Article

Statistics of Weibull Record-Breaking Events

Robert Shcherbakov [1,2]

1   Department of Earth Sciences, Western University, London, ON N6A 5B7, Canada
2   Department of Physics and Astronomy, Western University, London, ON N6A 3K7, Canada

Abstract: The statistics of record-breaking events plays an important role in the analysis of natural physical systems. It can provide an additional insight into the mechanisms and the occurrence of extreme events. In this work, the statistical aspects of the record-breaking events drawn from the Weibull distribution are considered and analyzed in detail. It is assumed that the underlying sequences of events are independent and identically distributed (*i.i.d.*). Several statistical measures of record-breaking events are analyzed. Exact analytical expressions are derived for the statistics of records. Particularly, the distributions of record magnitudes and the corresponding average magnitudes of records in case of Weibull distributed events are derived exactly for any specific record order and time step. In addition, a convolution operation is used to derive a recursive formula for the distribution of times of the occurrence of records. The analytical results are compared with the Monte Carlo simulations and their validity is confirmed. The numerical simulations also reveal that the finite-size effects strongly affect the statistics of records and need to be considered during the analysis of numerical experiments or empirical data.

Keywords: record-breaking events; interevent times; Weibull distribution; Monte Carlo simulations; finite-size effects

MSC: 60G70; 60E05; 65C05

1. Introduction

Measurements of physical characteristics in various natural phenomena, in many cases, can be considered as a sampling of a stochastic process in time and magnitude domains. This sampling produces time series which reflect the dynamics of the underlying physical process. From these measurements, sequences of record-breaking events with respect to their size (magnitude) can be extracted where each such an event is larger (smaller) than all previous events [1–5]. The most prominent example is the daily temperature measurements and the corresponding high (low) temperatures which were recorded during a particular historical time interval. From the sequence of daily temperatures it is possible to extract a subsequence of record-breaking temperatures. The analysis of these record-breaking temperatures is of critical importance to understand future trends and variations in weather patterns and climate changes [3,6–11].

Record-breaking events extracted from the physical measurements, thought experiments, or computer simulations form a subsequence in time and are distinguished based on their magnitudes. At a given time step a record-breaking event is defined as the largest among all previous records. Among the following events, the event that exceeds the previous record becomes a record-breaking event [1,3–5,8]. The statistical analysis of records was developed for the sequences extracted from the independent and identically distributed (*i.i.d.*) random variables and was based on extreme value statistics [1,4,5,12]. For such *i.i.d.* sequences it is known that some statistical measures of records are independent of their underlying distribution and can be derived analytically [2]. However, the effects of correlations and memory between events introduce complications to the theory.

In recent years some progress has been made in analyzing *i.i.d.* random sequences with time-varying underlying distributions as well as non-*i.i.d.* random sequences with the presence of correlations to study the statistics of their record-breaking events. The daily record temperatures were analyzed in Philadelphia to establish the trends and correlations in their variations [7]. Records drawn from independent random variables but with progressively broadening or sharpening distributions were investigated [9]. To consider the effects of correlations in time series, record-breaking events were extracted from the sequences generated by random walks and Lévy flights [13]. Record-breaking events were observed and studied in the models and experiments describing the processes of rupture and failure [14–19].

The standard model of the occurrence of the records assumes that a single event is added at each time step. The generalization of this model can also be considered where the occurrence of events grows stochastically in time. Particularly, several models with deterministic growth of events were analyzed [20]. The effects of long-term correlations was studied in the context of extreme events to quantify how the distribution of maxima is affected by the length and the presence of persistence in the time series [21]. The same authors also analyzed the statistics of return intervals between extremal events extracted from the long-term correlated time series [22].

Record-breaking statistics has been also applied to seismicity. To analyze clustering both in space and time, the record-breaking statistics was used to quantify the recurrence times between earthquakes [23]. This was generalized to events occurring in space and time by analyzing their recurrences which form a record-breaking process [24]. By assuming that global earthquakes are independent and their magnitudes follow exponential distribution, the sequences of record-breaking earthquakes were extracted and analyzed for world-wide earthquakes with magnitudes greater than 5.5 [25,26]. Record-breaking events can also be studied in the context of natural time analysis [27–30]. Recently, there has been interest in developing the forecasting or nowcasting approaches related to natural seismicity where record-breaking events can also play a prominent role [31–38].

In the present work, the sequences of *i.i.d.* random numbers following the Weibull distribution were generated and the corresponding subsequences of record-breaking events were extracted to analyze their statistical properties. The main goal of the work was to derive analytically and confirm through numerical Monte Carlo simulations several statistical measures describing the distribution of magnitudes and the temporal structure of record-breaking events. The temporal structure of the record-breaking events does not depend on the underlying distribution function from which the records are extracted. In the work, a convolution operation was used to derive the recursive formula for the distribution of times of the occurrence of records. In addition, the non-normalized cumulative log-normal distribution function was used to approximate the average time of the occurrence of the kth record. On the other hand, the distribution of magnitudes and the corresponding averages of record-breaking events are distribution specific. In this respect, the Weibull distribution was used to study several statistical measures of records. As a result, the distribution of the magnitudes of the records, the average magnitudes, the average of the values of records at given time steps, were derived analytically and confirmed through numerical simulations.

The Weibull distribution plays a prominent role in the studies of various problems in physics, geophysics, and engineering. It has been reported that the interoccurrence of characteristic earthquakes on a single fault follows the Weibull distribution [39,40]. It has been also shown that recurrence statistics in the long-range correlated time series follows the stretched exponential distribution. The stretched exponential distribution is the Weibull distribution with the shape parameter β in the range $0 < \beta < 1$. The stretched exponential distribution also plays an important role in the context of the nucleation phenomena [41].

The paper has the following structure. In Section 2, the basic known facts concerning the statistics of record-breaking events extracted from sequences of *i.i.d.* random variables are introduced. Several fundamental expressions for different measures of records are

derived. In Section 3 the analysis of record-breaking events generated from the Weibull distribution is presented. Several analytical results are presented and confirmed through numerical simulations. Section 4 concludes the analysis.

2. Statistics of Record-Breaking Events

In this section, I provide an overview of several known fundamental statistical measures that characterize the record-breaking events. This measures are independent of the underlying distribution from which the records are drawn and valid when events are *i.i.d.*. In addition, I derive an expression for ...

