and β are two positive real numbers.

(g-h) *Average Lévy Lorentz gas (ALL)*. We consider a RW on a 1d lattice with position dependent reflection or transmission probabilities

$r(x)$ or $t(x)$. In the subdiffusive model (resp. superdiffusive model), the transmission coefficient $t(x)$ (resp. reflection coefficient $r(x)$) is taken to be proportional to $|x|^{-\alpha}$ at large distance $|x|$ from the origin.

Data analysis

In this section we provide the method developed to determine the walk dimension of the time series presented in Fig. 3 as well as numerical checks of their stationarity.

(i) *Walk dimension determination*: In order to obtain the walk dimension d_w in a time series, we apply the Detrending Moving Average (DMA) method^{61,62}, which consists in evaluating the typical fluctuations in a window of size ℓ regardless of any bias or deterministic trend. More precisely, for a dataset $(X_t)_{t=0, \dots, N}$, we consider the windows of size up to ℓ_{\max} , compute the window averages $x_t^\ell = \frac{1}{\ell} \sum_{i=0}^{\ell-1} X_{t-i}$, and the typical fluctuation for a window of size ℓ , $F(\ell) = \sqrt{\frac{1}{N-\ell_{\max}} \sum_{t=\ell_{\max}}^N (X_t - x_t^\ell)^2}$. When several trajectories are available, we consider the average fluctuation over all the trajectories (for telomeres, vacuoles and microspheres in agarose data). If the data behave as a RW of walk dimension d_w , then $F(\ell) \propto \ell^{1/d_w}$. We obtain the value of $1/d_w$ via the DMA method to each dataset.

(ii) *Check of stationarity*: In order to check that the data are stationary, we compare the MSD obtained from the increments $\{x_t = X_{t+T} - X_T\}_{T \leq N/4, t}$ in the first quarter of the data and the increments $\{x_t = X_{t+T} - X_T\}_{3N/4 \leq T, t}$ in the last quarter of the data.

(iii) *Record ages in datasets*: Record ages are obtained by starting the subtrajectories at values of t equally spaced at intervals at least 200 time steps long, and observing successive records occurring in the subtrajectory. First return times are obtained by starting the subtrajectories at any value of time.

Data availability

The simulation data of this study are generated based on the code deposited in a GitHub repository⁶⁵ located at <https://github.com/LeoReg/RecordAges>.

The data of the Hadley Centre Central England Temperature (HadCET) project are available at <https://www.metoffice.gov.uk/hadobs/hadcet/>. The data of the European Climate Assessment & Dataset (ECA&D) project are available at <https://www.ecad.eu/>. The volcanic soil temperature data are available at Ref. 30. River discharge data are available at <https://portal.grdc.bafg.de/applications/>. The GenBank database is available at <https://www.ncbi.nlm.nih.gov/genbank/>. The data of traffic traces are available at <http://ita.ee.lbl.gov/html/contrib/BC.html>. Experimental trajectories of fBm realisations are available upon request by the authors of Ref. 26. Experimental cell migration trajectories are available upon request by the authors of Ref. 54.

Code availability

The codes used to generate the simulation data presented in this study as well as the code to analyse the experimental data have been deposited in a GitHub repository located at <https://github.com/LeoReg/RecordAges>.

References

1. Majumdar, S. N., Pal, A. & Schehr, G. Extreme value statistics of correlated random variables: a pedagogical review. *Phys. Rep.* **840**, 1–32 (2020).
2. Kearney, M. J. Record statistics for a discrete-time random walk with correlated steps. *J. Stat. Mech.* **2020**, 023206 (2020).
3. Godrèche, C. & Luck, J.-M. Record statistics of integrated random walks and the random acceleration process. *J. Stat. Phys.* **186**, 4 (2022).

