Statistical properties of record-breaking temperatures

William I. Newman*
Department of Earth and Space Sciences, Department of Physics and Astronomy, and Department of Mathematics, University of California, Los Angeles, California 90095, USA

Bruce D. Malamud†
Department of Geography, King's College London, Strand, London WC2R 2LS, United Kingdom

Donald L. Turcotte‡
Department of Geology, University of California, Davis, California 95616, USA
(Received 1 August 2010; revised manuscript received 8 October 2010; published 9 December 2010)

A record-breaking temperature is the highest or lowest temperature at a station since the period of time considered began. The temperatures at a station constitute a time series. After the removal of daily and annual periodicities, the primary considerations are trends (i.e., global warming) and long-range correlations. We first carry out Monte Carlo simulations to determine the influence of trends and long-range correlations on record-breaking statistics. We take a time series that is a Gaussian white noise and give the classic record-breaking theory results for an independent and identically distributed process. We then carry out simulations to determine the influence of long-range correlations and linear temperature trends. For the range of fractional Gaussian noises that are observed to be applicable to temperature time series, the influence on the record-breaking statistics is less than 10%. We next superimpose a linear trend on a Gaussian white noise and extend the theory to include the effect of an additive trend. We determine the ratios of the number of maximum to the number of minimum record-breaking temperatures. We find the single governing parameter to be the ratio of the temperature change per year to the standard deviation of the underlying white noise. To test our approach, we consider a 30 yr record of temperatures at the Mauna Loa Observatory for 1977–2006. We determine the temperature trends by direct measurements and use our simulations to infer trends from the number of record-breaking temperatures. The two approaches give values that are in good agreement. We find that the warming trend is primarily due to an increase in the (overnight) minimum temperatures, while the maximum (daytime) temperatures are approximately constant.

DOI: 10.1103/PhysRevE.82.066111 PACS number(s): 89.90.+n, 92.70.Mn, 05.45.Pq, 92.70.Gt

I. INTRODUCTION

Global warming has received a great deal of attention from a wide variety of studies. These studies have been summarized by Solomon *et al.* [1]. The consensus is that anthropogenic global warming is occurring but that there is considerable stochastic scatter. One observation supporting global warming is the excess of daily record-breaking maximum temperatures over daily record-breaking minimum temperatures [2]. A record-breaking maximum temperature is defined as an observed temperature that is higher than any previous maximum temperature at a prescribed weather station since the period of time considered began. A similar definition applies for a record-breaking minimum temperature. We will begin by reviewing some previous contributions to this topic and briefly discuss the techniques that we will bring to bear here before proceeding to discuss other aspects of the literature and the course we will pursue in this paper.

A systematic study of maximum record-breaking temperatures for the United States has been given by Meehl *et al.* [3]. For nearly 2000 weather stations, record-breaking maximum and minimum temperatures were determined for each day of the year, 1950–2006. For the 7 yr period from January 1, 2000 to December 31, 2006 and for all 2000 stations considered together, there were 291 237 record-breaking maximum temperatures and 142 420 record-breaking minimum temperatures, a ratio of approximately two to one.

Temperature measurements at a weather station constitute a time series. In order to remove daily periodicities, only daily minimum and maximum temperatures are considered. Moreover, in order to remove annual periodicities, we consider the sequence of temperatures on a specified day of the year. The time series of maximum (or minimum) temperatures on a specified day of the year is characterized by trends (i.e., global warming or cooling) and long-range correlations.

There is a rich literature on the topic of record breaking statistics [4]. Tata [5] introduced a basic theory of record-breaking statistics for a sequence of variables drawn from a continuous independent and identically distributed (i.i.d.) process. An example of such a process is a Gaussian white-noise time series. Tata [5] showed that the results are independent of the statistical distribution of measured values. Glick [6] presented several applications including record-breaking temperatures. Several authors [7,8] have considered record-breaking temperatures in terms of global warming.

Redner and Petersen [8] considered 126 yr of daily record-breaking temperature data for Philadelphia and showed that they were well approximated by a Gaussian pro-

*win@ucla.edu
†bruce.malamud@kcl.ac.uk
‡dlturcotte@ucdavis.edu

cess. They carried out Monte Carlo simulations of record-breaking temperature using parameters applicable to their data. They concluded that their data could not be used to establish a warming trend. Benestad [7] approached the study of record-breaking temperatures in a somewhat different way. Considering mean monthly temperatures at a variety of sites, he showed that the number of record-breaking values moving forward in time is significantly greater than the number of record-breaking values moving backward in time, a result consistent with global warming.

Redner and Peterson [8] showed that the Philadelphia data set could be approximated as a Gaussian white noise. We first consider the applicable record-breaking statistics for this case. However, other authors have shown that temperature time series exhibit correlations [9]. For time spans of 10^0-10^5 yr, these time series can be represented as fractional Gaussian noises. Using Monte Carlo simulations of fractional Gaussian noise processes, we obtain the influence of the correlations on the record-breaking statistics. For time spans of 10^1-10^2 yr, it is widely accepted [2] that temperature time series have an increasing temporal trend primarily due to the influence of anthropogenic greenhouse gases [10,11]. Again, we utilize Monte Carlo simulations to obtain the influence of temperature trends on the record-breaking statistics.

To apply our results, we will first consider the 30 yr time series of maximum and minimum daily temperatures at the National Oceanic and Atmospheric Administration (NOAA) Mauna Loa Observatory (MLO) on the Big Island, Hawaii [13]. The observatory established the systematic increase in CO_2 due to anthropogenic inputs. The well-mixed atmosphere at this isolated high-elevation weather station has been accepted [11] as representative of global average values of CO_2. We utilize our Monte Carlo simulations to determine temperature trends from the record-breaking data and compare the results with the directly measured trends. We next utilize our simulations to associate temperature trends with the record-breaking temperature statistics obtained by Meehl *et al.* [3] for some 2000 stations in the United States for the 56 yr period of 1950–2006.

Finally, we compare our results with other studies of global warming. Of particular interest will be the generally accepted decrease in the diurnal temperature range (DTR) [2]. Karl *et al.* [2], using minimum temperatures from stations covering 50% of the northern hemisphere and 10% of the southern hemisphere, observed a narrowing in the DTR from 1951 to 1990.

It should be noted that our studies of record-breaking temperatures are restricted to relatively short periods, 30 and 56 yr. Trends over these periods are generally attributed to anthropogenic global warming. However, Lennartz and Bunde [12] showed that there is a small probability that these trends are due to long-period correlations.

II. RECORD-BREAKING STATISTICS FOR AN INDEPENDENT AND IDENTICALLY DISTRIBUTED PROCESS

In this paper, we consider a time series of maximum and minimum temperatures on a specified day of the year at a specified location. Initially, we will not consider temporal correlations or trends in the time series nor will we consider the specific nature of the observed distribution function. This time series is considered to be a sequence of i.i.d. random variables. We derive an expression for the number of record-breaking events $\langle n_{rb}(n)\rangle$ in the sequence as well as the corresponding standard deviation. The basic analysis for $\langle n_{rb}\rangle$ was given by Tata [5] and clearly illustrated in a tutorial by Glick [6]. In this section, we will reproduce salient portions of their derivation.