Physical observations or computer simulations can produce a sequence of measurements of a particular observable, $\{m(t_i)\}$, at specific instances of time t_i. Examples abound such as daily temperature measurements, concentrations of carbon dioxide in the atmosphere, flood areas, sport events with the corresponding records, occurrence of earthquakes and volcanic eruptions, etc. These measurements can be considered as a stochastic variable. A record-breaking event $x(t_n)$ up to time t_n, has the largest magnitude among all previous events, $x(t_n) = \max\{m(t_1), m(t_2), \ldots, m(t_{n-1})\}$ [1]. A subsequent event becomes a record-breaking one if it exceeds the current record-breaking event. In this work, a discrete time $n = 1, 2, 3, \ldots$ is assumed to mark the times of the occurrence or generation of events with the simplified notation: $x(t_n) \equiv x(n)$. A subscript $k = 1, 2, 3, \ldots$ is used to mark the record-breaking events in a sequence. For example, x_k specifies the magnitude of the kth record-breaking event. As a result, for a given sequence of random events one can extract the subsequence of the record-breaking events: $\{x_k(n)\} = x_1(n_1), x_2(n_2), x_3(n_3), \ldots$, where $n_1 < n_2 < n_3 < \ldots$

Several fundamental measures can be defined to study the statistics of record-breaking events. The theory of record-breaking events typically assumes that they are *i.i.d.* random numbers [4,5]. This is a direct result that the records are extracted from the sequence of *i.i.d.* events drawn from a given distribution with a density function $f(x)$. The distribution can be bounded or unbounded depending on the problem. Record-breaking events extracted from a bounded distribution will be bounded as well. The probability for the records to not exceed x can be written as

$$F(x) = \int_{x_{\min}}^{x} f(x') dx' \tag{1}$$

where $x_{\min}$ specifies the lower bound of the distribution function.

2.1. Frequency-Magnitude Statistics of Record-Breaking Events

It is possible to compute the distribution of record magnitudes for each order k. The probability density function for the kth record has the form [7]:

$$p_k(x) = \left[\int_{x_{\min}}^{x} \frac{p_{k-1}(x')}{1 - F(x')} dx' \right] f(x) \tag{2}$$

where $F(x)$ is the distribution function given in Equation (1) from which the records are drawn. Equation (2) is a recursive formula to compute the distribution of magnitudes for the kth record given the distribution for the $(k-1)$st record-breaking event. The distribution of the first record, p_1, coincides with the distribution from which the random variables are drawn, $p_1(x) = f(x)$. For the second record order $k = 2$, by noticing that $p_1(x) = f(x) = \frac{dF}{dx}$, one can compute

$$
\begin{aligned}
p_2(x) &= \left[\int_{x_{\min}}^{x} \frac{dF(x')}{1 - F(x')} \right] f(x) = \left[-\int_{0}^{F(x)} d[\ln(1 - F)] \right] f(x) \\
&= -\ln[1 - F(x)] f(x).
\end{aligned}
\tag{3}
$$

This generalizes for an arbitrarily order k and can be proved by induction that the general form of Equation (2) is [42]

$$p_k(x) = f(x) \frac{\{-\ln[1 - F(x)]\}^{k-1}}{(k-1)!}. \tag{4}$$

Equation (4) is valid for records drawn from *i.i.d.* random variables with the underlying distribution function $F(x)$.

Using the above derived Equation (4), one can also compute the average magnitude, $\langle x_k \rangle$, of the kth record-breaking event

$$
\begin{aligned}
\langle x_k \rangle &= \int_S x' p_k(x') \, dx' \\
&= \frac{1}{(k-1)!} \int_S x' \{-\ln[1 - F(x')]\}^{k-1} dF(x') \quad k = 1, 2, \ldots,
\end{aligned} \tag{5}
$$

where $S = [x_{\min}, x_{\max}]$ is the support of the distribution function $F(x)$. In the next section we will show that the integral in Equation (5) can be computed exactly in the case of Weibull distributed random variables.

Similarly, the average magnitude $\langle x(n) \rangle$ of a record-breaking event at a given time step n has the form:

$$\langle x(n) \rangle = \int_S x \, q(x, n) \, dx \tag{6}$$

where $q(x, n)$ specifies the probability density function for the records to occur at time n. Therefore, $q(x, n)dx$ is a probability to have the magnitude of the record to be between x and $x + dx$ at a time step n. This probability density function $q(x, n)$ is related to $F(x)$ as [8]

$$q(x, n) = n[F(x)]^{n-1} f(x). \tag{7}$$

Noticing that $f(x) = \frac{dF}{dx}$, Equation (6) can be written in the following form

$$\langle x(n) \rangle = n \int_S x \, [F(x)]^{n-1} \, dF(x) = \int_0^1 x \, d[F(x)]^n. \tag{8}$$

The reviewed results, Equations (4)–(8), are valid only for *i.i.d.* random variables.

2.2. Temporal Structure of Record-Breaking Events

To characterize the occurrence of records in time, one can estimate the average number of record-breaking events, $\langle N_n \rangle$, that occurred up to a time step n. The quantity N_n is a random variable. In case of *i.i.d.* events from which the sequence of record-breaking events is extracted, the probability for the jth record-breaking event is $P_j = \frac{1}{j}$ [4]. Therefore, the probability decreases harmonically with increasing time steps.

It can be shown that the average number of records, $\langle N_n \rangle$, is [4]

$$\langle N_n \rangle = \sum_{j=1}^n P_j = H_n \simeq \gamma + \ln(n) + \mathcal{O}(1/n) \qquad \text{for} \qquad n \to \infty \tag{9}$$

where $\gamma \approx 0.577215665\ldots$ is the Euler–Mascheroni constant and H_n is a harmonic number, $H_n = \sum_{j=1}^n \frac{1}{j}$, Ref. [43]. This signifies that the average number of records increases as $\langle N_n \rangle \sim \ln n$ for large n and is independent of the underlying distribution function $F(x)$. When the records are extracted from processes with memories or long range correlations, the average number can deviate from the growth given in Equation (9).

