Climatic Change (2012) 113:1001–1024
DOI 10.1007/s10584-011-0370-8

Climate change: a new metric to measure changes in the frequency of extreme temperatures using record data

Lalith Munasinghe · Tackseung Jun · David H. Rind

Received: 25 November 2010 / Accepted: 16 November 2011 / Published online: 10 December 2011
© Springer Science+Business Media B.V. 2011

Abstract Consensus on global warming is the result of multiple and varying lines of evidence, and one key ramification is the increase in frequency of extreme climate events including record high temperatures. Here we develop a metric—called "record equivalent draws" (RED)—based on record high (low) temperature observations, and show that changes in RED approximate changes in the likelihood of extreme high (low) temperatures. Since we also show that this metric is independent of the specifics of the underlying temperature distributions, RED estimates can be aggregated across different climates to provide a genuinely global assessment of climate change. Using data on monthly average temperatures across the global landmass we find that the frequency of extreme high temperatures increased 10-fold between the first three decades of the last century (1900–1929) and the most recent decade (1999–2008). A more disaggregated analysis shows that the increase in frequency of extreme high temperatures is greater in the tropics than in higher latitudes, a pattern that is not indicated by changes in mean temperature. Our RED estimates also suggest concurrent increases in the frequency of both extreme high and

Electronic supplementary material The online version of this article (doi:10.1007/s10584-011-0370-8) contains supplementary material, which is available to authorized users.

L. Munasinghe
Department of Economics, Barnard College, Columbia University, 3009 Broadway, New York, NY 10027, USA

T. Jun (✉)
Department of Economics, Kyung Hee University, 1 Hoegi-dong, Dongdaemoon-gu, Seoul, 130-701, Korea
e-mail: tj32gm@gmail.com

D. H. Rind
NASA Goddard Institute for Space Studies, 2880 Broadway, New York, NY 10025, USA

extreme low temperatures during 2002–2008, a period when we observe a plateauing of global mean temperature. Using daily extreme temperature observations, we find that the frequency of extreme high temperatures is greater in the daily minimum temperature time-series compared to the daily maximum temperature time-series. There is no such observable difference in the frequency of extreme low temperatures between the daily minimum and daily maximum.

1 Introduction

Extreme temperatures are not only an important dimension of climate but there is wide spread recognition that increased frequency of extreme temperatures leads to major societal impacts including adverse health effects (Githeko and Ndegwa 2001; Curriero et al. 2002; Hay et al. 2002) and diminished crop productivity (Ciais et al. 2005; Battisti and Naylor 2009). In this paper we present a global assessment of climate change by analyzing changes in extreme high and extreme low temperatures over the past century. We do so by constructing a metric on the basis of record theory and temperature record data to measure changes in the likelihood of extreme temperatures.

The idea that the empirical frequency of record breaks can be exploited to make inferences about changes in the underlying distribution is certainly not new (Yang 1975; Ballerini and Resnick 1985; Munasinghe et al. 2001). The basics of record theory (Arnold et al. 1998) imply that if the underlying distribution is time invariant—that is, independent and identically distributed (i.i.d.)—then the likelihood of a record break in a given time period is simply the inverse of the elapsed time periods over which records are kept. Put simply, the chance of a record break in the second time period is 1/2, in the third period it is 1/3, etc. Clearly, with the passage of time it becomes increasingly difficult to break a record from a stationary distribution (Rényi 1962). The actual pattern of temperature record breaks over the last one hundred years, however, does not match this pattern predicted by a stationary distribution. As a matter of fact, the 10 hottest years—in terms of recorded annual global mean temperature—occurred between 1998–2010. The discrepancy between the observed and predicted record break frequencies of course suggests a change in the underlying temperature distribution. For example, a higher than predicted observed frequency of high temperature record breaks implies a positive shift in the underlying temperature distribution.

The technical challenge is how to measure a shift in the underlying distribution that would be consistent with the discrepancy between the observed and predicted record break frequencies. Our solution is to estimate an implied sampling rate—what we call "record equivalent draws" (RED)—from the same distribution that would match the empirical frequency of record breaks. The idea is that we pretend to sample more (or less) random draws from the same original distribution until the predicted record break rate equals the higher (or lower) observed record break rate. Note that since the record break process for any given time-series is independent of the specifics of the underlying distribution, the empirical record break frequency (for given time period) can be constructed from observations that have different underlying distributions. We link RED to extreme outcomes by showing that an

increase (decrease) in RED converges to a corresponding increase (decrease) in the frequency of an extreme outcome.

Climate scientists have addressed global warming not only by analyzing mean temperature changes but also by analyzing changes in extreme temperatures using a variety of statistical methods. Meehl et al. (2009) analyzed the ratio of the number of daily high and daily low temperature record breaks (using U.S. data) and showed that this ratio is around 2 in the most recent decade. They explain this finding by arguing that high temperature record breaks are declining by less than what is predicted by a stationary distribution since the late 1970s. Benestad (2003) examined the record break frequency of global mean temperatures over the past century and showed that the number of actual high temperature record breaks is larger than the theoretically predicted number under the assumption of a stationary distribution. This disparity in the actual versus predicted number of high temperature record breaks rejects the "no warming" hypothesis.

Methodologically, the RED metric is closely related to extreme value theory (EVT). The key result of EVT is that the maximum from a sufficiently large number of draws form a well-defined generalized extreme value (GEV) distribution (Gumbel 1958). Several recent studies (Zwiers and Kharin 1998; Kharin and Zwiers 2000; Nogaj et al. 2006; Yiou et al. 2008) have used temperature time-series data to fit a GEV distribution by estimating location and scale parameters that characterize the tails of the density function. Changes in the underlying distribution of the tails are inferred on the basis of estimated changes in these parameters using either parametric or nonparametric methods (Hall and Tajvidi 2000). Some studies have also used the concept of return period to assess the risk of extreme events based on GEV constructions (Kharin and Zwiers 2000; Laurent and Parey 2007; Zwiers et al. 2011).

Several other studies have analyzed trends of various indices related to extreme temperatures including daily minimum and maximum temperatures (Karl et al. 1993; Easterling et al. 1997; Bonsal et al. 2001; Frich et al. 2002; Griffiths et al. 2005; Alexander et al. 2006; Choi et al. 2009; Zwiers et al. 2011). The basic methodology here is to estimate a trend line of say the daily minimum temperature for each weather station and then compute the fraction of weather stations (across regions or seasons) that have a significant positive or negative trend line. The overall findings from these trend analyses (and also from EVT studies) are that the frequency of extreme warm events have increased while the frequency of extreme cold events have decreased in the second half of the last century, and that there is a larger increase in the minimum temperatures compared to the maximum temperatures.

We hope that the RED metric proposed in this paper and our findings will complement the methods and findings from these earlier studies. The major difference between our metric and previous estimation methods is the distribution-free property of RED. For example, changes in mean temperature can of course be highly informative of local changes in temperature, but aggregating mean changes in temperature across climate environments as diverse as the tropics and Antarctica is more challenging. Moreover, the mean temperature in any location in the world does not follow the global average (Trenberth et al 2007). The RED metric can address whether warming (or cooling) is truly global, and it also allows a comparison of changes in the frequency of extreme temperatures across vastly different climates as represented by geographic regions and seasons.

2 Methods

We begin the methods section with a non-technical illustration of RED. In the more formal analysis to follow we present the basic record framework and then show how we can carefully ignore the i.i.d. assumption in order to incorporate changes in the underlying distribution. Next we detail the empirical estimation of RED by maximum likelihood techniques. To link the RED estimates to changes in extreme temperatures, we prove that changes in RED converge asymptotically to changes in the frequency of extreme temperatures. Finally, we highlight the inefficiency of our current estimation of RED using only record break information, and suggest an alternative estimation strategy using the rank ordering of temperature observations.