From a probabilistic standpoint, we assume that n events have taken place, where an event corresponds to the outcome of a measurement, in our case, a daily maximum or minimum temperature. The first event automatically constitutes the first record-breaking event. With a second event, the likelihood that the second event is record breaking is 1/2, thereby making the average number of record-breaking events $1+\frac{1}{2}$. With a third event, the likelihood that the third event is record breaking is 1/3, thereby making the average number of record-breaking events $1+\frac{1}{2}+\frac{1}{3}$. If we were to generalize this argument, we would observe following n events (or trials) that the average number of record-breaking events would be

$$1+\frac{1}{2}+\frac{1}{3}+\cdots+\frac{1}{n}=\sum_{i-1}^{n}\frac{1}{i}\approx \ln(n)+\gamma, \tag{1}$$

where the latter is a well-known approximation for the sum of the reciprocals of sequential integers and $\gamma \approx 0.577\,215\,664\,9$ is the Euler-Mascheroni constant [6].

In order to proceed, we need to develop a more precise description of this as an uncorrelated process. The outcome of each new event, after the first, determines whether the prevailing record has been broken. We will employ the variable x_i, for $i=2,\ldots,$ to each successive new event where we assign x_i a value of zero, if the ith event is *not* record breaking, and assign it a value of one, if the ith event is record breaking. It follows that the probability that the ith event is record breaking is $1/i$, while the probability that it is not is $1-1/i$. Accordingly, the mean value of x_i over many realizations of the process, which we will call $\langle x_i\rangle$, is

$$\langle x_i\rangle = 0\times\left(1-\frac{1}{i}\right)+1\times\frac{1}{i}=\frac{1}{i}, \tag{2}$$

as we noted earlier. What this means in practical terms is that, if we were to collect and average many examples of the time series, then we would obtain $\langle x_i\rangle$ for the mean. With these definitions and results, it is convenient to extend our definition of the random variables to include the first event, thereby establishing x_1 as always 1, i.e., the probability that it is zero vanishes, and that

$$\langle x_1\rangle = 1. \tag{3}$$

With these definitions, we can now define a new random variable $n_{rb}(n)$ as the number of record-breaking events in $n>0$ trials, i.e.,

$$n_{rb}(n) \equiv \sum_{i=1}^{n} x_i = 1 + x_2 + \cdots + x_n. \quad (4)$$

It immediately follows that the mean of n_{rb} is

$$\langle n_{rb}(n)\rangle = \sum_{i=1}^{n} \langle x_i \rangle = 1 + \frac{1}{2} + \cdots + \frac{1}{n} \approx \ln(n) + \gamma, \quad (5)$$

as we observed in Eq. (1). Since $n_{rb}(n)$ is defined in terms of a set of independent random variables, an extension of the central limit theorem (known as the Lindeberg-Feller theorem [14]) assures us that $n_{rb}(n)$ for a given value of n will be approximately normally distributed. We next determine the variance of $n_{rb}(n)$ for $n>1$.

First, we note, for any $i=1,\ldots,n$, that

$$\langle (x_i - \langle x_i \rangle)^2 \rangle = \left(0 - \frac{1}{i}\right)^2 \left(1 - \frac{1}{i}\right) + \left(1 - \frac{1}{i}\right)^2 \frac{1}{i} = \frac{1}{i}\left(1 - \frac{1}{i}\right). \quad (6)$$

Here, the expected value $\langle x_i \rangle$ inside the left-hand side of the expression is regarded as a constant, while the terms in x_i are treated as random variables. Meanwhile, the right-hand side specifically enumerates the two specific outcomes and their respective probabilities. Moreover, we note that, for $i \neq j$, $i=1,\ldots,n$, and $j=1,\ldots,n$,

$$\langle [x_i - \langle x_i \rangle][x_j - \langle x_j \rangle] \rangle = 0, \quad (7)$$

since the ith and jth events are independent.

Therefore, we observe for $n>1$ that

$$\begin{aligned}\langle [n_{rb}(n) - \langle n_{rb}(n)\rangle]^2 \rangle &= \left\langle \left[\sum_{i=1}^{n} (x_i - \langle x_i \rangle)\right]\left[\sum_{j=1}^{n} (x_j - \langle x_j \rangle)\right] \right\rangle \\ &= \left\langle \sum_{i=1,j=1}^{n} (x_i - \langle x_i \rangle)(x_j - \langle x_j \rangle) \right\rangle \\ &= \sum_{i=1}^{n} \langle [x_i - \langle x_i \rangle]^2 \rangle. \end{aligned} \quad (8)$$

In this calculation, we explicitly expand the left-hand side of the expression. We then combine all terms and conclude by again noting that the $i \neq j$ terms do not contribute to the summations. Accordingly, this leaves

$$\langle [n_{rb}(n) - \langle n_{rb}(n)\rangle]^2 \rangle = \sum_{i=1}^{n} \frac{1}{i}\left(1 - \frac{1}{i}\right) = \sum_{i=1}^{n} \frac{1}{i} - \sum_{i=1}^{n} \frac{1}{i^2}. \quad (9)$$

For *any* time series with n values, the mean value and standard deviation of the record-breaking values are given in Eqs. (5) and (9).

The results are independent of the standard deviation of the time series itself—and all other features of the time series distribution function. Remarkably, this feature of record-breaking statistics allows us to test for i.i.d. behavior without explicitly knowing the statistical distribution of the underlying random variable. However, by estimating the standard error for a record-breaking process and comparing the i.i.d. theory with observations, we can assess whether the underlying time series is i.i.d. or, as we shall see, contains a systematic trend.

We now assume that the values in our simulated temperature time series have a Gaussian distribution consistent with the analysis of observations performed by Redner and Petersen [8]. If sequences of maximum and minimum daily temperatures can be represented as Gaussian white noises, with no trend present, the mean number of record-breaking maximum temperatures as a function of n will on average be equal to the mean number of record-breaking minimum temperatures as a function of n. We will confirm these results in the next section using Monte Carlo simulations.

III. INFLUENCE OF LONG-RANGE CORRELATIONS ON RECORD-BREAKING STATISTICS

A number of studies [9] have shown that temperature time series are well approximated by fractional Gaussian noises. Fractional Gaussian noises have a power-law dependence of the power-spectral density $S(f)$ on the frequency f [15],

$$S(f) \propto f^{-\beta}, \quad (10)$$

with a power-law exponent β. When $\beta=0$, we have a Gaussian white noise, and with $\beta=1$ we have a pink noise. In the range $-1<\beta<1$, we have fractional noises. With $\beta>0$, the time series exhibits long-range correlated, also called long-range persistence or memory, and with $\beta<0$, the time series are long-range antipersistent.

We use the Fourier filtering technique to generate fractional Gaussian noise [8,15]. A Gaussian white noise is generated, a discrete Fourier transformation is carried out, and the resulting Fourier coefficients are filtered using Eq. (10) and the new set of coefficients generate a fractional noise using an inverse discrete Fourier transform.

We utilize our fractional Gaussian noises to obtain record-breaking statistics as a function of β. For each value of β, we generate 1024 realizations of a time series, each with $n=2^{17}=131\,072$ values. We empirically determine the mean number of record-breaking values $\langle n_{rb}\rangle$ as a function of n. Without an added trend, the number of record-breaking maximum values is observed to be equal to the number of record-breaking minimum values for all values of β. The dependence of $\langle n_{rb}\rangle$ on n is given in Fig. 1.

The results for $\beta=0$ are identical to the i.i.d. random variable result given in Eq. (5), as expected. In the range $0<\beta<1/2$, the deviations from the i.i.d. random values are less than 10%. As β approaches unity, the number of record-breaking events increases significantly. Studies of long-range correlations in temperature time series [9] generally give $0.3<\beta<0.5$ for time periods of decades, similar to those examined later in this paper.