4. Godrèche, C., Majumdar, S. N. & Schehr, G. Record statistics of a strongly correlated time series: random walks and Lévy flights. *J. Phys. A: Math. Theor.* **50**, 333001 (2017).
5. Kumar, A. & Pal, A. Universal framework for record ages under restart. *Phys. Rev. Lett.* **130**, 157101 (2023).
6. Sabhapandit, S. Record statistics of continuous time random walk. *Europhys. Lett.* **94**, 20003 (2011).
7. Benigni, L., Cosco, C., Shapira, A. & Wiese, K. J. Hausdorff dimension of the record set of a fractional brownian motion. *Electron. Commun. Probab.* **23**, 22 (2018).
8. Lacroix-A-Chez-Toine, B. & Mori, F. Universal survival probability for a correlated random walk and applications to records. *J. Phys. A: Math. Theor.* **53**, 495002 (2020).
9. Aliakbari, A., Manshour, P. & Salehi, M. J. Records in fractal stochastic processes. *Chaos* **27**, 033116 (2017).
10. Witze, A. Extreme heatwaves: Surprising lessons from the record warmth. *Nature* **608**, 464–465 (2022).
11. Ambraseys, N. N. Value of historical records of earthquakes. *Nature* **232**, 375–379 (1971).
12. Ben-Naim, E. & Krapivsky, P. L. Statistics of superior records. *Phys. Rev. E* **88**, 022145 (2013).
13. Coumou, D., Robinson, A. & Rahmstorf, S. Global increase in record-breaking monthly-mean temperatures. *Clim. Change* **118**, 771–782 (2013).
14. Chandler, K. N. The distribution and frequency of record values. *J. R. Stat. Soc. Ser. B Methodol.* **14**, 220–228 (1952).
15. Nevzorov, V. B. Records. *Theory Probab. Appl.* **32**, 201–228 (1988).
16. Eliazar, I. & Klafter, J. Record events in growing populations: Universality, correlation, and aging. *Phys. Rev. E* **80**, 061117 (2009).
17. Krug, J. Records in a changing world. *J. Stat. Mech.* **2007**, P07001 (2007).
18. Gouet, R., Lafuente, M., López, F. J. & Sanz, G. Exact and asymptotic properties of δ -records in the linear drift model. *J. Stat. Mech.* **2020**, 103201 (2020).
19. Majumdar, S. N. & Ziff, R. M. Universal record statistics of random walks and lévy flights. *Phys. Rev. Lett.* **101**, 050601 (2008).
20. Majumdar, S. N., Schehr, G. & Wergen, G. Record statistics and persistence for a random walk with a drift. *J. Phys. A: Math. Theor.* **45**, 355002 (2012).
21. Godrèche, C., Majumdar, S. N. & Schehr, G. Universal statistics of longest lasting records of random walks and Lévy flights. *J. Phys. A: Math. Theor.* **47**, 255001 (2014).
22. Ben-Naim, E. & Krapivsky, P. L. Persistence of random walk records. *J. Phys. A: Math. Theor.* **47**, 255002 (2014).
23. Bray, A. J., Majumdar, S. N. & Schehr, G. Persistence and first-passage properties in nonequilibrium systems. *Adv. Phys.* **62**, 225–361 (2013).
24. Klafter, J. & Sokolov, I. M. *First steps in random walks: from tools to applications* (OUP Oxford, 2011).
25. Franosch, T. et al. Resonances arising from hydrodynamic memory in brownian motion. *Nature* **478**, 85–88 (2011).
26. Krapf, D. et al. Spectral content of a single non-brownian trajectory. *Phys. Rev. X* **9**, 011019 (2019).
27. Weiss, M. Single-particle tracking data reveal anticorrelated fractional brownian motion in crowded fluids. *Phys. Rev. E* **88**, 010101 (2013).
28. Reverey, J. F. et al. Superdiffusion dominates intracellular particle motion in the supercrowded cytoplasm of pathogenic *acanthamoeba castellanii*. *Sci. Rep.* **5**, 11690 (2015).
29. Di Crescenzo, A., Martinucci, B. & Mustaro, V. A model based on fractional brownian motion for temperature fluctuation in the Campi Flegrei caldera. *Fractal Fract.* **6**, 421 (2022).
30. Sabbarese, C. et al. Continuous radon monitoring during seven years of volcanic unrest at Campi Flegrei caldera (Italy). *Sci. Rep.* **10**, 9551 (2020).
31. Brody, D. C., Syroka, J. & Zervos, M. Dynamical pricing of weather derivatives. *Quant. Finance* **2**, 189 (2002).
32. Zhang, Q., Xu, C.-Y., Chen, Y. D. & Yu, Z. Multifractal detrended fluctuation analysis of streamflow series of the Yangtze river basin, China. *Hydrol. Process.* **22**, 4997–5003 (2008).
33. Movahed, M. S. & Hermanis, E. Fractal analysis of river flow fluctuations. *Physica A* **387**, 915–932 (2008).
34. Peng, C.-K. et al. Long-range correlations in nucleotide sequences. *Nature* **356**, 168–170 (1992).
35. Peng, C.-K. et al. Mosaic organization of dna nucleotides. *Phys. Rev. E* **49**, 1685–1689 (1994).
36. Leland, W. & Wilson, D. High time-resolution measurement and analysis of lan traffic: Implications for lan interconnection. In *IEEE INFCOM'91. The conference on Computer Communications. Tenth Annual Joint Conference of the IEEE Computer and Communications Societies Proceedings*, 1360–1366 (IEEE, 1991).
37. Fowler, H. & Leland, W. Local area network characteristics, with implications for broadband network congestion management. *IEEE J. Sel. Areas Commun.* **9**, 1139–1149 (1991).
38. Leland, W. E., Taqqu, M. S., Willinger, W. & Wilson, D. V. On the self-similar nature of ethernet traffic. In *Conference proceedings on Communications architectures, protocols and applications*, 183–193 (Association for Computing Machinery, San Francisco, California, USA, 1993).
39. Gabel, A. & Redner, S. Random walk picture of basketball scoring. *J. Quant. Anal. Sports* **8**, <https://doi.org/10.1515/1559-0410.1416> (2012).
40. Bouchaud, J.-P. & Georges, A. Anomalous diffusion in disordered media: Statistical mechanisms, models and physical applications. *Phys. Rep.* **195**, 127–293 (1990).
41. Magdziarz, M., Weron, A., Burnecki, K. & Klafter, J. Fractional brownian motion versus the continuous-time random walk: A simple test for subdiffusive dynamics. *Phys. Rev. Lett.* **103**, 180602 (2009).
42. Mandelbrot, B. B. & Van Ness, J. W. Fractional brownian motions, fractional noises and applications. *SIAM Rev.* **10**, 422–437 (1968).
43. Schulz, J. H. P., Barkai, E. & Metzler, R. Aging renewal theory and application to random walks. *Phys. Rev. X* **4**, 011028 (2014).
44. Levernier, N., Bénichou, O., Guérin, T. & Voituriez, R. Universal first-passage statistics in aging media. *Phys. Rev. E* **98**, 022125 (2018).
45. Majumdar, S. N., Bray, A. J., Cornell, S. & Sire, C. Global persistence exponent for nonequilibrium critical dynamics. *Phys. Rev. Lett.* **77**, 3704 (1996).
46. Levernier, N., Mendes, T., Bénichou, O., Voituriez, R. & Guérin, T. Everlasting impact of initial perturbations on first-passage times of non-markovian random walks. *Nat. Commun.* **13**, 5319 (2022).
47. Vezzani, A., Barkai, E. & Burioni, R. Single-big-jump principle in physical modeling. *Phys. Rev. E* **100**, 012108 (2019).
48. Carpentier, D. & Le Doussal, P. Glass transition of a particle in a random potential, front selection in nonlinear renormalization group, and entropic phenomena in Liouville and sinh-Gordon models. *Phys. Rev. E* **63**, 026110 (2001).
49. Schütz, G. M. & Trimper, S. Elephants can always remember: Exact long-range memory effects in a non-markovian random walk. *Phys. Rev. E* **70**, 045101 (2004).
50. Sapozhnikov, V. B. Self-attracting walk with $v < 1/2$. *J. Phys. A: Math. Gen.* **27**, L151 (1994).
51. Davis, B. Reinforced random walk. *Probab. Theor. Rel. Fields* **84**, 203–229 (1990).
52. Barbier-Chebbah, A., Benichou, O. & Voituriez, R. Anomalous persistence exponents for normal yet aging diffusion. *Phys. Rev. E* **102**, 062115 (2020).
53. Barbier-Chebbah, A., Bénichou, O. & Voituriez, R. Self-interacting random walks: Aging, exploration, and first-passage times. *Phys. Rev. X* **12**, 011052 (2022).