2.1 A non-technical illustration of RED

How do we measure changes in the frequency of extreme temperatures? We will illustrate our approach and the RED metric with a simple example. Assume a large and fixed number of weather stations report the temperature on January 1 at 12 Noon annually from 1900 to 2008. If the underlying temperature distributions remain unchanged then the temperature observations from each station in 1900 and 1901 will have the same probability of being the "record" high temperature. Hence we would expect 50% of all stations to report a record high temperature in 1901. However, suppose 75% of the stations in fact report a record high temperature. This would suggest the underlying distribution in 1901 is warmer. The key question is, "how do we quantify this warming" Consider the following hypothetical scenario: whereas the recorded temperature in 1900 was from a single "draw",[1] suppose the recorded temperature in 1901 in each station was the maximum of three independent draws from the identical 1900-temperature distribution. Note of course that the implied 1901 distribution is clearly different from the 1900 distribution since the former distribution is the maximum of three draws from the latter distribution. But more importantly, 1901 represents a warmer climate in the sense that a temperature reporting in 1901 is more likely to exceed any given temperature than a temperature reporting in 1900. To get back to our hypothetical scenario, if this were the case then 75% of the stations would be expected to report a record high temperature in 1901. Now suppose in 1902, 50% of the stations report a temperature higher than all previous years' temperatures. Given that the total draws in 1900 and 1901 add up to four (one in 1900 and three in 1901), a total of four hypothetical draws from the 1900-temperature distribution in 1902 are required to predict that 50% of the stations will report a record high temperature in 1902 compared to all previous years. By proceeding similarly to 2008, the number of hypothetical draws for each year can be computed using the observed fraction of record high temperatures in each year across all the weather stations.

It is this number of hypothetical draws that we call "record equivalent draws" (RED). To be terminologically specific, if the records refer to high (low) temperature

[1]The term "draw" in this paper refers to the sampling from a distribution as opposed to the sometimes British English usage of the term to refer to a tie.

records then we call it RED for high (low) temperature, and refer to it as "RED-H" (RED-L) for short. In Section 2.2.4 below we formally show that the change in RED-H and RED-L approximate the change in the frequency of "extreme" high and "extreme" low temperatures, respectively.

2.2 Derivation and estimation of RED

2.2.1 Standard results from record theory

Let $k \in \{(i, j)\}$ denote a composite index for month i and grid j. We start with the single-draw framework that implies a stationary temperature distribution such that for each k, a single temperature observation is drawn from a continuous temperature distribution function $F^k(.)$ in each year t.[2] Let $\{x_t^k\}$ denote a sequence of such temperature draws that are independent random variables for each k, where $t = 1, 2, \ldots$. We call a temperature observation a "record high temperature" at time t for k if it is the highest temperature observed up to time t. We can similarly define a record low temperature at time t for k. The first temperature draw is trivially a high and low record temperature. Under these assumptions it does not matter whether we talk about record highs or record lows (Chandler 1952), and hence we refer only to record high temperatures in the following exposition.

To formally define a record, first denote V_r^k as the r-th time for record high temperature in k and define it inductively as follows:

$$V_r^k = \min\left\{t | t > V_{r-1}^k, x_t^k > x_{V_{r-1}^k}^k\right\}, V_1^k = 1.$$

And then denote e_t^k as the record high indicator for k at year t and define it as follows: $e_t^k = 1$ if for some r, $V_r^k = t$; and $e_t^k = 0$ otherwise.

The key result (Rényi 1962) we use is that under the i.i.d. assumption of the temperature distribution the record indicators $e_2^k, e_3^k, \ldots$, are independent Bernoulli variables with $\Pr\left(e_t^k = 1\right) = 1/t$ for each k. The intuition is that a new record is set at time t if and only if x_t^k is the maximum of the first t realizations for an given k. Since each temperature observation is drawn from the same stationary distribution, each realization has the identical probability of being the maximum. This probability is independent of the underlying temperature distribution and hence does not depend on the specifics of the distribution such as the mean or the standard deviation. This feature, due to the nonparametric nature of record processes, will be used to aggregate record observations from different months and geographic regions, i.e., across different month-grid pairs. Therefore from now on we drop the index k from the definition of the temperature distribution and simply denote it as $F(.)$.

2.2.2 Modeling changes in the underlying distribution

Here we introduce the idea of time-varying multiple draws in order to drop the i.i.d. assumption and incorporate changes in the underlying distribution. Let $F_t(.)$ denote

[2]We assume that the temperature distribution is continuous. Otherwise we need to deal with ties of temperature observations that happen with positive probability, which unnecessarily complicates the presentation of Rényi (1962) result below.

the temperature distribution at time t so that the distribution is now time dependent. For example, if time period s is warmer than time period t it means that $F_s(.)$ is in some sense a warmer distribution than $F_t(.)$. For some purposes it may be convenient to model warming as an increase in the mean of the distribution. However, mean-increasing change is not convenient for our purposes since our focus is on changes in extreme temperatures. Hence we will concentrate on what we call "logarithmically proportional" warming to incorporate the idea that the new distribution is generated by taking the maximum of multiple draws from the old distribution (Munasinghe et al. 2001). For $s > t$, we say that warming between t and s is logarithmically proportional if there is some positive constant $\gamma(t, s) > 1$ such that for all x,

$$\ln F_s(x) = \gamma(t, s) \ln F_t(x),$$

which reduces to:

$$F_s(x) = (F_t(x))^{\gamma(t,s)}.$$

By induction, this becomes:

$$F_s(x) = (F_1(x))^{\gamma(1,s)}. \tag{1}$$

If $\gamma(1, s) > 1$ implies warming then Eq. 1 says that observing a temperature lower than x is less likely at time s than at time 1.

Put differently, equation Eq. 1 shows that the distribution $F_s(x)$ is generated by taking the maximum of $\gamma(1, s)$ draws from the initial (or base) distribution $F_1(x)$. Therefore the number of draws at any time is measured relative to the number of draws in the base time period (which can be normalized to 1). Note that the temperature distributions in the base time period and the latter time period s are different even though the distribution at time s is a direct transformation of the base distribution. Moreover, since the multiple draws in any given time period are assumed to arrive from the same base distribution, Rényi's result from above holds in this multi-draw context: the record indicators $e_2^k, e_3^k, \ldots$, are independent Bernoulli variables with $\Pr\left(e_t^k = 1\right) = \gamma(1, t) / \sum_{\tau=1}^{t} \gamma(1, \tau)$ (See Nevzorov 1985 for a formal treatment of this result). As a consequence, the number of draws at any time period is independent of the base temperature distribution.

Note that EVT shares the same theoretical foundation with our approach up to this point. From here EVT is extended by establishing the asymptotic distribution of the tails for a sufficiently large number of draws. Subsequently, the changes in the distribution are examined by estimating changes in the parameters that characterize this asymptotic distribution under the assumption of a large value for $\gamma(t, s)$ for all t and s. By contrast, our approach infers changes in the distribution on the basis of the estimates of the number of these draws in each period that "best" fit the observed record break data.

2.2.3 Estimation of RED

Let c_t (RED) be the empirical counterpart of γ in year t, defined as the area-weighted (by the cosine of latitude) sum of record indicators over the twelve time series (representing the twelve months in a year) and grids in year t. Define $\lambda_t = \sum_{\tau=1}^{t} c_\tau$.

Then the probability that a new record will be set in year t is c_t/λ_t. If c_t is constant then this probability simplifies to $1/t$, the classical case with a constant draw rate. Our empirical strategy is to estimate the $\{c_t\}$ sequence by maximum likelihood estimation (MLE). Let n_t and b_t denote the area-weighted number of distinct temperature observations and the number of new records at year t, respectively. Define $f_t = b_t/n_t$. Note that f_t is an unbiased estimator of c_t/λ_t. Therefore $\{c_t\}$ can be estimated as follows: fix $\hat{c}_1 = 1$ and let:

$$\hat{c}_t = \frac{f_t}{1 - f_t}\hat{\lambda}_{t-1}, \text{ where } \hat{\lambda}_{t-1} = \textstyle\sum_{\tau=1}^{t-1}\hat{c}_\tau.$$

Denote $\mathbf{L}_t\left(\hat{c}\right)$ as the logarithm of the likelihood of observing precisely f_t proportion of new record high temperatures in year t with the vector of parameters $\hat{c} = \left(\hat{c}_1, \hat{c}_2, ..., \hat{c}_T\right)$, where T is the last year for each of the time series:

$$\mathbf{L}_t\left(\hat{c}\right) \approx b_t \ln \hat{c}_t + (n_t - b_t) \ln \hat{\lambda}_{t-1} - n_t \ln \hat{\lambda}_t. \tag{2}$$

We use maximum likelihood techniques to estimate the vector of $\hat{c}$ so as to maximize $\sum_{t=1}^{T}\mathbf{L}_t\left(\hat{c}\right)$.

For stations with temperature data that start at different time periods, we need to slightly modify the above procedure. We group the time series that start at time h as the time series of cohort h where $h \in \{1, 2, ..., T-1\}$. Then we compute the logarithm of the likelihood function defined by the Eq. 2 for each cohort, and aggregate all such functions to form the new likelihood function to maximize.