The running sum of a Gaussian white noise with $\beta=0$ is a Brownian walk with $\beta=2$ [15]. The running sums of fractional Gaussian noises in the range $-1<\beta<1$ give fractional Brownian walks in the range $1<\beta<3$. We have utilized the running sums of the fractional Gaussian noises produced above to generate fractional Brownian walks. The resulting dependence of $\langle n_{rb}(n)\rangle$ on n is given in Fig. 2.

It is seen that there is a strong increase in the number of record-breaking events with increasing β. The results for β

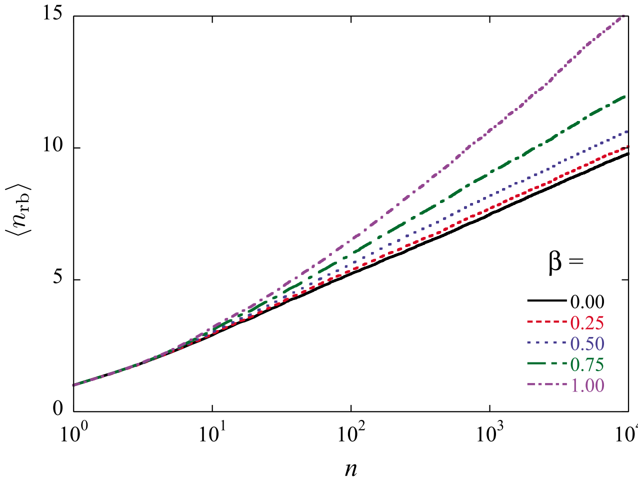

FIG. 1. (Color online) Dependence of the mean number of record-breaking values $\langle n_{\rm rb} \rangle$ as a function of the number of events n. These results do not depend on using maximum or minimum values. Results are shown for fractional Gaussian noises with $\beta = 0$, 0.25, 0.50, 0.75, and 1.00. The results for the white noise $\beta = 0$ are identical to the i.i.d. random variable theory given in Eq. (5).

$=2.0$, 2.5, and 3.0 can be roughly approximated by a power-law relation over a given range of n,

$$\langle n_{\rm rb}(n) \rangle \propto n^{\zeta}, \tag{11}$$

with $\zeta = 0.49$, 0.69, and 0.83, respectively. These results for ζ were obtained by fitting a power-law to $\langle n_{\rm rb} \rangle$ uniformly sampling n over the range from 10^3 to 10^4. A recognized property of fractional Brownian walks is that the standard deviation σ of the walk has a power-law dependence on the number of values in the walk [15],

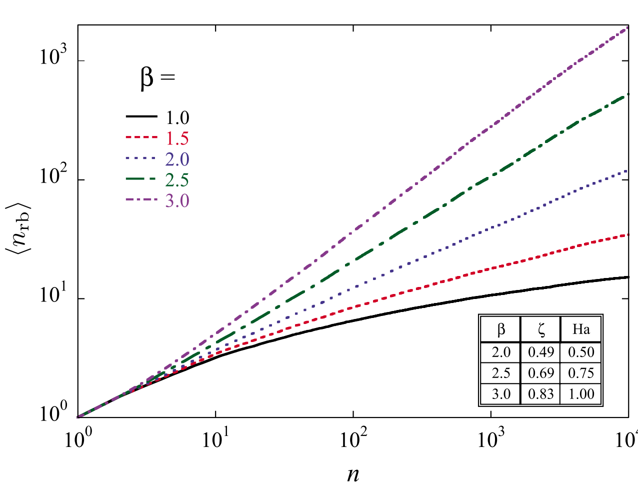

FIG. 2. (Color online) Dependence of the mean number of record-breaking values $\langle n_{\rm rb} \rangle$ as a function of the number of events. Results are shown for fractional Brownian walks $\beta = 1.0$, 1.5, 2.0, 2.5, and 3.0. As β increases, the corresponding curve initially preserves $\langle n_{\rm rb} \rangle = 1$ when $n = 1$, but rises above the previously displayed above curve as n increases. The results for $\beta = 2.0$, 2.5, and 3.0 are compared with the power-law correlation given in Eq. (11) for ζ and Eq. (12) for Ha.

$$\sigma \propto n^{\rm Ha}, \tag{12}$$

where Ha is the Hausdorff measure. The Hausdorff measure is related in turn to the power-spectral density exponent β by

$$\mathrm{Ha} = \tfrac{1}{2}(\beta - 1). \tag{13}$$

For a Brownian walk ($\beta = 2$), we have Ha=0.5 and we found (Fig. 2) $\zeta = 0.48$. However, for $\beta = 2.5$, we have Ha=0.75 and $\zeta = 0.69$, and for $\beta = 3.0$ we have Ha=1.0 and $\zeta = 0.83$. Thus, the Hausdorff measure appears to be related to ζ, but there is no strong correspondence. What has emerged from this analysis is the appreciation that long-range correlations can have an influence upon record-breaking statistics after a large number of events n and, especially, as fractional Gaussian noises become increasingly nonwhite. For the range of β observed in temperature time series [9], $0.3 < \beta < 0.5$, the influence of long-range correlations appears to be small.

IV. INFLUENCE OF A LINEAR TREND ON RECORD-BREAKING STATISTICS FOR i.i.d. PROCESSES

We next consider the influence of a linear trend on record-breaking statistics. We consider a time series z_i, where $i = 1, 2, \ldots, n$. For the applications that we consider in this paper, this could be the maximum (or minimum) temperature observed each day. In order to introduce a linear trend to our time series, we assume that $\langle z_i \rangle$ increases linearly with i, namely,

$$\langle z_i \rangle = \alpha(i - 1) + \langle z_1 \rangle, \tag{14}$$

where z_1 is our starting value and the slope of the trend line α is the change in $\langle z_i \rangle$ from step i to step $i+1$. The values of z_i are selected randomly from a specified distribution with the mean given by Eq. (14) and with unit standard deviation.

In order to better understand the role of a linear trend, let us rederive Eq. (2), which we stated invoking "Occam's razor" from first principles, this time explicitly introducing the underlying (one-point) distribution function. In so doing, we are extending the theory developed by Tata [5] and elucidated by Glick [6]. Suppose that we are measuring a random variable (like the temperature as mentioned earlier) z, such that $-\infty \le z < \infty$, which is i.i.d. and described, in general, by some cumulative distribution function $P(z)$ and a probability density function $p(z) = dP(z)/dz \ge 0$. The differentiability of $P(z)$ is not essential but is assumed for convenience. In our application, z represents the temperature. We begin by assuming that there is no trend. For the ith event, it follows for $i > 1$ that

$$\langle x_i \rangle = \int_{-\infty}^{\infty} p(z_i) dz_i \left\{ \int_{-\infty}^{z_i} p(z_{i-1}) dz_{i-1} \times \int_{-\infty}^{z_i} p(z_{i-2}) dz_{i-2} \cdots \int_{-\infty}^{z_i} p(z_1) dz_1 \right\}, \tag{15}$$

since $\langle x_i \rangle$ is the probability that $z_i \ge z_{i-j}$ for $j = 1, \ldots, i-1$. It follows immediately that this can be rewritten as

$$\langle x_i\rangle = \int_{z_i=-\infty}^{\infty} dP(z_i)P^{i-1}(z_i) = \int_{P=0}^{1} P^{i-1}dP = \frac{1}{i}, \quad (16)$$

as expected from Eq. (2).