To quantify the variability of the average number of records, one can compute the variance [9]

$$
\begin{aligned}
\mathrm{Var}(N_n) \;\;&=\;\; \langle (N_n - \langle N_n \rangle)^2 \rangle = \sum_{j=1}^{n} \left(\frac{1}{j} - \frac{1}{j^2} \right) \\
&\simeq\;\; \gamma + \ln(n) - \frac{\pi^2}{6} + \mathcal{O}(1/n), \quad n \to \infty.
\end{aligned}
\tag{10}
$$

For *i.i.d.* record-breaking events, the ratio of the variance, Equation (10), to the mean, Equation (9), approaches unity as $n \to \infty$ and the distribution of the number of events at each time step N_n become Poisson with the mean value $\ln n$. This signifies that the occurrence of records follows a log-Poisson process [9].

The distribution of times between two subsequent record-breaking events (interevent times) characterizes the process of the occurrence of records. The interevent time between kth and $(k-1)$st record-breaking event is defined as $m = t_k - t_{k-1}$. For the records drawn from *i.i.d.* random events, the distribution of interevent times is independent of the underlying distribution of magnitudes and the corresponding non-normalized histogram follows a power law: $G(m) = 1/m$, [8]. This power-law distribution is obtained by considering all interevent times between records in a given sequence.

In addition, it is possible to consider the probability that the kth record is broken after m time steps. The corresponding distribution function, $w_k(m)$, provides a more detailed structure of times between consecutive records [5,7]

$$
w_k(m) = \int_0^\infty p_k(x)\,[F(x)]^{m-1}[1 - F(x)]\,dx \quad \text{for} \quad m \ge 1
\tag{11}
$$

where $F(x)$ is the underlying distribution function of the random variables from which records are drawn. The expression, $F^{m-1}(x_k)[1 - F(x_k)]$, gives the probability that the previous record, kth, is broken after $m \ge 1$ time steps and the new $(k+1)$st record has the value x_{k+1}. Equation (11) is obtained by averaging this probability over all possible values of x.

The probability that the 1st record ($k = 1$) is broken after m time steps, $w_1(m)$, can be computed explicitly. This can be achieved by using the fact that $p_1(x) = f(x) = \frac{dF}{dx}$. Substituting this into Equation (11) and performing integration by parts, one has

$$
w_1(m) = \int_0^1 F^{m-1}(1 - F)\,dF = \frac{1}{m(m+1)} \quad \text{for} \quad m \ge 1.
\tag{12}
$$

This distribution is a power-law; as a result, the average interevent time between the first and the second records is infinite, $\langle m_1 \rangle = \sum_{m=1}^{\infty} m\,w_1(m) = \infty$.

Using Equation (4), Equation (11) can be written using the cumulative distribution function $F(x)$ with the result

$$
w_k(m) = \frac{1}{(k-1)!} \int_0^1 [-\ln(1 - F)]^{k-1}\,F^{m-1}(1 - F)\,dF.
\tag{13}
$$

The integration can be performed explicitly and one obtains [42]:

$$
w_k(m) = \sum_{l=1}^{m} (-1)^{l+1} \frac{(m-1)!}{(m-l)!(l-1)!} \frac{1}{(l+1)^k}.
\tag{14}
$$

For $k = 2$ and 3, one has

$$w_2(m) = \frac{H_m}{m} - \frac{H_{m+1}}{m+1} \tag{15}$$

$$w_3(m) = -\frac{1}{2(m+1)^3} + \frac{\pi^2}{12m(m+1)}$$

$$+ \frac{H_m^2}{2m} - \frac{H_{m+1}^2}{2(m+1)} - \frac{\Psi(m+1)}{2m(m+1)} \tag{16}$$

where H_m is a harmonic number and $\Psi(m+1) = -\gamma + H_m$ is a digamma function [43]. The obtained results illustrate that, for the records drawn from the *i.i.d.* random variables, the probability distribution $w_k(m)$ is independent of the magnitude distribution $F(x)$ of records.

Finally, it is also possible to define and analyze the probability distribution, $u_k(n)$, for the time of the occurrence of the kth record-breaking event at a given time step n. By knowing this probability distribution, one can compute the average time $\langle n_k \rangle = \sum_{n=1}^{\infty} n\, u_k(n)$ of the occurrence of the kth record-breaking event. It is obvious that the first event is always a record-breaking event; as a result, $\langle n_1 \rangle = 1$. The distribution for the time of the occurrence of the second record-breaking event, $u_2(n)$, is the same as the distribution, $w_1(n)$, of interevent times between the first ($k = 1$) and second ($k = 2$) records. The occurrence times of the kth record, n_k, is a random variable. It can be computed as a sum of two random variables, $n_k = n_{k-1} + m_{k-1}$, where n_{k-1} is the occurrence time of the $(k-1)$st record and m_{k-1} is the interevent time between $(k-1)$ and k records. Therefore, the distribution of times of the occurrence of the kth record, $u_k(n)$, can be computed recursively using the discrete convolution of the two densities $u_{k-1}(n)$ and $w_{k-1}(m)$ with the result:

$$u_k(n) = \sum_{m=1}^{\infty} u_{k-1}(n - m)\, w_{k-1}(m). \tag{17}$$

In practice, this distributions cannot be evaluated explicitly, except for $k = 2$, where it coincides with $w_1(n)$. Instead, one can use a numerical approximation by using long but finite sequences of events to perform the convolution operation and computing the distributions recursively for specific values of k. This can also be achieved through Monte Carlo simulations of events drawn from a well-known distribution to compute explicitly the distributions of the occurrence times. This is going to be illustrated in the next section.

Next, I consider the record-breaking events extracted from the sequences of random variables drawn from the Weibull distribution.

3. Weibull Record-Breaking Events

A particular example of record-breaking events can be analyzed by constructing the sequence of *i.i.d.* random numbers drawn from the Weibull distribution. The probability density function, $f(x)$, for the Weibull distribution is

$$f(x) = \frac{\beta}{\tau}\left(\frac{x}{\tau}\right)^{\beta-1} \exp\left[-\left(\frac{x}{\tau}\right)^{\beta}\right] \tag{18}$$

where β and τ are the shape and scaling parameters, respectively. When $\beta = 1$, this reduces to the exponential distribution. In the case of $0 < \beta < 1$, this defines the stretched-exponential distribution. The corresponding distribution function is given by

$$F(x) = \int_{x_{\min}}^{x} f(x')dx' = 1 - \exp\left[-\left(\frac{x}{\tau}\right)^{\beta}\right]. \tag{19}$$

In order to investigate various statistical measures of record-breaking events drawn from the Weibull distribution, I performed Monte Carlo simulations and compared numerical results with theoretical ones. As stated in the previous section, the temporal structure

of record breaking events is independent from the underlying distribution. On the other hand, the distribution of magnitudes and the corresponding averages are not. In the case of the Weibull distribution, they can be derived analytically. This is illustrated in this section.