There is one technical issue with the observed temperature data sets that we still need to address. Since temperature data are recorded only up to a 10th of a decimal point we have several ties in our time-series. Ties in temperature observations violate the continuity assumption of the underlying temperature distribution. To resolve this issue of ties, we perform Monte Carlo type simulations. In each simulation ties are randomly broken prior to the estimation of RED. This experiment is repeated 1,000 times, and the mean and 95% confidence intervals (CI) of RED estimates from this 1,000 experiments are computed for each year.

2.2.4 Deriving the frequency of extreme temperatures from RED

The focus of this paper is not record temperatures per se, but rather the estimation of changes in the frequency of extreme temperatures over time. Hence we need to link the relationship between our RED estimates and the frequency of extreme temperatures. Here we show that the change in RED converges to the change in the frequency of a given temperature as this temperature goes to an extreme limit of the distribution.

First we normalize RED-H in the base year to one and assume that the RED-H in the current year is equal to $n > 0$. Suppose the distribution of the underlying temperature distribution in the base year is continuous, differentiable, and strictly increasing in the range $[\alpha, \beta]$, where α and β can be infinity. Let $\mathbf{Z}$ denote the temperature random variable with the corresponding cumulative distribution function (cdf) $F_{\mathbf{Z}}(.)$, and let $\mathbf{Z}_{(n)}$ denote the nth order statistic (i.e. the maximum of n draws) with the corresponding cdf denoted as $F_{\mathbf{Z}_{(n)}}(.)$. Assuming that $\mathbf{Z}_1, \mathbf{Z}_2, ..., \mathbf{Z}_n$ are from

the same initial distribution $F_Z(.)$ as given in Eq. 1,[3] the probability density of the nth order statistic is given by:

$$f_{\mathbf{Z}_{(n)}}(x) = \frac{dF_{\mathbf{Z}_{(n)}}}{dx}(x) = nF_Z^{n-1}(x)\frac{dF_Z(x)}{dx} = nF_Z^{n-1}(x)\, f_Z(x).$$

Hence the probability of observing a temperature greater than x from a maximum of n draws is given by:

$$g(x) = \int_{x\geq\alpha}^{\beta} nF_Z^{n-1}(z)\, f_Z(z)\, dz.$$

Similarly, the probability of observing a temperature greater than x from a single draw is given by:

$$h(x) = \int_{x\geq\alpha}^{\beta} f_Z(z)\, dz.$$

Then the ratio $g(x)/h(x)$ represents the change in the frequency of observing a temperature greater than x in the current year compared to the base year. Since both $g(x)$ and $h(x)$ are differentiable and converge to zero, the limit of the ratio $g(x)/h(x)$ as x approaches an "extreme" high value is obtained by L'Hospital's rule:

$$\lim_{x\to\beta}\frac{g(x)}{h(x)} = \lim_{x\to\beta}\frac{g'(x)}{h'(x)} = \lim_{x\to\beta}\frac{nF_Z^{n-1}(x)\, f_Z(x)}{f_Z(x)} = \lim_{x\to\beta} nF_Z^{n-1}(x) = n.$$

Therefore changes in RED-H approximate the change in frequency of observing an "extreme" high temperature. The same argument holds for the frequency of observing an extreme low temperature derived from changes in RED-L. Note that the observed maximum or minimum temperature doesn't necessarily correspond to this "extreme" temperature since the latter refers to the "limit" of the underlying distribution.

To complete the proof we need to show that the ratio approaches this limit *monotonically* as x approaches extreme values. We rewrite the ratio $g(x)/h(x)$ as follows:

$$\frac{g(x)}{h(x)} = \frac{\int_{x\geq\alpha}^{\beta} nF_Z^{n-1}(z)\, f_Z(z)\, dz}{\int_{x\geq\alpha}^{\beta} f_Z(z)\, dz} = \frac{1-F_Z^n(x)}{1-F_Z(x)} = \sum_{t=1}^{n} F_Z^{t-1}(x).$$

If $F(.)$ is strictly increasing then this ratio is also strictly increasing in x for any n, which completes the proof.

2.3 An alternative approach to estimate RED

One shortcoming of our current estimation of RED is the fact that we do not use all the information available in our temperature data. More specifically, we only use record break information and not also the magnitude of temperature changes from one time period to the next. Hence our estimates of RED are inefficient. The question is whether we can use this additional information while preserving

[3] As mentioned earlier, although we assume that the n draws come from the same initial distribution the transformed distribution by taking the maximum of n such draws clearly represents a change in the underlying distribution.

the distribution-free property of RED. One alternative approach is to rank order temperature observations, and then estimate a time series of RED that maximizes the likelihood of the rank order of the entire history of temperature observations.

This rank-order approach clearly incorporates more information than the record-break approach, though not quite the exact magnitudes of temperature changes. For instance, the temperature observation in the third year may be a record high among the first three temperature observations, but when considered against the entire history of temperature data, this high record may be also the third lowest. By using this additional fact, the estimated number of draws associated with this temperature observation will be smaller under the rank-order approach. In other words, the rank-order approach uses more information by looking at the overall standing of a temperature observation in relation to the entire time series of temperature observations.

The rank-order method still maintains all the basic properties of our current record-break approach, including the distribution-free property. Moreover, the rank-order method is especially useful when sample sizes are small which is likely to be the case when we study geographically restricted areas. Note that a record is less likely to be broken simply due to the passage of time, and this problem is further exacerbated when we have small sample sizes because we may not observe even a single record break toward the end of a long time series where the predicted probability of a record break is low. This problem can be avoided by using the rank ordering of temperature observations which, unlike record break information, does not depend on the time elapsed nor the sample size, i.e., temperature observations in a time series can be always ranked by its magnitude.

The downside of this rank-order approach is that it is very costly in terms of computation resources due to maximum likelihood routines that use multinomial regressions. The current RED estimates using record data only require binomial regressions. However, initial simulation results suggest that as long as we have sufficiently large sample sizes the RED estimates are very similar across these two estimation methods. In our future research we hope to implement the more efficient rank-order estimation routine in order to assess changes in extreme climate events.

3 Data

The two most widely used temperature data types are daily extremes and monthly average temperatures. The type of data that is more appropriate will typically depend on the kind of impacts we wish to study. We choose the monthly average land surface air temperature (LSAT) as our main data primarily due to the limited availability of high quality daily temperature data (especially in the tropics and in the early time periods of the last century). However we do compare RED estimates from the monthly average temperature data with RED estimates from the best-available daily extreme temperature data (see Section 4.4).

3.1 Monthly average temperature data

The monthly average temperature data we use for our main analysis are the $0.5^\circ \times 0.5^\circ$ gridded global network of time-series between 1900–2008 provided by the

Center for Climate Research at the University of Delaware (called here by DelCCR2). This data set is available from the University of Delaware website.[4] The DelCCR2 is constructed from major historical LSAT data including homogenized GHCN-Monthly version 2. The monthly average temperature in each grid is either directly observed from weather stations or interpolated from nearby stations (Willmott and Matsuura 1995).

We select this data set as our primary source over others (e.g. HadCRUT3v) for two specific reasons. First, it has a complete data record, whereas the other sources have missing data entries due to unreliable station data. Missing data can lead to biased RED estimates because we do not know whether an unobserved temperature is a new record or not, and hence whether a subsequent temperature in the same time series is a record or not. Second, the DelCCR2 data have better spatial coverage, especially for low latitude regions. However, the reason that the DelCCR2 does not have missing data entries is because it uses interpolation over much longer distances than compared to say HadCRUT3v. This implies that some of the data observations in the DelCCR2 are likely to have a larger uncertainty associated with them compared to HadCRUT3v. To address this issue we restrict our analysis to grids with latitudes higher than 24°S.

To construct the annual time-series of record high and record low temperatures for each month-grid pair, we create a $2° \times 2°$ grid network that contains (a maximum of) 16 monthly average temperatures from the original $0.5° \times 0.5°$ grid network. From each of these expanded grids we select the maximum and minimum monthly average temperatures to create the annual time-series of record high and record low temperatures, respectively.[5] In other words, we estimate RED-H for the maximum of monthly average temperatures ("RED-H Mx" for short), and RED-L for the minimum of monthly average temperatures ("RED-L Mn" for short).[6] Since there is no missing temperature information in DelCCR2, each grid has 12 record high and 12 record low time series corresponding to each month. In our sample we have 3, 996 $2° \times 2°$ grids geographically distributed as follows: 1,432 grids at latitudes higher than 55°N; 1,391 between 24°N and 55°N; 630 between 0° and 24°N; and 542 between 24°S and 0°.