We now introduce a linear trend as described above in Eq. (14). Further, we must generalize our probability distributions $p(z)=dP(z)/dz$ to accommodate specifically time, i.e., the event number i. Henceforth, we will employ the distributions $p(z,i)=dP(z,i)/dz$. Accordingly, it follows that the average value of z at event i, or z_i, satisfies

$$\langle z_i\rangle = \int_{-\infty}^{\infty} zp(z,i)dz. \quad (17)$$

Intuitively, as we are focused on event i, it follows for $j=0,\ldots,i-1$ that we can write

$$p(z,i-j) = p_i(z-j\alpha), \quad (18)$$

where $p_i(z)=dP_i(z)/dz$ describes the distribution of the ith event z which we have referred to as z_i. What this means is that with each event going further back in time, our distribution function has shifted by a unit α. Therefore, using Eqs. (17) and (18), we directly recover Eq. (14).

We then observe that

$$\langle x_i\rangle = \int_{-\infty}^{\infty} p_i(z_i)dz_i\left\{\int_{-\infty}^{z_i} p_i(z_{i-1}-\alpha)dz_{i-1} \times \int_{-\infty}^{z_i} p_i(z_{i-2}-2\alpha)dz_{i-2}\cdots\int_{-\infty}^{z_i} p_i[z_1-(i-1)\alpha]dz_1\right\}. \quad (19)$$

This can be rewritten as

$$\langle x_i\rangle = \int_{-\infty}^{\infty} p_i(z_i)dz_i\left\{\int_{-\infty}^{z_i+\alpha} p_i(z_{i-1})dz_{i-1} \times \int_{-\infty}^{z_i+2\alpha} p_i(z_{i-2})dz_{i-2}\cdots\int_{-\infty}^{z_i+(i-1)\alpha} p_i(z_1)dz_1\right\}. \quad (20)$$

So, we can write

$$\langle x_i\rangle = \int_{z_i=-\infty}^{\infty} dP_i(z_i)P_i(z_i+\alpha)P_i(z_i+2\alpha)\cdots P_i[z_i+(i-1)\alpha]. \quad (21)$$

We observe that the derivative of $\langle x_i\rangle$ with respect to α is *always* positive for any α or i and, hence, n. This follows because the derivative of P_i with respect to its argument is non-negative. We now want to understand how this quantity varies as α and/or n become large.

Recalling that $P_i(z)$ is a nondecreasing function, two immediate results follow:

(1) If α is positive,

$$P_i(z_i) \le P_i(z_i+\alpha) \le P_i(z_i+2\alpha) \le \cdots \le P_i[z_i+(i-1)\alpha] \le 1,$$

and we can replace the P_i terms in the preceding with unity and observe the bounding result that

$$\langle x_i\rangle \le \int_0^1 dP_i(z_i)\times 1\cdots 1 = 1. \quad (22)$$

(2) If α is negative (or, conversely, corresponding to positive α but for seeking minimum temperatures),

$$P_i(z_i) \ge P_i(z_i-\alpha) \ge P_i(z_i-2\alpha) \ge \cdots \ge P_i[z_i-(i-1)\alpha] \ge 0,$$

and we can replace the P_i terms in Eq. (21) with zero and observe the bounding result that

$$\langle x_i\rangle \ge \int_0^1 dP_i(z_i)\times 0\cdots 0 = 0, \quad (23)$$

since the P_i terms are bounded from below by zero.

Thus, with a strong linear temperature trend, we could expect that the number of record-breaking high-temperature events will increase more rapidly than in the i.i.d. situation with a limiting value of $\langle n_{\mathrm{rb}}(n)\rangle \le n$. This is clear, as we have remarked, since the first derivative of $\langle z_i\rangle$ with respect to any $\alpha>0$ and for any $n>0$ is positive. Similarly, we could expect that the number of record-breaking low-temperature events will increase less rapidly than in the i.i.d. situation with a limiting value of $n_{\mathrm{rb}}=1$. An important outcome of these theoretical considerations is that the ratio of the number of maximum temperatures to the number of minimum temperatures will depend on the trend slope α. For $\alpha>0$ and the inequalities shown before Eq. (22) where the relevant probabilities are greater than or equal to those in the trend-free situation, it follows that $\langle x_i\rangle$ will be larger than the expected value of $1/i$ derived in Eq. (16). Similarly, considering record lows is mathematically equivalent to the case $\alpha<0$, and the inequalities following Eq. (22) apply where the probabilities are now less than or equal to those in the trend-free situation. This then establishes that $\langle x_i\rangle$ for record-breaking lows will be less than the expected value of $1/i$. Consequently, the ratio of the expected values of the number of record-breaking highs to the number of record-breaking lows will be greater than 1 and can be expected to increase with n.

Another intuitive way of appreciating this analytical result emerges by noting that the monotonicity in the cumulative probability distribution guarantees that $P(x_i+n\alpha)\to 1$ and $P(x_i-n\alpha)\to 0$ as $n\to\infty$ with $\alpha>0$ in Eq. (21). This assures that the number of record-breaking highs will exceed the number of record-breaking lows. Therefore, for $\alpha=0$, i.e., no trend, the ratio should be unity but will increase in response to increasing trend slope α. The preceding discussion also facilitates the analytical calculation of $\langle n_{\mathrm{rb}}\rangle$ for some special distribution functions. In addition, it provides insight into the emergence of a linear trend in $\langle n_{\mathrm{rb}}(n)\rangle$ as a function of n, particular for $n\gg\alpha^{-1}$.

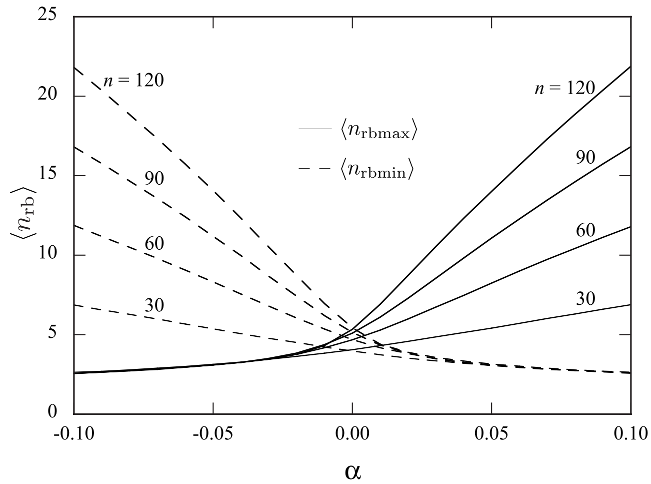

FIG. 3. Mean numbers of record-breaking maximum values $\langle n_{\mathrm{rbmax}}\rangle$ and record-breaking minimum values $\langle n_{\mathrm{rbmin}}\rangle$ are given as functions of the trend slope α for $n=30$, 60, 90, and 120 values.

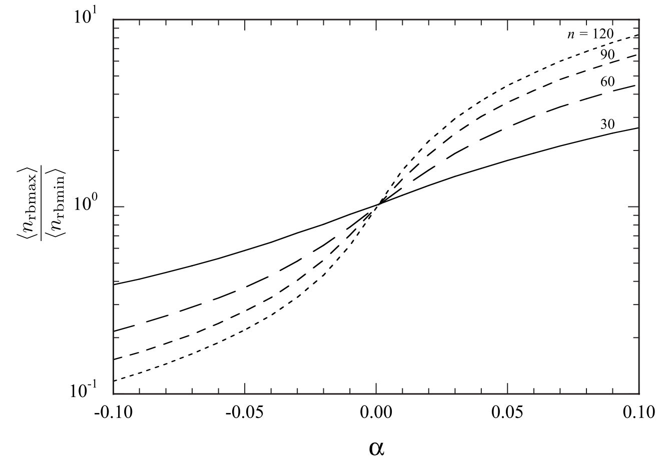

FIG. 4. Ratios of the mean number of record-breaking maximum values to the mean number of record-breaking minimum values $\langle n_{\mathrm{rbmax}}\rangle/\langle n_{\mathrm{rbmin}}\rangle$ are given as a function of the slope α for $n=30$, 60, 90, and 120 values.