54. d’Alessandro, J. et al. Cell migration guided by long-lived spatial memory. *Nat. Commun.* **12**, 4118 (2021).
55. Radice, M., Onofri, M., Artuso, R. & Cristadoro, G. Transport properties and ageing for the averaged lévy-lorentz gas. *J. Phys. A: Math. Theor.* **53**, 025701 (2019).
56. Radice, M., Onofri, M., Artuso, R. & Pozzoli, G. Statistics of occupation times and connection to local properties of non-homogeneous random walks. *Phys. Rev. E* **101**, 042103 (2020).
57. Barthelemy, P., Bertolotti, J. & Wiersma, D. S. A Lévy flight for light. *Nature* **453**, 495–498 (2008).
58. Lim, S. C. & Muniandy, S. V. Self-similar gaussian processes for modeling anomalous diffusion. *Phys. Rev. E* **66**, 021114 (2002).
59. Saxton, M. J. Anomalous subdiffusion in fluorescence photobleaching recovery: a monte carlo study. *Biophys. J.* **81**, 2226–2240 (2001).
60. Stadler, L. & Weiss, M. Non-equilibrium forces drive the anomalous diffusion of telomeres in the nucleus of mammalian cells. *New J. Phys.* **19**, 113048 (2017).
61. Höll, M., Kiyono, K. & Kantz, H. Theoretical foundation of detrending methods for fluctuation analysis such as detrended fluctuation analysis and detrending moving average. *Phys. Rev. E* **99**, 033305 (2019).
62. Alessio, E., Carbone, A., Castelli, G. & Frappietro, V. Second-order moving average and scaling of stochastic time series. *Eur. Phys. J. B* **27**, 197–200 (2002).
63. Amaya, D. et al. Marine heatwaves need clear definitions so coastal communities can adapt. *Nature* **616**, 29–32 (2023).
64. Zarfaty, L., Barkai, E. & Kessler, D. A. Discrete sampling of extreme events modifies their statistics. *Phys. Rev. Lett.* **129**, 094101 (2022).
65. Régnier, L., Dolgushev, M. & Bénichou, O. Record ages of non-markovian scale-invariant random walks. “<https://zenodo.org/badge/latestdoi/682057871>” (2023).
66. Régnier, L., Dolgushev, M., Redner, S. & Bénichou, O. Complete visitation statistics of one-dimensional random walks. *Phys. Rev. E* **105**, 064104 (2022).
67. Régnier, L., Dolgushev, M., Redner, S. & Bénichou, O. Universal exploration dynamics of random walks. *Nat. Commun.* **14**, 618 (2023).