First, I illustrate the known results for the evolution of the records drawn from any underlying distribution. The average number of record-breaking events $\langle N_n \rangle$, which occurred up to time step n, is shown in Figure 1a and follows Equation (9). Next, the index of dispersion of record average numbers, which is defined as the ratio of the variance to the mean value of records, is shown in Figure 1b as solid symbols and is computed as the ratio of Equation (10) to Equation (9) at any time step n.

The probability density function, $p_k(x)$, for the magnitude of the kth record-breaking event can be computed analytically using Equations (4), (18), and (19) and has the form:

$$p_k(x) = \frac{1}{(k-1)!} \frac{\beta}{\tau} \left(\frac{x}{\tau}\right)^{k\beta-1} \exp\left[-\left(\frac{x}{\tau}\right)^{\beta}\right] \qquad k = 1, 2, \ldots, \qquad (20)$$

with

$$p_1(x) = f(x) \quad \text{and} \quad p_2(x) = \frac{\beta}{\tau}\left(\frac{x}{\tau}\right)^{2\beta-1} \exp\left[-\left(\frac{x}{\tau}\right)^{\beta}\right]. \qquad (21)$$

The comparison of the Monte Carlo simulations of the Weibull random variables and Equation (20) for the several record orders $k = 1, 2, \ldots, 10$ is given in Figure 2a for $\beta = 4.0$ and $\tau = 1.0$. This confirms the validity of our simulation results. It also shows the deviations from the theoretical distributions given by Equation (20) starting for orders larger than $k \geq 8$. This is attributed to the finiteness, $T = 10^6$, of the generated sequences.
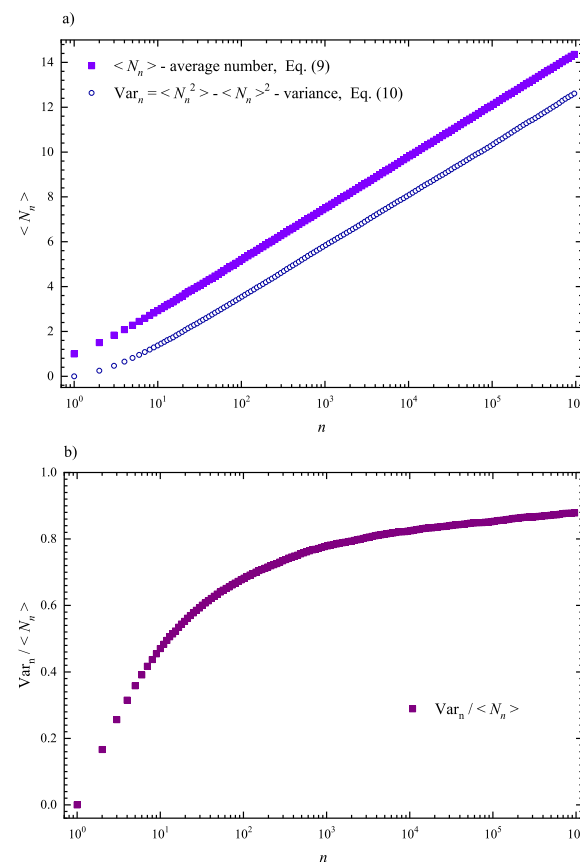

Figure 1. (**a**) The average number, $\langle N_n \rangle$, of record-breaking events versus time step n by Equation (9). (**b**) The index of dispersion of the record numbers is defined as the ratio of the variance to the mean value of the record numbers N_n. This is given as the ratio of Equation (10) to Equation (9).
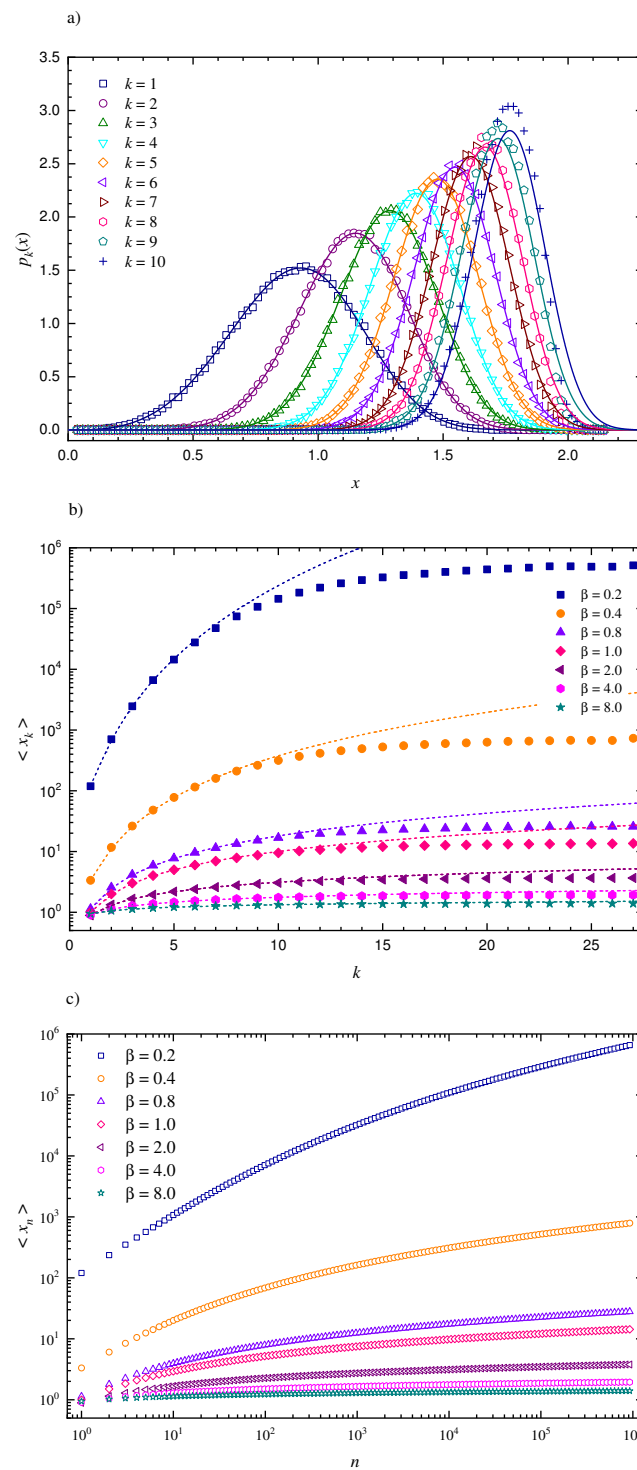
Figure 2. (**a**) The distribution of values of the *k*th record breaking event, $p_k(x)$. Symbols are from Monte Carlo simulations of Weibull random variables with $\beta = 4.0$ and $\tau = 1.0$ for sequences of $T = 10^6$ time steps and averaged over 10^5 realizations. The solid curves are given by Equation (20) for different values of record orders $k = 1, 2, \ldots$; (**b**) the mean value $\langle x_k \rangle$ of the *k*th record-breaking event versus record order, *k*. The dashed curves are given by Equation (22); (**c**) the average magnitude, $\langle x_n \rangle$, of record-breaking events versus a time step *n*. The symbols correspond to the numerical evaluation of Equation (24) for the corresponding values of the Weibull parameters.