For each month-grid pair we take 1929 as the base year and substitute its temperature with the average of the maximum (minimum) monthly average temperatures from 1900 to 1929 (for that month-grid). This thirty-year average temperature should be long enough to make valid comparisons with the future changes in the underlying distribution of temperatures. For the base year 1929 we normalize RED to 1 for both samples. Given this normalization RED should be interpreted as follows: if for example RED-H Mx in year 2000 is 10, then an extreme high temperature in terms of monthly averages is ten times more likely in year 2000 than it is in first three decades of the past century (See Section 2.2.4).

[4] http://climate.geog.udel.edu/~climate/html_pages/Global2_Ts_2009/README.global_t_ts_2009.html

[5] As defined earlier, the record high (low) temperature at time t is the highest (lowest) temperature observed up to time t.

[6] Although one can also estimate RED-H for the minimum of monthly average temperatures and vice versa, here we are interested in how the "extreme" monthly average temperatures become more extreme over time in order to examine changes in the tails of the temperature distribution.

We use HadCRUT3v data (Brohan et al. 2006) to test the robustness of our findings. These data also help to address whether the use of land only data in DelCCR2 leads to biased estimates of global temperature changes. Since the original $5° \times 5°$ grid of this data is too coarse to construct two separate time-series for the maximum and minimum temperatures, we construct record high and record low indicators from the single time-series of monthly average temperatures for each month-grid pair.[7]

As mentioned before, estimates of RED based on HadCRUT3v data could be biased because this data source has missing temperature information. Hence we construct our time-series of record indicators only on the basis of observed temperature data and ignore the years where temperature data are missing. To further minimize the impact of missing data we restrict our comparison to the period 1969–2008, where we have fewer missing data points. We substitute mean temperature in 1969 with the average of mean temperatures from 1940 to 1969. We also exclude month-grid pairs that have less than 10 years of data or more than 5 gaps in the time series. This leaves us with 1,376 grids in total (468 in landmass and 908 in sea).

3.2 Daily extreme temperature data

The data we used (called here by ETCCDI) for daily extreme temperatures are based on work by CCl/CLIVAR/JCOMM Expert Team (ET) on Climate Change Detection and Indices (ETCCDI) (Peterson 2005). The data set is available from http://cccma.seos.uvic.ca/ETCCDI/index.shtml and has been widely used to examine changes in extreme temperatures and precipitations (Griffiths et al. 2005; Alexander et al. 2006; Choi et al. 2009). The data set has monthly station level information of daily extreme temperatures. The station level data is preferred over the gridded data to examine changes in daily extreme temperatures since the latter data are smoothed and thus the record information in this data can be artificial.

To compare with our monthly average temperature data, we use the following four indices based on daily extreme temperatures: monthly maximum of the daily maximum temperature (TXx), monthly minimum of the daily maximum temperature (TXn), monthly maximum of the daily minimum temperature (TNx), and monthly minimum of the daily minimum temperature (TNn). It should be noted that a direct comparison of RED based on daily extreme temperatures and RED based on monthly average temperatures is not possible since monthly average temperatures contain both daily maximum and minimum temperatures, and thus the variability of daily extreme temperatures may not be identified in the monthly average temperatures.

We use station data at latitudes higher than 24°S (to be compatible with the monthly data) and exclude month-station pairs that have less than 10 years of data or more than 5 gaps in the time series. We also exclude station data that are identified as inhomonegeous. For each of the four series we substitute the mean temperatures from 1940 to 1969 as the base year temperature for 1969. Although we have station data up to 2007 in ETCCDI, we restrict our analysis to 2004 because (with the

[7] Although the use of the same time series for both high and low record samples implies that for any month-grid pair we cannot set both a high and low temperature record in any given time period, this restriction does not qualitatively change our results as long as the time-series is sufficiently long.

exception of the TXx time series) there are no station data in the tropics—i.e. in the latitudes between 24°S and 24°N—from 2004 to 2007. Hence our daily extreme temperature data span from 1969 to 2004.

The stations in ETCCDI are unevenly distributed across the global landmass, and hence it is difficult to say whether the RED estimates are globally representative. The two solutions to this problem are either to thin the station network (Frich et al. 2002) or create a grid network (Kiktev et al. 2003; Alexander et al. 2006; Caesar et al. 2006). If we thin the station network then we lose stations in places where we do not have many stations to begin with. On the other hand, if we grid the station network then temperature information is smoothed. As an alternative, we attach appropriate weightings for each station to aggregate *all* available station data as follows. We create a 5° × 5° grid network and compute the total number of stations in each grid. Then we weight the record break information of each station by the inverse of the total number of stations in the grid to which it belongs in our formulation of the likelihood function. By doing so, we still use station level data directly to estimate RED without smoothing the temperature information. The RED estimates do not vary much as we choose different grid resolutions,[8] and hence we only present results based on the 5° × 5° grid network.

4 Results

4.1 Global analysis

In Fig. 1 global annual estimates of RED-H Mx and RED-L Mn from 1929 to 2008 are shown in panels (a) and (b), respectively. Note that in both panels the figures show the 95% CI bands and the five-year moving-averages of the mean RED (both from 1,000 simulations of random tie-breaks). Global warming is indicated by an increase (decrease) in RED-H Mx (RED-L Mn). Since the RED-H Mx and RED-L Mn estimates do not suggest a linear trend over the entire time period of study, we first identify "change-points" (Lund and Reeves 2002) in time where the linear fit before and after the change-points are significantly different. The change-points we identify on a yearly basis are: 1986 and 1996 for RED-H Mx, and 1976, 1985 and 2004 for RED-L Mn.

The striking finding is that the global RED-H Mx has increased 10-fold between the first three decades of the last century (our base period) and the most recent decade (based on five-year moving-averages). In the context of global warming this implies that the likelihood of observing an extreme high temperature (measured in terms of monthly averages) has increased 10 times over the last hundred years. To further summarize our findings, RED-H Mx shows a mild upward trend until the mid 1980s, where the RED-H Mx about doubled from the base period. However, since 1986 the increase in RED-H Mx has been dramatic and also more volatile. On the other hand, the magnitude of changes in the global RED-L Mn (Fig. 1b) is much smaller compared to RED-H Mx over the past century. The detailed changes in RED-L Mn show a more complex pattern compared to the changes in RED-H

[8]The grid resolutions we tested include 2° × 2°, 5° × 5°, 8° × 8° and 10° × 10°.

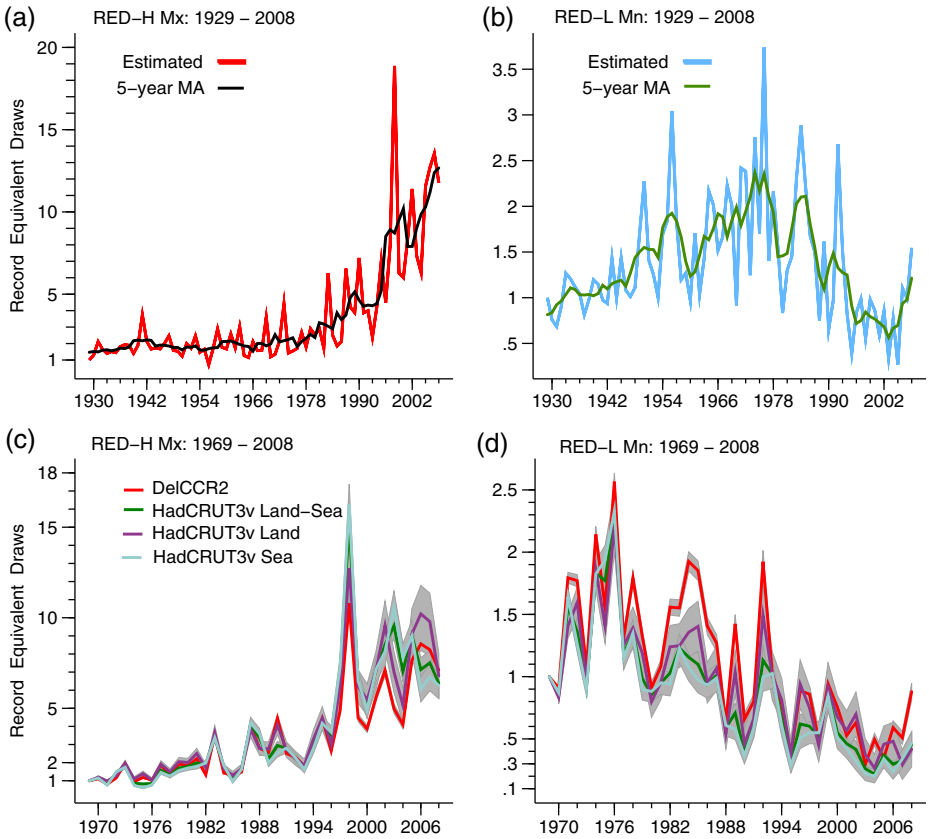

Fig. 1 Global annual estimates of RED from 1929 to 2008. RED-H Mx (*red*) are in panel (**a**) and RED-L Mn (*blue*) are in panel (**b**). The *black line* in (**a**) and the *green line* in (**b**) are the five-year moving averages (MA) of RED-H Mx and RED-L Mn estimates, respectively. In panel (**c**–**d**), the RED estimates based on DelCCR2 and HadCRUT3v data from 1969 to 2008 are compared. The *gray bands* around the RED estimates are 95% CI from a single tie-break simulation. All RED are normalized to 1 in 1929 in panel (**a**–**b**) and in 1969 in panel (**c**–**d**)