For a temperature time series, we associate z with T and for our studies the increment in time between values is $\Delta t=1$ yr. From Eq. (14), we can write

$$\alpha=\frac{\langle z_i\rangle-\langle z_1\rangle}{i-1}\rightarrow\frac{1}{\sigma(T)}\Delta T, \tag{24}$$

where $\sigma(T)$ is the standard deviation of the temperature time series about its mean and ΔT is the change in temperature associated with the linear trend during one time step, that is, 1 yr. Therefore, $\Delta T\equiv(dT/dt)\Delta t$, where dT/dt is the linear trend in temperature and Δt is 1 yr. Our reference distribution for our Monte Carlo simulations will be a Gaussian white noise with a standard deviation $\sigma(z_i)=1$.

We utilize Monte Carlo simulations to determine the mean number of record-breaking events as a function of α. For each case, we run 1024 simulations and the mean number of record-breaking maximum values $\langle n_{\mathrm{rbmax}}\rangle$ and record-breaking minimum values $\langle n_{\mathrm{rbmin}}\rangle$ are given in Fig. 3 as functions of the trend slope α.

Results are given for $n=30$, 60, 90, and 120 values. For $\alpha=0$, we have $\langle n_{\mathrm{rbmax}}\rangle=\langle n_{\mathrm{rbmin}}\rangle$, and the values are identical to the i.i.d. random variable results given in Eq. (5). The mean numbers of record-breaking maximum values $\langle n_{\mathrm{rbmax}}\rangle$ for a positive value of α are equal to the mean numbers of record-breaking minimum values $\langle n_{\mathrm{rbmin}}\rangle$ for the same amplitude but negative value of α. We observe, as α changes sign, that the roles of temperature maxima and minima are reversed yielding the symmetry present in Fig. 3. Of particular interest are the ratios of the mean numbers of record-breaking maximum values to the mean number of record-breaking minimum values $\langle n_{\mathrm{rbmax}}\rangle/\langle n_{\mathrm{rbmin}}\rangle$. These are given as a function of α in Fig. 4 with $n=30$, 60, 90, and 120 values.

The results are sensitive to both the slope α and the length of the record n. The symmetry observed in Fig. 3 is reflected in Fig. 4—since the roles of the numerator and denominator are interchanged when α changes sign, and this particular symmetry is mirrored in the logarithmic representation employed here.

In the results given above, we have determined the mean cumulative number of record breaking values that occur during n events, namely, $\langle n_{\mathrm{rbmax}}\rangle$ and $\langle n_{\mathrm{rbmin}}\rangle$. We now turn our attention to the mean number of record breaking values that occur during the nth event, namely, $\langle\Delta n_{\mathrm{rbmax}}\rangle$ and $\langle\Delta n_{\mathrm{rbmin}}\rangle$. Specifically, we will consider the ratio $\langle\Delta n_{\mathrm{rbmax}}\rangle/\langle\Delta n_{\mathrm{rbmin}}\rangle$ associated with the nth event. In terms of record-breaking temperatures, this is the ratio of $\langle\Delta n_{\mathrm{rbmax}}\rangle/\langle\Delta n_{\mathrm{rbmin}}\rangle$ during the nth year after the start of the temperature time series. We will compare our results with the data on record-breaking temperatures given by Meehl *et al.* [3]. The values of $\langle\Delta n_{\mathrm{rbmax}}\rangle/\langle\Delta n_{\mathrm{rbmin}}\rangle$ are given in Fig. 5 as a function of n for $\alpha=0.005$, 0.01, 0.02, and 0.03.

In order to obtain smooth results from our Monte Carlo simulations, $2^{20}=1\,048\,576$ realizations of the Gaussian process were employed for each value of α and n. We performed

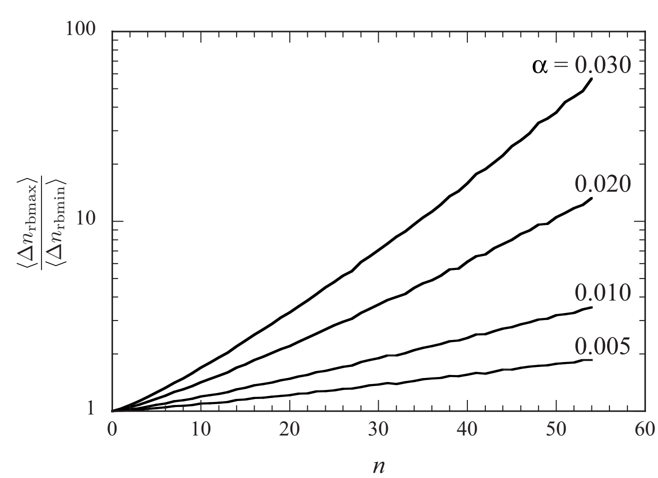

FIG. 5. Ratios of the mean number of record-breaking maximum values to the mean number of record-breaking minimum values $\langle\Delta n_{\mathrm{rbmax}}\rangle/\langle\Delta n_{\mathrm{rbmin}}\rangle$ during year n are given as a function of n for slopes $\alpha=0.00$, 0.01, 0.02, and 0.03.

an exponential fit to our Monte Carlo simulations and obtained the exponential relation

$$\frac{\langle \Delta n_{\mathrm{rbmax}} \rangle}{\langle \Delta n_{\mathrm{rbmin}} \rangle} = \exp[2.34\alpha n]. \tag{25}$$

The agreement of Eq. (25) with the simulations shown in Fig. 5 is quite good, with regression coefficients in excess of 0.998. The ratio $\langle \Delta n_{\mathrm{rbmax}} \rangle / \langle \Delta n_{\mathrm{rbmin}} \rangle$ has a very strong super-exponential dependence on the length of the record n.

V. RECORD-BREAKING TEMPERATURES AT THE MAUNA LOA OBSERVATORY

We now apply the results we have obtained to temperature observations obtained at the MLO, Big Island, Hawaii. This is a NOAA benchmark observatory where high-quality hourly temperature measurements have been carried out since 1977 [13]. An advantage of this data set is that the temperature measurements have a precision of 0.1 °C. Most temperature measurements in other data sets have been carried out to the nearest °F or °C and record-setting values are often duplicated in subsequent years.

The MLO is at an altitude of 3397 m above sea level and provides a wide range of atmospheric data that are relatively unperturbed by continental and local (e.g., "heat island") anthropogenic activities [13]. The observations that established the systematic increase in CO_2 were carried out at the MLO [11]. The well-mixed atmosphere at this isolated high-elevation observatory has been widely taken as representative of global average values.

We utilized the hourly data to obtain daily maximum and minimum temperatures. A day is defined to begin at 10:00 local time and end at 09:00 local time the next day. It is important to note that 3.7% of the hourly data for the 30 yr period are missing. Interpolation was used to replace the missing data. If one to three successive days were missing, the values for the adjacent days were averaged. If more days were missing, the data for the adjacent years for that day were averaged.

We analyzed temperature data collected between January 1, 1977 and December 31, 2006. We first considered the maximum and minimum temperatures for each day of the year numbered sequentially from 1 to 365. We did not include temperatures on February 29 of leap years. We considered the 30 values for the years considered and obtain the maximum likelihood (least-squares) best-fit linear trend to these daily values. The results are given in Fig. 6.