The mean value, $\langle x_k \rangle$, of the kth record-breaking event can be evaluated exactly using Equation (20) with the result:

$$\langle x_k \rangle = \int_0^\infty x' p_k(x') \, dx' = \tau \frac{\Gamma\left(k + \frac{1}{\beta}\right)}{(k-1)!} \qquad k = 1, 2, \ldots, \tag{22}$$

where $\Gamma(x)$ is the gamma function. When $\beta = 1$, this reduces to the known result for the exponential distribution, $\langle x_k \rangle = \tau k$, [7]. The comparison of the record-breaking events constructed from the Monte Carlo simulations of the Weibull random variables and a plot of Equation (22) is given in Figure 2b for several values of β and $\tau = 1.0$. The simulated values start to deviate from the theoretical ones starting from the 8th record order for these particular simulations where I have used sequences of $T = 10^6$ time steps. The finiteness of the sequences plays an important role in the statistics of the record breaking events and has to be taken into account when comparing with the analytical results. This is also evident in Figure 2a where the distributions of magnitudes of record-breaking events deviate from ones given by Equation (20) starting from the record order $k = 8$.

The probability density function $q(x, n)$, Equation (7), can be computed analytically in the case of the Weibull random variables. Using Equations (7), (18), and (19), one obtains:

$$q(x, n) = n \frac{\beta}{\tau} \left(\frac{x}{\tau}\right)^{\beta-1} \exp\left[-\left(\frac{x}{\tau}\right)^\beta\right] \left\{1 - \exp\left[-\left(\frac{x}{\tau}\right)^\beta\right]\right\}^{n-1}. \tag{23}$$

The average value $\langle x(n) \rangle$ of the record-breaking events at time step n can be derived by substituting Equation (23) into Equation (6)

$$\langle x(n) \rangle = n \beta \int_0^\infty \left(\frac{x}{\tau}\right)^\beta \exp\left[-\left(\frac{x}{\tau}\right)^\beta\right] \left\{1 - \exp\left[-\left(\frac{x}{\tau}\right)^\beta\right]\right\}^{n-1} dx. \tag{24}$$

The integral in Equation (24) can be evaluated analytically with the result (see Appendix A):

$$\langle x(n) \rangle = \frac{\tau}{\beta} \Gamma\left(\frac{1}{\beta}\right) \sum_{k=1}^n (-1)^{k+1} \binom{n}{k} k^{-\frac{1}{\beta}}. \tag{25}$$

It is worthwhile to note that $\langle x(1) \rangle = \frac{\tau}{\beta} \Gamma\left(\frac{1}{\beta}\right)$ is equal to the mean of the Weibull distribution for given parameters τ and β.

For $\beta = 1$ and $\beta = 1/2$, Equation (25) has the form:

$$\langle x(n) \rangle = \tau H_n \qquad \text{for} \qquad \beta = 1 \tag{26}$$

$$\langle x(n) \rangle = \tau \left[\frac{\pi^2}{6} + H_n^2 - \psi^{(1)}(n+1)\right] \qquad \text{for} \qquad \beta = \frac{1}{2} \tag{27}$$

where $\psi^{(1)}(n+1)$ is a polygamma function [43].

It is also possible to obtain the asymptotic limit of Equation (24) in the case of large time steps:

$$\langle x(n) \rangle \simeq \frac{\tau}{\beta} \Gamma\left(\frac{1}{\beta}\right) (H_n)^{1/\beta} \simeq \frac{\tau}{\beta} \Gamma\left(\frac{1}{\beta}\right) [\gamma + \ln(n)]^{1/\beta} \text{ for } n \to \infty. \tag{28}$$

Using the above Equation (24), one can compute the average value $\langle x(n) \rangle$ of the record-breaking events at different time steps n. The results are illustrated in Figure 2c where the values are computed using the numerical integration of Equation (24) for several values of β and fixed $\tau = 1.0$.

In addition, I estimated the distributions for which the kth record is broken after m time steps, $w_k(m)$, from the numerical simulations of Weibull random variables. I also

compared them with ones given by Equation (13), which is valid for *i.i.d.* random variables drawn from any underlying distribution function. The results of Monte Carlo simulations and evaluation of Equation (13) are shown in Figure 3a for the first several orders of $k = 1, 2, \ldots, 10$, which confirm the derived formulas.