Mx (where we do not observe a downward trend of any significance over the past century). For example, RED-L Mn estimates in 2008 are not much different from eighty years earlier with less obvious trend lines during this period. The RED-L Mn doubled between the first three decades of the last century and the mid-1970s, and between the mid-1970s and 2004 RED-L Mn estimates decreased by more than half. This overall downward trend of RED-L Mn is consistent with findings from other studies using global data (Karl et al. 1993; Easterling et al. 1997). The increase in RED-L Mn since 2004 represents a cooling in terms of monthly averages, but it merely goes back to the same level as in the base period.

The apparent doubling of RED-H Mx and RED-L Mn between 1929 (where both RED are fixed at 1) and the 1970s (where both RED are about 2) is not inconsistent with studies on extreme temperatures that in particular have not found evidence of cold peaks in the 1960s and 1970s. Due to the unavailability of global data on

extreme temperatures prior to 1951 (Jones et al. 1999)[9] we can only compare our RED estimates with previous studies in the post 1950 time periods. Hence when we restrict our comparison to the 1950–1970 time period, we do not observe a significant time trend in either RED-H Mx or RED-L Mn, which is similar to findings from previous studies on extreme temperatures (Karl et al. 1993; Easterling et al. 1997). Moreover, directly comparing RED based on monthly data with findings from previous studies using daily extreme temperature data is not straightforward. Note that monthly data are constructed by averaging daily temperatures over the month and the daily extreme temperature data are the highest (or the lowest) observed daily temperature (over the course of a month). Hence the monthly RED-L is based not only on the extreme cold temperature in a month but also on all the other temperature observations in the month. Thus the two seemingly contradictory findings, namely, that monthly RED-L increases from 1929 to the 1970s and that there are no extreme-cold temperature peaks in the 1960s and the 1970s, perhaps suggest that the cooling during this period may have occurred in a more persistent manner that is not necessarily reflected in extreme cold temperatures.

In Fig. 1c–d, we present RED estimates based on HadCRUT3v and DelCCR2 data from 1969 to 2008[10] to check the robustness of our findings.[11] The fact that the 95% CIs fall within an extremely narrow band in Fig. 1a–b suggests that even a single simulation is sufficient to effectively break ties. Hence a single random tie-break is used to estimate RED for both HadCRUT3v and DelCCR2 in Fig. 1c–d.[12] The patterns and magnitudes of RED estimates are similar across these data implying that our original estimates are robust across these different data sources. The simple correlation coefficients between RED estimates of HadCRUT3v and DelCCR2 are indeed very high: .88 $\sim$ .96 for RED-H Mx and .82 $\sim$ .89 for RED-L Mn.

A second robustness check relates to the issue of spatial correlation of our temperature data. We implicitly assume that temperature distributions are independent across grids. However, temperature observations from neighboring grids can be temporally correlated, and this correlation in turn can bias our RED estimates by duplicating the same information. We test this spatial correlation hypothesis by comparing RED estimates from different grid sizes. The idea is that if the grid resolution is coarse enough, then temperature observations across grids are less likely to be temporally correlated. We test this implication by estimating RED across more coarsely gridded resolutions over the globe (in latitudes higher than 24°S)—4° $\times$ 4°(1,209 grids), 6° $\times$ 6° (590 grids) and 8° $\times$ 8° (364 grids).[13] As before, to construct

[9]Prior to the 1950s longer time series data on extreme temperatures are available only for a few countries, including the US, Canada and the former USSR. In this limited geographic context (prior to the 1950s) there is neither a strong upward nor downward trend, nor any uniform trends across the different regions (Karl et al. 1984, 1991, 1993; Kumar et al. 1994; Weber et al. 1994; Bonsal et al. 2001).

[10]The comparison of DelCCR2 and HadCRUT3v (on landmass) from 1930 to 1969 is found in the online supplementary material.

[11]For comparison purposes, we reconstruct the record high and record low indicators from the single time-series for each month-grid pair in DelCCR2 as in HadCRUT3v.

[12]Implementing multiple simulations for random tie-breaks in temperature observations is computationally costly when the data set is an unbalanced panel.

[13]All estimates of RED are statistically significant at less than 1% level of significance even in the case of the coarsest data set with 8° $\times$ 8° resolution.

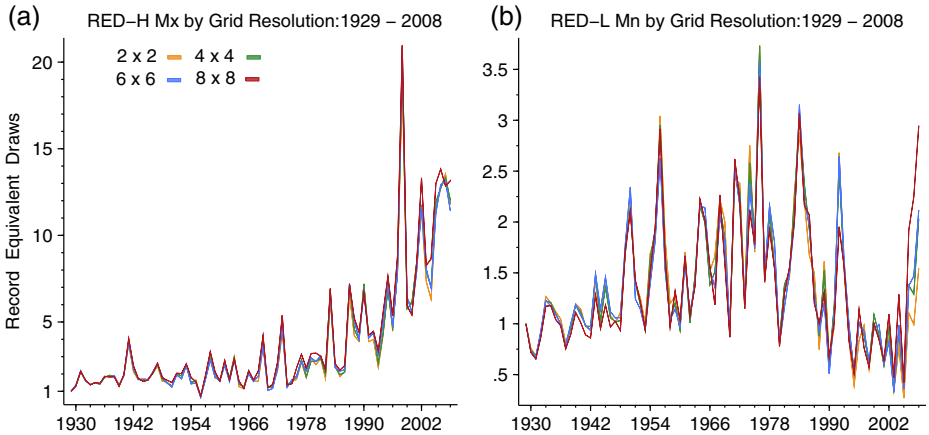

Fig. 2 Global annual estimates of RED by grid resolution from 1929 to 2008. RED-H Mx (in panel **a**) and RED-L Mn (in panel **b**) across four different resolutions –2° × 2°, 4° × 4°, 6° × 6° and 8° × 8°—are presented

the annual time-series of record high and low temperatures, we select the maximum and minimum monthly average temperature among the original grids included in each expanded grid (e.g. maximum of 64 grids in the 4° × 4° case) to create the record high and record low time series, respectively. Figure 2 shows that there is no discernible differentiation in the 95% CI bands (based on 1,000 simulations of random tie-breaks) across the four data sets with different resolutions.[14] These results suggest that even though spatial correlation of temperature data may not be zero, its effect on RED estimates is minimal.

4.2 Regional and seasonal analysis

Disaggregated analyses of RED provide some new information about climate change especially when compared with mean-based results (Jones and Briffa 1992). Here we estimate RED-H Mx and RED-L Mn across four regions (by latitude) and two seasons (winter/spring and summer/fall in the Northern Hemisphere (NH)). In Fig. 3 (regional breakdowns) and Fig. 4 (seasonal breakdowns), we show the five-year moving averages of the mean RED from 1,000 simulations of random tie-breaks, and the five-year moving averages of the mean LSAT as deviations from the mean temperature between 1961–1990.

In the regional comparison (Fig. 3), mean LSAT suggest that warming is somewhat less severe in the tropics (latitudes between 24°S and 24°N) compared with other regions (latitudes higher than 24°N): increases of .58°C and 1.05°C, respectively, from 1929 to 2008 are based on the difference between the first and last 5 years' mean LSAT. This is consistent with previous findings in the literature (Hunter et al. 1993; Trenberth et al 2007). In contrast, RED-H Mx in the past decade compared to the first three decades a century ago show more than a 16-fold increase

[14]The exception is RED-L Mn for 2007 and 2008 where it is higher in the 8° × 8° grid resolution than it is across the other grid resolutions.