As expected, there is a considerable scatter; however, the trends for the minimum temperatures are consistently more positive (warming) than the maximum temperatures. In Fig. 7, we give the variance of the daily values about the linear trend line given in Fig. 6.

We observe that the variance is greater for temperature maxima than for temperature minima, and that there is a pronounced seasonal effect with enhanced variation during the winter months. The standard deviation for the maximum temperature is $\delta_{\max} = \sqrt{6.038} = 2.457$ °C, and the mean standard deviation for the minimum temperature is $\delta_{\min} = \sqrt{3.221} = 1.795$ °C. We will require these values when we compare trend results to record-breaking simulation results. The mean variance of the maximum temperature trends is significantly larger than the mean variance of the minimum temperature trends.

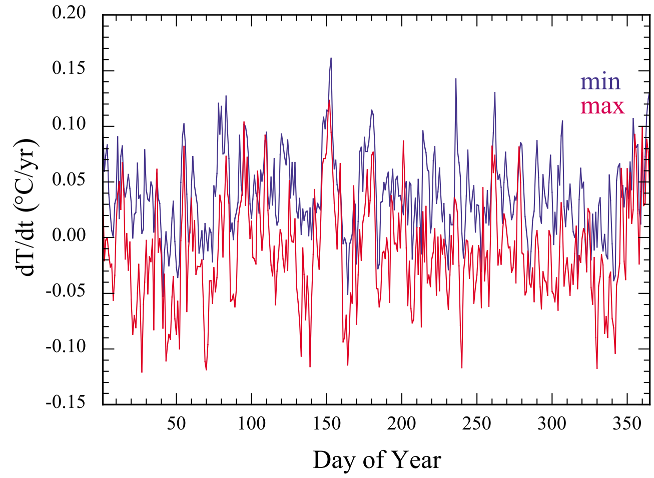

FIG. 6. (Color online) The best-fit linear temperature trends dT/dt for 30 yr are given for the 365 days of the year. The trends of both maximum daily temperatures $dT_{\max}/dt$ (lower curve) and minimum daily temperatures $dT_{\min}/dt$ (upper curve) are given.

We next consider the annual mean maximum and minimum temperatures at the MLO for the 30 yr period of 1977–2006. These are given in Fig. 8 along with the least-squares best-fit linear trends.

The slope of the trend line for the mean maximum annual temperature is $dT_{\max}/dt = -0.0129$ °C yr^{-1}, a small amount of cooling. The slope of the trend line for the mean minimum annual temperature is $dT_{\min}/dt = 0.0388$ °C yr^{-1}, a moderate amount of warming. We take the mean warming at this station to be the average of these values, $dT/dt = 0.0130$ °C yr^{-1}.

The diurnal temperature range (DTR) is the difference between maximum and minimum temperatures in a day (i.e., a given 24 h period). We will call the annual mean DTR the

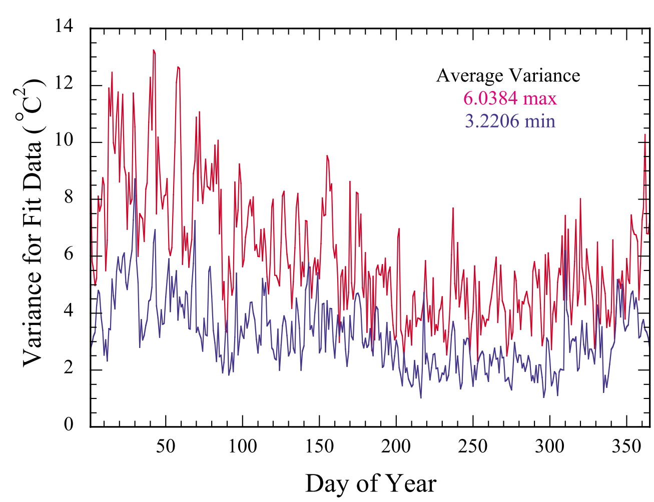

FIG. 7. (Color online) The variance of the temperature values about the linear trends in Fig. 1 is given for both the maximum (upper curve) and minimum (lower curve) daily temperatures.

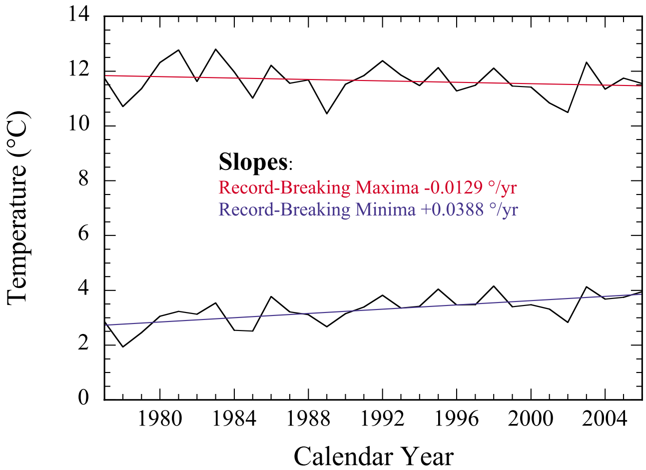

FIG. 8. (Color online) The mean annual maximum and minimum temperatures are given as functions of time from 1977 to 2006.

mean of the daily DTR values for the year, which is equivalent to taking the difference between (a) the mean of the daily maximum values in a year and (b) the mean of the daily minimum values for the same year. Subtracting the two trends shown in Fig. 8, we find that the trend in DTR is $d(\mathrm{DTR})/dt = -0.0517\ ^\circ\mathrm{C\ yr}^{-1}$. The temperature range is systematically decreasing. Our values are also tabulated in Table I. Giambelluca *et al.* [16] studied temperature records from 21 stations in Hawaii for the period of 1975–2006. For four high-altitude stations (elevations >900 m), they found a yearly DTR change $d(\mathrm{DTR})/dt = -0.036\ ^\circ\mathrm{C\ yr}^{-1}$, somewhat less than the value we obtain. The relevance of these two values to global warming will be discussed after we consider record-breaking temperatures.

Clearly, there is a scatter in the annual temperature data presented in Fig. 8. We will now determine whether the record-breaking analysis is consistent with the trend data given above. We determine the number of record-breaking maximum and record-breaking minimum temperatures.

We next consider the record-breaking statistics for the MLO temperature data. We consider the time series of 30 maximum and 30 minimum temperatures for each day of the year, beginning on January 1, 1977. We obtain the cumulative numbers or record-breaking temperatures. The 365 values are averaged to give $\langle n_{\mathrm{rbmax}} \rangle$ and $\langle n_{\mathrm{rbmin}} \rangle$ as functions of year from 1997 to 2006. The results are given in Fig. 9.

Also included in Fig. 9 is the number of record-breaking temperatures for an i.i.d. random process as given in Eq. (5). Importantly, the variance calculated in Eq. (9) provides a measure of the degree of variability that we can expect to see in a single realization of a stochastic process from its expected record-breaking statistics. However, when we employ data averaged over many, i.e., 365, realizations, the measure of fluctuations should be expected to diminish by a factor of $\sqrt{365} \approx 19$, bringing observations into much closer agreement with i.i.d. random variable theory if we are observing an independent and identically distributed process. This is shown in the inset plot in Fig. 9, whose function is to demonstrate that the departures observed in the low-temperature data from i.i.d. theory are unacceptably large. There, the i.i.d. theory result is shown as a solid line. Flanking this curve on each side are dashed lines showing departures of one, two, and three standard errors, respectively. Since the record-breaking random variable $n_{\mathrm{rb}}(n)$ is approximately Gaussian, we have a relatively quantitative measure present for the

TABLE I. dT/dt and DTR trends represented by $d(\mathrm{DTR})/dt$ with units of °C yr^{-1}.