I also computed the distribution of interevent times between all subsequent record-breaking events, $g(m)$. These distributions were constructed by counting all interevent times for all record orders k. The results are illustrated in Figure 3b for the Weibull random variables for several values of β and $\tau = 1.0$. For comparison, I also plot as a dashed line the non-normalized distribution $G(m) = 1/m$. The finite size effects are also present in the distributions $g(m)$. For large values of m, they are influenced by the finiteness of the interval $T = 10^6$.
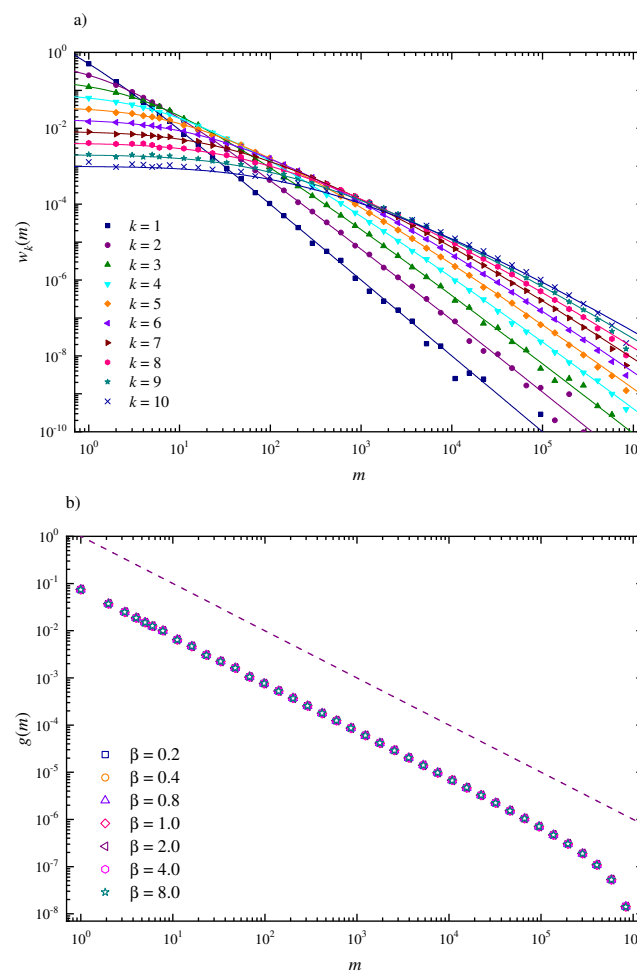

Figure 3. (**a**) The probability density functions, $w_k(m)$ of records for a given order k which are broken after m time steps. Numerical simulations (symbols) as well as analytical results (solid curves, Equation (13)) are shown for the first several orders of $k = 1, 2, \ldots, 10$. (**b**) The distribution of interevent times, $g(m)$, between successive record-breaking events for several values of model parameters are given. For reference, the non-normalized histogram $G(m) = 1/m$ is given as a dashed line.

The probability density functions, $u_k(n)$, for the time of occurrences of the kth record-breaking event versus a time step n are shown in Figure 4a for the first several orders of $k = 2, \ldots, 10$. As mentioned above, the distribution function $u_2(n)$ coincides with the distribution function $w_1(m)$. This is confirmed by plotting Equation (12) as a solid purple line in Figure 4a. The subsequent solid green curve for the ($k = 3$)th record breaking event was computed using the recursive formula, Equation (17). For higher-order record distributions, the computations using the recursive formula become very time consuming.

In addition, the average times $\langle n_k \rangle$ of the occurrence of the kth record-breaking event are given in Figure 4b. These times are independent of the parameters of the Weibull distribution. This is related to the fact that the temporal structure of the record-breaking events drawn from *i.i.d.* random variables does not depend on the underlying distribution from which random variables are drawn.

To approximate the functional form of the variability of the average times $\langle n_k \rangle$, the following function was considered and fitted to the simulated data given in Figure 4b:

$$\phi(x) = A \int_0^x \frac{1}{\sqrt{2\pi}\,\sigma\,u} e^{-\frac{[\ln(u)-\mu]^2}{2\sigma^2}}\, du\,. \tag{29}$$

The estimated parameters from the fit were $A = 634,263 \pm 216,565$, $\mu = 2.87 \pm 0.35$, and $\sigma = 0.44 \pm 0.42$ with the corresponding 95% confidence intervals. The corresponding Equation (29) is plotted as a solid black line in Figure 4b. As a result, the average time for the occurrence of the kth record is $\langle n_k \rangle = \phi(k)$. Equation (29) is in fact the cumulative distribution function of the log-normal distribution multiplied by the parameter A.
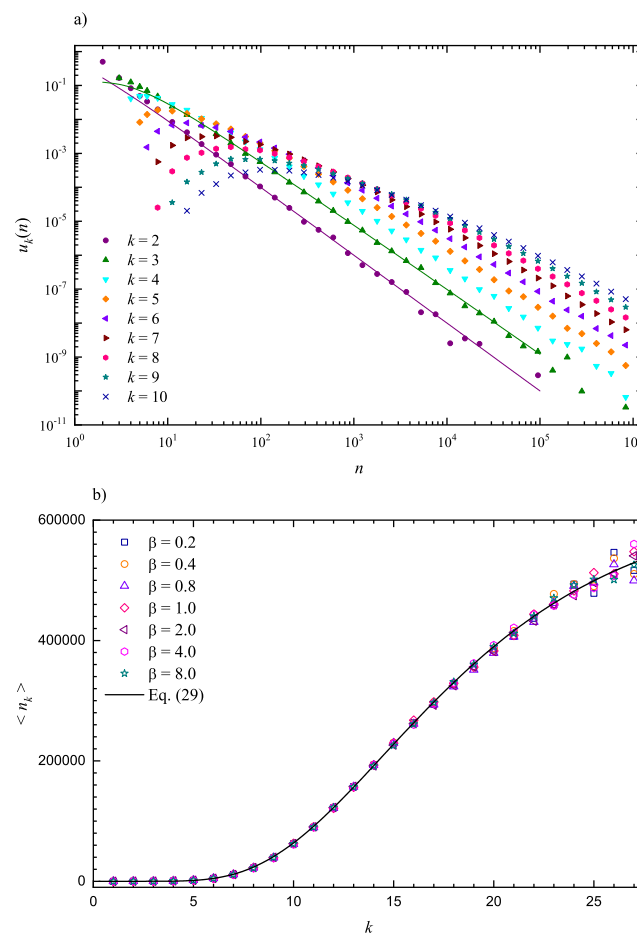

Figure 4. (**a**) The distribution of times of the kth record breaking event, $u_k(n)$, are shown for the first several orders of $k = 2, 3, \ldots, 10$. The solid lines were computed using the recursive formula, Equation (17); (**b**) the average time $\langle n_k \rangle$ of the occurrence of the kth record-breaking event for several values of the Weibull model parameters.

4. Conclusions

In this work, I analyzed both analytically and numerically the statistics of record-breaking events extracted from the sequences of *i.i.d.* random variables drawn from the Weibull distribution. I derived several analytical results concerning the magnitude

distribution and the corresponding averages of records and confirmed them through numerical simulations. The numerical simulations revealed that the finiteness of the sequences considered, T, played an important role for higher record orders, k. This is particularly evident in Figure 2a,b, where one observes significant deviations from the theoretical results for orders larger than $k \geq 8$. This is attributed to the fact that the statistics for higher order records come from the entries generated closer to the end of the sequences. Therefore, the finiteness of the sequences influences these statistics.