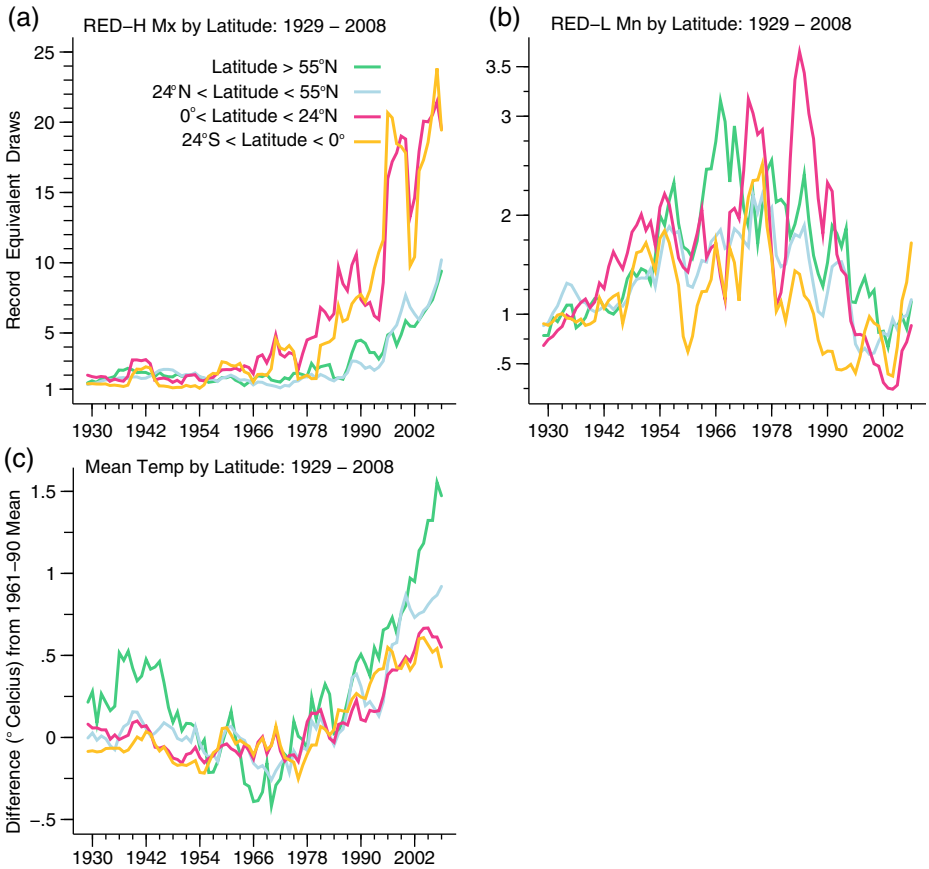

Fig. 3 Annual estimates of RED by latitude from 1929 to 2008. The five-year moving-averages of RED estimates (RED-H Mx in panel **a** and RED-L Mn in panel **b**) and the five-year moving-averages of mean temperatures in panel (**c**) are presented

in the tropics and only about a 7-fold increase in the extratropics. The apparent discrepancy of RED-H Mx and the mean LSAT across regions can be explained as follows. The fact that absolute changes in mean temperature are smaller in the tropics relative to the extratropics could simply imply that the underlying base temperature distribution in the tropics has a relatively smaller variability. However, note that any regional RED estimate is a measure of how that regional distribution evolves over time. Hence the higher RED-H Mx estimates in the tropics say, roughly speaking, that the right tail of the temperature distribution in the tropics became relatively "thicker" to its own base distribution compared to the right tail of the temperature distribution in the extratropics.[15]

[15]In more technical terms, recall that a RED-H Mx estimate of 16 in the tropics means that the transformed distribution is generated by taking the maximum of 16 draws from its base distribution, and a RED-H Mx estimate of 7 in extratropics means that the transformed distribution is only from a maximum of 7 draws (See Eq. 1).

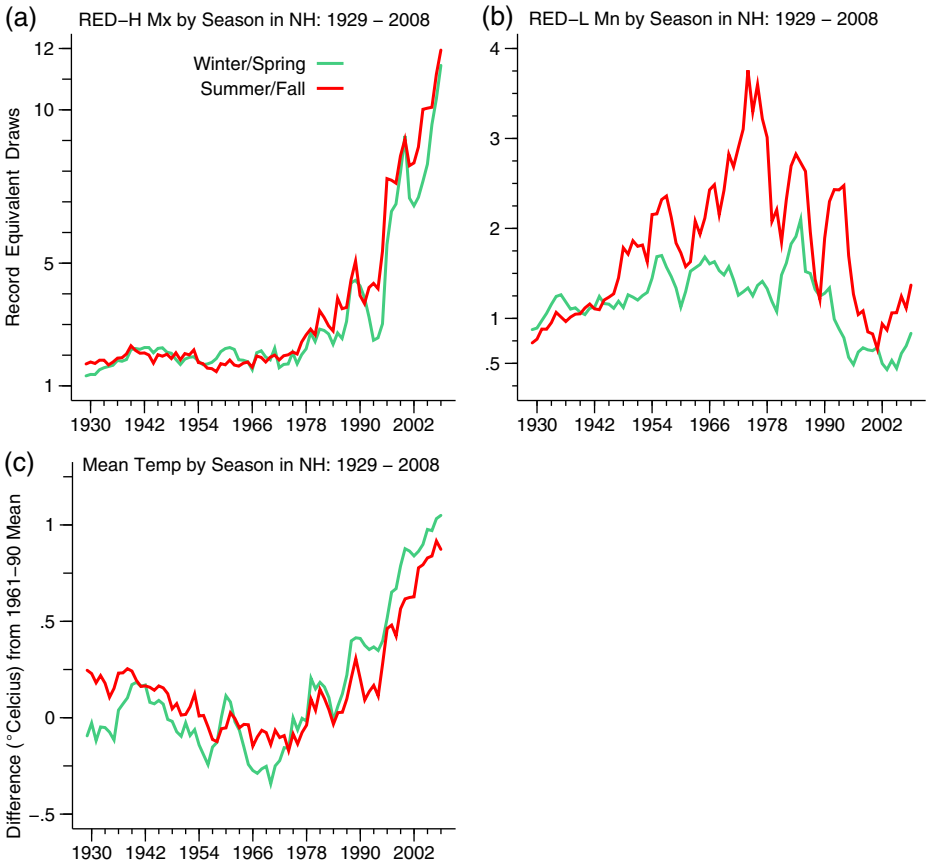

Fig. 4 Annual estimates of RED by season in the Northern Hemisphere from 1929 to 2008. The five-year moving-averages of RED estimates (RED-H Mx in panel **a** and RED-L Mn in panel **b**) and the five-year moving-averages of mean temperatures in panel (**c**) are presented

Periods since 2002 might indicate the action of opposing surface temperature forcings, as anthropogenic warming was counterbalanced by reduced solar irradiance and a generally negative phase of the PDO and ENSO cycles, compared to the time periods 1976–1998 (See Fig. 1 in Lean and Rind 2009). Moreover, the historical pattern of response to these two forcings suggests that the solar effect on mean temperature is greatest outside of the tropics, while the PDO/ENSO is greatest in the tropics (Lean and Rind 2008). Therefore the general congruence between the increase in RED-L Mn between 2002–2008 in subtropical/tropical regions (Fig. 3b) and a generally negative phase of the PDO and ENSO cycles during this period relative to time periods 1976–1998 suggests that the Pacific Ocean temperature variability might be primarily responsible.

In the seasonal comparison (Fig. 4), we find that RED-H Mx are not different between winter/spring and summer/fall in the Northern Hemisphere. However, mean LSAT shows that warming has been greater in winter/spring than summer/fall: increases of 1.09°C and .66°C, respectively, from 1929 to 2008. This difference across

seasons in mean LSAT is also found in the literature (Gordon 1992; Hunter et al. 1993; Zhang et al. 2000; Bonsal et al. 2001; Lund et al. 2001; Trenberth et al 2007) and may be related to the higher RED-L Mn in summer/fall compared to winter/spring (Fig. 4b).

Therefore in terms of climate change impacts, changes in the tropics and summer/fall may be at least as important as the changes in the extratropics and winter/spring. These results are not indicated by examining changes in the mean temperature.