Mauna Loa Observatory, Hawaii 1977–2006 Direct measurements of temperature trends [13]	
dT_{max}/dt	−0.0129
dT_{min}/dt	0.0388
dT/dt	0.0130
$d(\mathrm{DTR})/dt$	−0.0517
Temperature trends inferred from record-breaking statistics (this paper)	
dT_{max}/dt	−0.0091
dT_{min}/dt	0.0381
dT/dt	0.0145
$d(\mathrm{DTR})/dt$	−0.0472
United States 2000 stations 1950–2004 Record-breaking temperatures per year [3]	
dT/dt	0.025
Global land surface 1950–2004 Direct measurements of temperatures trends [1]	
dT_{max}/dt	0.014
dT_{min}/dt	0.020
dT/dt	0.017
$d(\mathrm{DTR})/dt$	−0.006

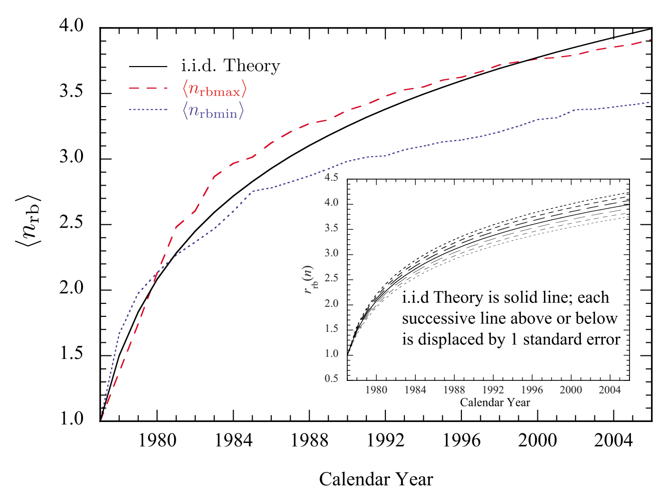

FIG. 9. (Color online) The average number of record-breaking maximum and minimum temperatures, $\langle n_{\mathrm{rbmax}} \rangle$ (dashed line) and $\langle n_{\mathrm{rbin}} \rangle$ (dotted line), as a function of time measured forward from January 1, 1977. The i.i.d. theory is also shown (solid line). The average is over the 365 days of the year. Also included as an inset is the number expected for an i.i.d. random process from Eq. (5).

quality of agreement between the observational data and i.i.d. theory. The number of record-breaking maximum temperatures is in reasonably good agreement with the i.i.d. random variable theory indicating the absence of a trend. The numbers of record-breaking minimum temperatures are substantially less than the i.i.d. random variable results indicating a warming trend. These results are in qualitative agreement with the direct measurements of trends given in Fig. 8. We will now provide a further quantification of the extent of this agreement.

To make a comparison between the directly measured temperature trends given in Fig. 8 and the number of record-breaking temperatures given in Fig. 9, we will consider the full 30 yr record of 1977–2006. For this period, we have $\langle n_{\text{rbmax}}\rangle = 3.90$ values and $\langle n_{\text{rbmin}}\rangle = 3.42$ values. The i.i.d. random value for this period from Eq. (5) is $\langle n_{\text{rbiid}}\rangle = 4.00$ values.

The influence of a linear trend on the mean number of record-breaking events was given in Fig. 3. We will utilize the simulation data given in Fig. 3 for a sequence of $n=30$ events. For small values of α, the simulations can be approximated by

$$\langle n_{\text{rbmax}}\rangle = 4.00 + 27.2\alpha, \qquad (26)$$

$$\langle n_{\text{rbmin}}\rangle = 4.00 - 27.2\alpha. \qquad (27)$$

Taking $\langle n_{\text{rbmax}}\rangle = 3.90$, we find from Eq. (26) that $\alpha_{\max} = -0.0037$. Furthermore, taking $\langle n_{\text{rbmin}}\rangle = 3.42$, we find from Eq. (27) that $\alpha_{\min} = 0.0213$.

In order to obtain the inferred linear temperature trends, we use Eq. (15). We first consider the inferred trend for the maximum temperatures. From Fig. 7, we obtain $\sigma_{\max} = \sqrt{6.04} = 2.45$ °C. Substituting this value and $\alpha_{\max} = -0.0037$ into Eq. (24), we find that $dT_{\max}/dt = -0.0091$ °C yr^{-1}. This compares with the directly measured value in Fig. 8, which is $dT_{\max}/dt = -0.0129$ °C yr^{-1}. We next consider the inferred trend for the minimum temperatures. From Fig. 7, we obtain $\sigma_{\min} = \sqrt{3.22} = 1.79$ °C. Substituting this value and $\alpha_{\min} = 0.0213$ into Eq. (24), we find that $dT_{\min}/dt = 0.0381$ °C yr^{-1}. This compares with the directly measured value in Fig. 8, which is $dT_{\min}/dt = 0.0388$ °C yr^{-1}. The inferred mean warming is $dT/dt = 0.0145$ °C yr^{-1} and the trend in the DTR is $d(\text{DTR})/dt = -0.0472$ °C yr^{-1}. These values are also given in Table I. The agreement between the direct measurements of temperature trends and the values inferred from the record-breaking statistics is quite good.

VI. RECORD-BREAKING TEMPERATURES IN THE U.S.

We next apply the results we have obtained to a comprehensive study of record-breaking temperatures in the United States. Meehl *et al.* [3] utilized data from nearly 2000 NCDC U.S. COOP network stations. The maximum and minimum daily temperatures were considered for the period of 1950–2004. For each day of the year, the occurrences of record-breaking values were determined. The numbers of record-breaking maximum and minimum temperatures were tabulated. In Fig. 10, we give numbers of record-breaking maximum temperatures to the numbers of record-breaking minimum temperatures $\langle \Delta n_{\text{rbmax}}\rangle / \langle \Delta n_{\text{rbmin}}\rangle$ for the period of 1950–2004 using the results of [3].

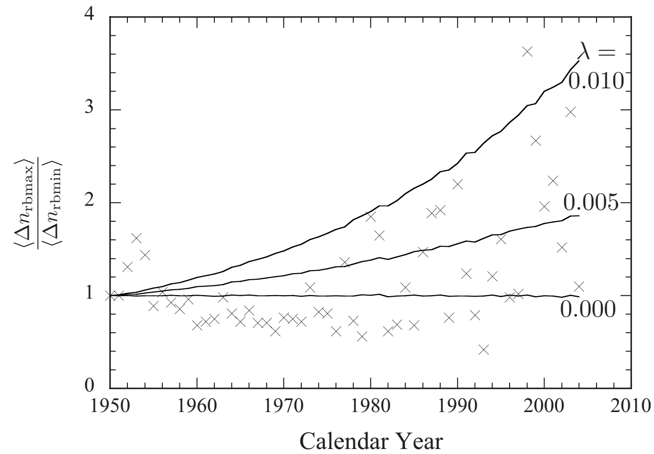

FIG. 10. Ratios of the numbers of record-breaking maximum temperatures to the numbers of record-breaking minimum temperatures $\langle \Delta n_{\text{rbmax}}\rangle / \langle \Delta n_{\text{rbmin}}\rangle$ on a yearly basis for nearly 2000 stations in the U.S. as given by [3]. Also included are our simulation results from Fig. 5 for $\lambda = 0$, 0.005, and 0.010.