I derived exact analytical expressions for the distribution of magnitudes, Equation (20), and the average magnitude, Equation (22), of the record-breaking events of a given order k. Similarly, I derived an exact analytical expression for the average record magnitude at a given time step n reported in Equation (25). This formula has simpler representations for particular values of β, and the expressions are reported in Equations (26) and (27) for $\beta = 1$ and $1/2$, respectively. I also provided the asymptotic form of Equation (25) for large values of n given in Equation (28). In addition, a recursive formula was derived for the distribution of record times, Equation (17). All these obtained results were compared to numerical simulations to confirm their validity.

The presented analysis confirmed that the temporal structure of the studied record-breaking events extracted from the Weibull random variables did not depend on the underlying distribution function. On the other hand, the magnitude distributions and the corresponding average values were controlled by the shape of the underlying distribution from which record sequences were extracted.

Data Availability Statement: No new data were created in this work.

Conflicts of Interest: The author declares no conflict of interest.

Appendix A. Derivation of Equation (25)

To evaluate the integral in Equation (24), I first rewrite it in terms of a new variable $y = \left(\frac{x}{\tau}\right)^{\beta}$

$$
\begin{aligned}
\langle x(n) \rangle &= n\beta \int_0^\infty \left(\frac{x}{\tau}\right)^\beta \exp\left[-\left(\frac{x}{\tau}\right)^\beta\right] \left\{1 - \exp\left[-\left(\frac{x}{\tau}\right)^\beta\right]\right\}^{n-1} dx \\
&= n\tau \int_0^\infty y^{\frac{1}{\beta}} e^{-y} \left(1 - e^{-y}\right)^{n-1} dy.
\end{aligned}
\tag{A1}
$$

Then, I expand the term which depends on n as a binomial sum and exchange the operations of integration and summation with the result:

$$
\begin{aligned}
\langle x(n) \rangle &= n\tau \int_0^\infty y^{\frac{1}{\beta}} e^{-y} \left[\sum_{k=0}^{n-1} (-1)^k \frac{(n-1)!}{(n-1-k)!k!} e^{-ky}\right] dy \\
&= n\tau \sum_{k=0}^{n-1} (-1)^k \frac{(n-1)!}{(n-1-k)!k!} \int_0^\infty y^{\frac{1}{\beta}} e^{-y(k+1)} dy.
\end{aligned}
\tag{A2}
$$

One can observe that the integral in Equation (A2) is proportional to the gamma function. By shifting the index of summation, I finally obtain:

$$
\langle x(n) \rangle = \frac{\tau}{\beta} \Gamma\left(\frac{1}{\beta}\right) \sum_{k=1}^{n} (-1)^{k+1} \binom{n}{k} k^{-\frac{1}{\beta}}.
\tag{A3}
$$