4.3 Comparison with mean temperature

The analysis so far shows that our RED estimates can behavior very differently from the mean temperature. To investigate the relationship between RED and the mean temperature we perform several simulation exercises. More specifically, we generate a random data set where we can explicitly control the parameters of the underlying distribution and hence the evolution of this distribution over time. In the basic setup, we randomly draw 400,000 observations for each period from the normal distribution. We interpret each of these observations as originating from a station. We then pair two observations (i.e. two stations) from which we select the maximum and minimum numbers to create the two time series of record highs and record lows, respectively. This mimics the construction of our $2^\circ \times 2^\circ$ grid data from the original $0.5^\circ \times 0.5^\circ$ grid data (The simulation results do not change qualitatively when two record time series are created from a single time series from each of 400,000 stations). We repeat this random number generation process for 40 periods.

In order to understand the specific trend patterns we observe, we parameterize a normal distribution over a 40 period as follows: from period 1 to 19 we fix the mean at zero and increase the standard deviation throughout. Then from period 20 to 40 we increase the mean throughout and fix the standard deviation to the same level as in period 19. More specifically, we set the mean, $m = 0$ for $1 \leq t \leq 19$ and $(t - 19)/9$ for $20 \leq t \leq 40$, and the standard deviation, $\sigma = (-2t^2 + 105t)/1000 + 3.6$ for $1 \leq t \leq 19$ and 4.9 for $20 \leq t \leq 40$. Our objective here is to generate data—by carefully considering changes in the underlying distribution—that could mimic the general observed trends in mean temperature, RED-H Mx, and RED-L Mn during our period of study. Note that there are many different combinations of changes in mean and standard deviation that can generate the observed trends, and hence this is not an effort to pretend that these assumed changes correspond to reality.

Figure 5 shows the RED estimates based on simulated data. They clearly parallel the observed trends in Fig. 1a–b: in the period 1–19 the mean is invariant but both RED-H Mx and RED-L Mn are increasing—mimicking the 1929–1970 observed time period—and in period 20–40 the mean is trending upwards, RED-H Mx are increasing and RED-L Mn are decreasing—mimicking the post 1970s observed period.

In conclusion, these simulation results clearly illustrate that the mean temperature can be positively correlated with RED-H Mx and negatively correlated with RED-L Mn (holding everything else constant), and that in principle, concurrent changes in RED-H Mx and RED-L Mn can be associated with no changes in the mean. The latter observation raises the question of whether there are in fact periods where we

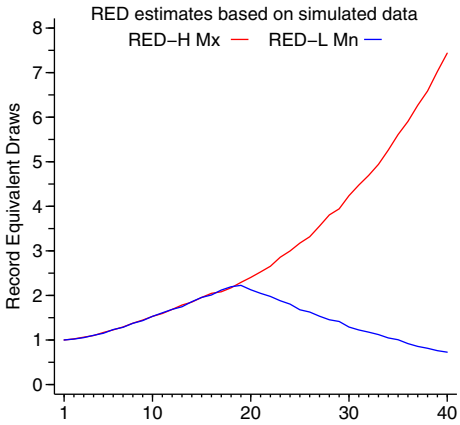

Fig. 5 RED estimates based on simulations. The RED are fixed at 1 in the first period

might observe concurrent increases (or decreases) in RED-H Mx and RED-L Mn with no discernible trend in the mean.

To address this question we use regression analyses to uncover changes in climate that may be hidden by simply looking at changes in mean temperature. Note that RED-H Mx and RED-L Mn are only moderately negatively correlated—with a simple correlation coefficient of −.37—suggesting potential periods of concurrent increases (or decreases) in the frequencies of both extreme high and extreme low temperatures. To identify periods where both RED-H Mx and RED-L Mn increase or decrease significantly and mean temperature changes are insignificant, we do the following exercise borrowed from trend analysis in the literature (Alexander et al. 2006; Choi et al. 2009). We run linear regressions (where autocorrelation in the error term is corrected by Cochrane-Orcutt AR(1))[16] of annual mean temperature and of RED-H Mx and RED-L Mn (using mean value of 1,000 Monte Carlo simulations) on a time variable that covers time-windows of length 5 to 25 years between 1929–2008. Through this exercise we find time periods where the time coefficient in the RED regressions for both high and low temperature have the same sign, and are statistically significant at 10% (or lower) levels of confidence, and where the time coefficient in the mean temperature regression is insignificant. In all cases, we require that the R-squared values exceed 10% to ensure a reasonable fit.[17]

Applying these criteria we identify the time window 2002–2008 where both RED-H Mx and RED-L Mn are increasing with almost no change in the mean

[16]Cochrane-Orcutt AR(1) regression (Cochrane and Orcutt 1949; Breusch and Pagan 1979) is simply the generalized least-squares method to estimate the parameters in a linear regression model in which the errors are assumed to follow a first-order autoregressive process AR(1). For the time series of our RED, the partial autocorrelation for lags larger than one is generally not significant and so we restricted it to an AR(1) process. Our procedure uses an iterative process to find the best autocorrelation coefficient. A similar procedure is used in Wang and Swail (2001), which is applied in Zhang and Zwiers (2004) for climate data.

[17]Note that the R-squared in Cochrane-Orcutt AR(1) cannot be compared with the counterpart in the OLS since it comes from the transformed dependent variable on the transformed independent variables.

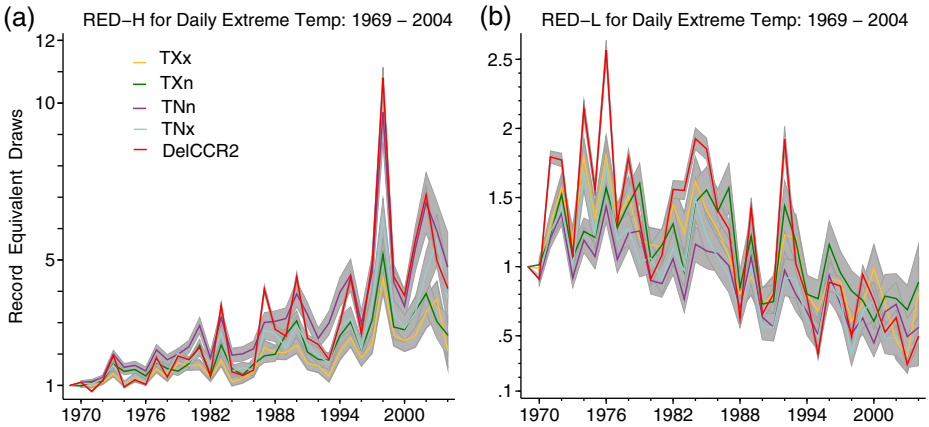

Fig. 6 Annual estimates of RED based on daily extreme temperatures from 1969 to 2004. RED-H and RED-L estimates are shown in panel (**a**) and (**b**), respectively. The *gray bands* around the RED estimates are 95% CI from a single tie-break simulation

temperature.[18] Specifically, we show that the plateauing in global mean temperature between 2002–2008 is associated with increases in the frequency of extreme high and extreme low temperatures of 13.3% and 22% per annum, respectively. Although one needs to be careful in generalizing this observation due to small sample sizes in this time window, the basic result is consistent with an increase in the variance of the temperature distribution in Europe, a conclusion based on heat waves during the summer of 2003 (Schäl et al. 2004).

4.4 Comparison with RED for daily extreme temperatures

In Fig. 6 we present annual estimates of RED for the four indices—TXx, TXn, TNx, and TNn—from 1969 to 2004 from the ETCCDI data set. For each of the four time series we present both RED-H and RED-L to examine whether breaking record high or record low temperatures dominates the trend for each time series since 1969. The RED estimates here are based on a single simulation of random tie-breaks (the gray bands around the RED estimates are 95% CI). RED estimates from 1969 to 2004 based on the monthly average temperature in DelCCR2 are also reproduced for easy comparison.

In Fig. 6b, the RED-L for the four time series of daily extreme temperatures are not distinguishable from each other, showing an overall decline from 1969 to 2004. There is also no difference between RED-L for daily extreme temperatures and RED-L Mn based on DelCCR2. In sum, the frequency of extreme low temperatures for each index in 2004 is less than the level three decades earlier.

In Fig. 6a, RED-H for the four daily extreme temperature time series show different magnitudes but the overall pattern of changes are similar. The RED-H for the daily minimum temperatures (TNn and TNx) have increased in general

[18]This observation is robust when the time window is restricted to 2002–2007.

more than the RED-H for the daily maximum temperatures (TXx and TXn). These observations are consistent with what others have found (Zhang et al. 2000; Yan 2002; Alexander et al. 2006; Choi et al. 2009): more areas on global landmass show significant increases in the daily minimum temperatures as opposed to the daily maximum temperatures.[19]

As the time trends in Fig. 6 illustrate RED estimates from both monthly average temperature and daily extreme temperature data are highly correlated. The simple correlations between RED estimates from the monthly and daily data are .94 ∼ .98 for RED-H and .80 ∼ .96 for RED-L. However, note that the increase in RED-H Mx based on DelCCR2 is generally greater than the increase in RED-H based on daily extreme temperature data. As noted earlier (Section 4.1), the finding that monthly RED-H increases are larger than daily RED-H increases suggests that the warming indicated by the daily extreme temperature data is perhaps more "persistent".