Acknowledgements

We thank T. Guérin, N. Levernier, and G. Oshanin for helpful discussions, G. Page for careful reading of the paper, and S. Majumdar for mentioning the similarity between the record age statistics and the statistics of the times between visits of new sites^{66,67}. We are thankful to D. Krapf, M. Weiss, F. Taheri and C. Selhuber-Unkel for providing us the experimental trajectories of fBm realisations used in Ref. 26. We thank J. d’Alessandro for providing us the experimental cell migration trajectories analysed in Ref. 54. We acknowledge the data providers in the Hadley Centre Central England Temperature (HadCET) and European Climate Assessment & Dataset (ECA&D) projects. We thank the authors of Ref. 30 for

giving access to the volcanic soil temperature data. We acknowledge the Global Runoff Data Centre (GRDC), 56068 Koblenz, Germany for providing the Elbe and Rhône rivers’ water debit data. We acknowledge the data providers of the GenBank database, hosted by the National Library of Medicine, as well as Jaenicke T., Diederich K.W., Haas W., Schleich J., Lichter P., Pfordt M., Bach A. and Vosberg H.P. who deposited the specific HUMBMYH7 sequence used in this study. We thank the authors of Ref. 36 for the data of traffic traces.

Author contributions

O.B., L.R. and M.D. contributed to analytical calculations. L.R. and M.D. performed numerical simulations. All the authors wrote the paper. O.B. conceived the research.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-023-41945-9>.

Correspondence and requests for materials should be addressed to Olivier Bénichou.

Peer review information *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023