References

1. Tata, M.N. On outstanding values in a sequence of random variables. *Z. Warsch. Verw. Geb.* **1969**, *12*, 9–20. [CrossRef]
2. Rényi, A. On Outstanding Values of a Sequence of Observations. In *Selected Papers of A. Rényi*; Akadémiai Kiadó: Budapest, Hungary, 1976; Volume 3.
3. Glick, N. Breaking Records and Breaking Boards. *Am. Math. Mon.* **1978**, *85*, 2–26. [CrossRef]
4. Nevzorov, V.B. *Records: Mathematical Theory*, 1st ed.; American Mathematical Society: Providence, RI, USA, 2001.
5. Arnold, B.C.; Balakrishnan, N.; Nagaraja, H.N. *Records*; Wiley: New York, NY, USA, 1998.
6. Benestad, R.E. How often can we expect a record event? *Climate Res.* **2003**, *25*, 3–13. [CrossRef]
7. Redner, S.; Petersen, M.R. Role of global warming on the statistics of record-breaking temperatures. *Phys. Rev. E* **2006**, *74*, 061114. [CrossRef]
8. Schmittmann, B.; Zia, R.K.P. "Weather" records: Musings on cold days after a long hot Indian summer. *Am. J. Phys.* **1999**, *67*, 1269–1276. [CrossRef]
9. Krug, J. Records in a changing world. *J. Stat. Mech.* **2007**, *2007*, P07011. [CrossRef]
10. Newman, W.I.; Malamud, B.D.; Turcotte, D.L. Statistical properties of record-breaking temperatures. *Phys. Rev. E* **2010**, *82*, 066111. [CrossRef] [PubMed]
11. Sena, E.T.; Koren, I.; Altaratz, O.; Kostinski, A.B. Record-breaking statistics detect islands of cooling in a sea of warming. *Atmos. Chem. Phys.* **2022**, *22*, 16111–16122. [CrossRef]
12. Nevzorov, V.B. Records. *Theory Probab. Appl.* **1987**, *32*, 201–228. [CrossRef]
13. Majumdar, S.N.; Ziff, R.M. Universal record statistics of random walks and Levy flights. *Phys. Rev. Lett.* **2008**, *101*, 050601. [CrossRef]
14. Varotsos, P.A.; Sarlis, N.V.; Tanaka, H.K.; Skordas, E.S. Similarity of fluctuations in correlated systems: The case of seismicity. *Phys. Rev. E* **2005**, *72*, 041103. [CrossRef] [PubMed]
15. Danku, Z.; Kun, F. Record breaking bursts in a fiber bundle model of creep rupture. *Front. Phys.* **2014**, *2*, 8. [CrossRef]
16. Pál, G.; Raischel, F.; Lennartz-Sassinek, S.; Kun, F.; Main, I.G. Record-breaking events during the compressive failure of porous materials. *Phys. Rev. E* **2016**, *93*, 033006. [CrossRef]
17. Jiang, X.; Liu, H.; Main, I.G.; Salje, E.K.H. Predicting mining collapse: Superjerks and the appearance of record-breaking events in coal as collapse precursors. *Phys. Rev. E* **2017**, *96*, 023004. [CrossRef] [PubMed]
18. Kundu, M.; Mukherjee, S.; Biswas, S. Record-breaking statistics near second-order phase transitions. *Phys. Rev. E* **2018**, *98*, 022103. [CrossRef]
19. Kádár, V.; Pál, G.; Kun, F. Record statistics of bursts signals the onset of acceleration towards failure. *Sci. Rep.* **2020**, *10*, 2508. [CrossRef]
20. Eliazar, I.; Klafter, J. Record events in growing populations: Universality, correlation, and aging. *Phys. Rev. E* **2009**, *80*, 061117. [CrossRef]
21. Eichner, J.F.; Kantelhardt, J.W.; Bunde, A.; Havlin, S. Extreme value statistics in records with long-term persistence. *Phys. Rev. E* **2006**, *73*, 016130. [CrossRef]
22. Eichner, J.F.; Kantelhardt, J.W.; Bunde, A.; Havlin, S. Statistics of return intervals in long-term correlated records. *Phys. Rev. E* **2007**, *75*, 011128. [CrossRef] [PubMed]
23. Davidsen, J.; Grassberger, P.; Paczuski, M. Earthquake recurrence as a record breaking process. *Geophys. Res. Lett.* **2006**, *33*, L11304. [CrossRef]
24. Davidsen, J.; Grassberger, P.; Paczuski, M. Networks of recurrent events, a theory of records, and an application to finding causal signatures in seismicity. *Phys. Rev. E* **2008**, *77*, 066104. [CrossRef]
25. Yoder, M.R.; Turcotte, D.L.; Rundle, J.B. Record-breaking earthquake intervals in a global catalogue and an aftershock sequence. *Nonlinear Proc. Geophys.* **2010**, *17*, 169–176. [CrossRef]
26. Van Aalsburg, J.; Newman, W.I.; Turcotte, D.L.; Rundle, J.B. Record-Breaking Earthquakes. *Bull. Seismol. Soc. Am.* **2010**, *100*, 1800–1805. [CrossRef]
27. Sarlis, N.V.; Skordas, E.S.; Varotsos, P.A. Heart rate variability in natural time and 1/f "noise". *Europhys. Lett.* **2009**, *87*, 18003. [CrossRef]
28. Rundle, J.B.; Luginbuhl, M.; Giguere, A.; Turcotte, D.L. Natural Time, Nowcasting and the Physics of Earthquakes: Estimation of Seismic Risk to Global Megacities. In *Earthquakes and Multi-Hazards Around the Pacific Rim, Vol. II*; Williams, C.A., Peng, Z., Zhang, Y., Fukuyama, E., Goebel, T., Yoder, M.R., Eds.; Springer International Publishing: Cham, Switzerland, 2019; pp. 123–136.
29. Varotsos, P.K.; Perez-Oregon, J.; Skordas, E.S.; Sarlis, N.V. Estimating the Epicenter of an Impending Strong Earthquake by Combining the Seismicity Order Parameter Variability Analysis with Earthquake Networks and Nowcasting: Application in the Eastern Mediterranean. *Appl. Sci.* **2021**, *11*, 10093. [CrossRef]
30. Christopoulos, S.R.G.; Varotsos, P.K.; Perez-Oregon, J.; Papadopoulou, K.A.; Skordas, E.S.; Sarlis, N.V. Natural Time Analysis of Global Seismicity. *Appl. Sci.* **2022**, *12*, 7496. [CrossRef]
31. Rundle, J.B.; Turcotte, D.L.; Donnellan, A.; Grant Ludwig, L.; Luginbuhl, M.; Gong, G. Nowcasting earthquakes. *Earth Space Sci.* **2016**, *3*, 480–486. [CrossRef]
32. Pasari, S. Nowcasting Earthquakes in the Bay of Bengal Region. *Pure Appl. Geophys.* **2019**, *176*, 1417–1432. [CrossRef]

33.   Shcherbakov, R.; Zhuang, J.; Zöller, G.; Ogata, Y. Forecasting the magnitude of the largest expected earthquake. *Nat. Commun.* **2019**, *10*, 4051. [CrossRef]

34.   Shcherbakov, R. Statistics and Forecasting of Aftershocks During the 2019 Ridgecrest, California, Earthquake Sequence. *J. Geophys. Res.* **2021**, *126*, e2020JB020887. [CrossRef]

35.   Pasari, S.; Sharma, Y. Contemporary Earthquake Hazards in the West-Northwest Himalaya: A Statistical Perspective through Natural Times. *Seismol. Res. Lett.* **2020**, *91*, 3358–3369. [CrossRef]

36.   Rundle, J.B.; Donnellan, A.; Fox, G.; Crutchfield, J.P.; Granat, R. Nowcasting Earthquakes: Imaging the Earthquake Cycle in California With Machine Learning. *Earth Space Sci.* **2021**, *8*, e2021EA001757. [CrossRef]

37.   Rundle, J.B.; Stein, S.; Donnellan, A.; Turcotte, D.L.; Klein, W.; Saylor, C. The complex dynamics of earthquake fault systems: New approaches to forecasting and nowcasting of earthquakes. *Rep. Prog. Phys.* **2021**, *84*, 076801. [CrossRef]

38.   Rundle, J.B.; Yazbeck, J.; Donnellan, A.; Fox, G.; Ludwig, L.G.; Heflin, M.; Crutchfield, J. Optimizing Earthquake Nowcasting With Machine Learning: The Role of Strain Hardening in the Earthquake Cycle. *Earth Space Sci.* **2022**, *9*, e2022EA002343. [CrossRef] [PubMed]

39.   Abaimov, S.G.; Turcotte, D.L.; Shcherbakov, R.; Rundle, J.B. Recurrence and interoccurrence behavior of self-organized complex phenomena. *Nonlinear Proc. Geophys.* **2007**, *14*, 455–464. [CrossRef]

40.   Abaimov, S.G.; Turcotte, D.L.; Shcherbakov, R.; Rundle, J.B.; Yakovlev, G.; Goltz, C.; Newman, W.I. Earthquakes: Recurrence and interoccurrence times. *Pure Appl. Geophys.* **2008**, *165*, 777–795. [CrossRef]

41.   Avrami, M. Kinetics of Phase Change. I General Theory. *J. Chem. Phys.* **1939**, *7*, 1103–1112. [CrossRef]

42.   Shcherbakov, R.; Davidsen, J.; Tiampo, K.F. Record-breaking avalanches in driven threshold systems. *Phys. Rev. E* **2013**, *87*, 052811. [CrossRef]

43.   Abramowitz, M.; Stegun, I.A. *Handbook of Mathematical Functions, with Formulas, Graphs, and Mathematical Tables*, 10th ed.; Dover: New York, NY, USA, 1972.