There is clearly a lot of scatter but an upward trend does appear, particularly in the largest values. These data can be directly compared with our simulation results given in Fig. 5. The results for $\lambda = 0.000$, 0.005, and 0.010 are given in Fig. 10. In order to obtain a temperature trend, we require a numerical value for the standard deviation of the temperature time series. The appropriate value has considerable variability. We take the value $\delta = 5$ °C given by Redner and Petersen [8] to be representative of continental stations. Taking $\lambda = 0.005$ and $\delta = 5$ °C in Eq. (15), we obtain $dT/dt = 0.025$ °C yr^{-1}. This is about a factor of 2 larger than our values for MLO. However, the smaller value cannot be ruled out because of the large scatter.

VII. DISCUSSION

The object of this paper has been to systematically study the statistics of record-breaking events with an emphasis on record-breaking temperatures. We first studied a Gaussian white-noise time series as a first-order model of the temperature time series for a specified day of the year at a specified weather station. The maximum and minimum temperatures constitute a yearly time series with a specified starting date. This is an i.i.d. random process, and the number of record-breaking temperatures increases approximately as the natural logarithm of the number of years n as given by Eq. (5). This result was first given by Glick [6].

Studies have shown that temperature time series exhibit long-range correlations that can be approximated as a fractional Gaussian noises [9]. We have studied the influence of long-range correlations on record-breaking statistics using Monte Carlo simulations. For the range of fractional Gaussian noises applicable to temperature time series, the influence on the record-breaking statistics is small.

The current interest in record-breaking temperatures is associated with global warming [7,8]. Specifically, greater numbers of record-breaking maximum temperatures are compared to the numbers of record-breaking minimum temperatures. In our Monte Carlo simulations, we superimposed linear temperature trends on a Gaussian white noise of yearly temperature values on a specified day of the year and a specified weather station. We found the single governing parameter to be the ratio of (a) the temperature change in a year to (b) the standard deviation of the temperature white noise as defined in Eqs. (14) and (24). We found an exponential sensitivity to the number of years in the record as given in Eq. (25).

In order to test our analyses, we have considered two data sets. The first is the 30 yr record of temperatures at the Mauna Loa Observatory, Hawaii. This station plays a unique role in global warming studies as the site that established the systematic increase in anthropogenic CO_2. The well-mixed atmosphere at this isolated high-elevation site has been taken to be representative of global average values. We determine the best-fit linear trends to the annual maximum and minimum temperatures for 1977–2006. We also infer temperature trends from the numbers of record-breaking maximum and minimum temperatures during the period. The directly measured and inferred trends are in good agreement with each other as shown in Table I.

It is of interest to compare our MLO values with globally averaged values. Many studies of global heating have been carried out and these have been summarized in Global Change 2007 [1]. Trend values for global land surface temperatures for the period of 1950–2004 are $dT_{max}/dt = 0.014\ °\mathrm{C\ yr}^{-1}$, $dT_{min}/dt = 0.020\ °\mathrm{C\ yr}^{-1}$, and $d(\mathrm{DTR})/dt = -0.006\ °\mathrm{C\ yr}^{-1}$. The mean trend $dT/dt = 0.017\ °\mathrm{C\ yr}^{-1}$ is close to our values of $dT/dt = 0.0130°$ and $0.0145\ °\mathrm{C\ yr}^{-1}$ as shown in Table I. The global measurements have a smaller decrease in DTR. This can possibly be attributed to buffering of low-elevation continental data versus the high-elevation MLO data.

We also consider the yearly ratios of record-breaking maximum to minimum temperatures for some 2000 U.S. stations for the period of 1950–2004 using results given by [3]. Taken together, we have shown using the theory of record-breaking statistics that stationary fluctuations alone cannot explain the temperature observations, but that a linear trend, particularly in the low-temperature data, is required to explain the observations.

ACKNOWLEDGMENTS

We wish to thank the referees for some helpful comments during the revision of this manuscript.

[1] S. D. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis, K. B. Averyt, M. Tignor, and H. L. Miller, *Climate Change 2007: The Physical Science Basis. Contribution of Working Group 1 to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* (Cambridge University Press, Cambridge, UK, 2007).

[2] T. R. Karl, P. D. Jones, R. W. Knight, G. Kukla, N. Plummer, V. Razuvayev, K. P. Gallo, J. Lindseay, R. J. Charlson, and T. C. Peterson, Bull. Am. Meteorol. Soc. **74**, 1007 (1993).

[3] G. A. Meehl, C. Tebaldi, G. Walton, D. Easterling, and L. McDaniel, Geophys. Res. Lett. **36**, L23701 (2009).

[4] V. B. Nezvorov, Theor. Probab. Appl. **32**, 201 (1986); B. Arnold, N. Balakrishnan, and H. N. Nagaraja, *Records* (John Wiley and Sons, New York, 1998).

[5] M. N. Tata, Wahrscheinlich Leit. **12**, 9 (1969).

[6] N. Glick, Am. Math. Monthly **85**, 2 (1978).

[7] R. E. Benestad, Clim. Res. **25**, 3 (2003); Glob. Planet. Change **44**, 11 (2004); EOS Trans. Am. Geophys. Union **89**, 389 (2008).

[8] S. Redner and M. R. Petersen, Phys. Rev. E **74**, 061114 (2006).

[9] E. Koscielny-Bunde, A. Bunde, S. Havlin, and Y. Goldreich, Physica A **231**, 393 (1996); J. D. Pelletier, J. Climate **10**, 1331 (1997); E. Koscielny-Bunde, A. Bunde, S. Havlin, H. E. Roman, Y. Goldreich, and H. J. Schellnhuber, Phys. Rev. Lett. **81**, 729 (1998); J. D. Pelletier and D. L. Turcotte, Adv. Geophys. **40**, 91 (1999); R. O. Weber and P. Talkner, J. Geophys. Res. **106**, 20131 (2001); K. Fraedrich and R. Blender, Phys. Rev. Lett. **90**, 108501 (2003); J. F. Eichner, E. Koscielny-Bunde, A. Bunde, S. Havlin, and H. J. Schellnhuber, Phys. Rev. E **68**, 046133 (2003).

[10] Carbon Dioxide Information Analysis Center, Atmospheric CO_2 concentrations (ppmv) derived from *in situ* air samples collected at Mauna Loa Observatory, Hawaii; available online at http://cdiac.ornl.gov/trends/co2/maunaloa.co2

[11] C. D. Keeling, R. B. Bacastow, A. E. Ekdahl, C. A. Guenther, L. S. Waterman, and J. F. S. Chin, Tellus **28**, 538 (1976).

[12] S. Lennartz and A. Bunde, Geophys. Res. Lett. **36**, L16706 (2009).

[13] National Oceanic and Atmospheric Administration (NOAA), Earth System Research Laboratory (ESRL), Mauna Loa hourly temperature data for 1 January 1997 to 7 February 2007; available online at ftp.cmdl.noaa.gov/met/

[14] W. Feller, *An Introduction to Probability Theory and Its Applications*, 3rd ed. (John Wiley and Sons, New York, 1968), Vol. 1, p. 229.

[15] D. L. Turcotte, *Fractals and Chaos in Geology and Geophysics*, 2nd ed. (Cambridge University Press, Cambridge, UK, 1997); B. D. Malamud and D. L. Turcotte, Adv. Geophys. **40**, 1 (1999).

[16] T. W. Giambelluca, H. F. Diaz, and M. S. A. Luke, Geophys. Res. Lett. **35**, L12702 (2008).