5 Conclusion

In this paper we have introduced a new metric to assess changes in the frequency of extreme temperatures and thus provided a precise quantification of global warming and cooling. Methodologically, our approach is capable of aggregating temperatures from heterogeneous distributions and has a clear interpretation of the changes in the temperature distribution. The inefficiency of our use of only record break information in computing RED can be corrected by rank ordering the data. However, this method is costly in terms of computational resources.

Substantively, while our RED estimates based on monthly average of land surface air temperatures reproduced the general patterns of warming and cooling indicated by changes in mean temperatures over the past century, it also provided new information on the spatial patterns of this warming, including new insights into climate changes, region and seasonal variability. In particular, while the tropical mean temperature did not increase as much as mean temperature in the extratropics, the number of high temperature record breaks was rising faster in the tropics than in the extratropics. Moreover, the well-known characteristic of global warming maximizing in the extratropics in winter was not reproduced in the pattern of high temperature record breaks, which shows, if anything, greater effect in the tropics with little seasonal variation. We also found congruence between the time series of RED estimates for monthly average temperature and for daily extreme temperatures, and the time evolution of these estimates are consistent with findings from previous studies.

The RED metric can be applied more widely to study other climate components such as sea surface temperature and precipitation where changes in the extremes are likely to have major societal and ecological impacts. For example, increases in extreme precipitation will impact not only people (Thibault and Brown 2008), but also the survival of various species (Rittenhouse et al. 2010). RED estimates from various climate components should be especially useful for comparison with

[19]This is also consistent with findings based on other indices. For example, days of warm summer nights have increased while frosty days have decreased (Frich et al. 2002; Griffiths et al. 2005; Alexander et al. 2006; Choi et al. 2009).

observed biological climate change responses (Walther et al. 2002; Parmesan and Yohe 2003; Rosenzweig et al. 2008) since they are more likely to be impacted by changes in the extremes than by changes in the mean (Easterling et al. 2000).

Acknowledgements DelCCR2, HadCRUT3v, and ETCCDI data were provided by the Center for Climatic Research at the University of Delaware, Climate Research Unit at the University of East Anglia, and Expert Team on Climate Change Detection and Indices, respectively. We gratefully acknowledge financial support from the Kyung Hee University Research Grant (2009) and Barnard College Presidential Research Award (2010). The primary research was conducted while the second author was a Research Fellow at Barnard College, and on sabbatical leave from Kyung Hee University. He would like to thank both institutions for providing invaluable support to conduct research on climate change. We thank three anonymous referees for their constructive criticisms and many helpful suggestions. We also thank Cynthia Howells, Brendan O'Flaherty, Stephanie Pfirman, Sanjay Tikku, Moonkyoung Um, and Jerry Welch for encouragement and generous feedback on earlier versions of the paper. Finally, we thank Claire Fram and Britt Johnson for excellent research assistance.

References

Alexander LV et al (2006) Global observed changes in daily climate extremes of temperature and precipitation. J Geophys Res 111:D05109. doi:10.1029/2005JD006290

Arnold BC, Balakrishnan N, Nagaraja HN (1998) Records. John Wiley, New York

Ballerini R, Resnick SI (1985) Records from improving populations. J Appl Probab 22:487–502

Battisti DS, Naylor RL (2009) Historical warnings of future food insecurity with unprecedented seasonal heat. Science 323:240–244

Benestad RE (2003) How often can we expect a record event? Clim Res 25:3–13

Bonsal BR, Zhang X, Vincent LA, Hogg WD (2001) Characteristics of daily and extreme temperatures over Canada. J Climate 14:1959–1976

Breusch TS, Pagan AR (1979) A simple test for heteroskedasticity and random coefficient variation. Econometrica 47:1287–1294

Brohan P, Kennedy JJ, Harris I, Tett SFB, Jones PD (2006) Uncertainty estimates in regional and global observed temperature changes: a new dataset from 1850. J Geophys Res 111:D12106. doi:10.1029/2005JD006548

Caesar J, Alexander L, Vose R (2006) Large-scale changes in observed daily maximum and minimum temperatures: Creation and analysis of a new gridded data set. J Geophys Res 111:D05101. doi:10.1029/2005JD006280

Chandler KN (1952) The distribution and frequency of record values. J Royal Stat Soc B 14:220–228

Choi G et al (2009) Changes in means and extreme events of temperature and precipitation in the Asia-Pacific network region: 1955–2007. Int J Climatol 29:1906–1925

Ciais Ph et al (2005) Europe-wide reduction in primary productivity caused by the heat and drought in 2003. Nature 437:529–533

Cochrane D, Orcutt GH (1949) Application of least squares regression to relationships containing auto-correlated error terms. J Am Stat Assoc 44:32–61

Curriero FC, Heiner K, Zeger S, Samet JM, Patz JA (2002) Temperature and mortality in 11 cities of the eastern United States. Am J Epidemiol 155:80–87

Easterling DR et al (1997) Maximum and minimum temperature trends for the globe. Science 277:364–367

Easterling DR et al (2000) Climate extremes: observations, modeling, and impacts. Science 289:2068–2074

Frich P et al (2002) Observed coherent changes in climatic extremes during the second half of the twentieth century. Clim Res 19:193–212

Githeko AK, Ndegwa W (2001) Predicting malaria epidemics in the Kenyan Highlands using climate data: a tool for decision makers. Glob Change Hum Health 2:54–63

Gordon AH (1992) Interhemispheric contrasts of mean global temperature anomalies. Int J Climatol 12:1–9

Griffiths GM et al. (2005) Change in mean temperature as a predictor of extreme temperature change in the Asia–Pacific region. Int J Climatol 25:1301–1330

Gumbel EJ (1958) Statistics of extremes. Columbia University Press, New York
Hall P, Tajvidi N (2000) Nonparametric analysis of temporal trend when fitting parametric models to extreme-value data. Stat Sci 15:153–167
Hay SI et al (2002) Climate change and the resurgence of malaria in the East African highlands. Nature 415:905–909
Hunter DE, Schwartz SE, Wagener R, Benkovitz CM (1993) Seasonal, latitudinal, and secular variations in temperature trend: evidence for influence of anthropogenic sulfate. Geophys Res Lett 20:2455–2458
Jones PD, Briffa KR (1992) Global surface air temperature variations during the twentieth century: part 1, spatial, temporal and seasonal details. Holocene 2:165–179
Jones PD et al (1999) Surface air temperature and its changes over the past 150 years. Rev Geophys 37:173–199
Karl TR, Kukla G, Gavin J (1984) Decreasing diurnal temperature range in the United States and Canada from 1941 through 1980. J Clim Appl Meteorol 23:1489–1504
Karl TR et al (1991) Global warming: evidence for asymmetric diurnal temperature change. Geophys Res Lett 18:2253–2256
Karl TR et al (1993) Asymmetric trends of daily maximum and minimum temperature. B Am Meteorol Soc 74:1007–1023
Kharin VV, Zwiers FW (2000) Changes in the extremes in an ensemble of transient climate simulations with a coupled atmosphere-ocean GCM. J Climate 13:3760–3788
Kiktev D, Sexton D, Alexander L, Folland C (2003) Comparison of modelled and observed trends in indicators of daily climate extremes. J Climate 16:3560–3571
Kumar RK, Kumar KK, Pant GB (1994) Diurnal asymmetry of surface temperature trends over India. Geophys Res Lett 21:677–680
Laurent C, Parey S (2007) Estimation of 100-year-return-period temperatures in France in a non-stationary climate: Results from observations and IPCC scenarios. Glob Planet Change 57:177–188
Lean JL, Rind DH (2008) How natural and anthropogenic influences alter global and regional surface temperatures: 1889 to 2006. Geophys Res Lett 35: L18701. doi:10.1029/2008GL034864
Lean JL, Rind DH (2009) How will Earth's surface temperature change in future decades? Geophys Res Lett 36:L15708. doi:10.1029/2009GL038932
Lund R, Reeves J (2002) Detection of undocumented changepoints: a revision of the two-phase regression model. J Climate 15:2547–2554
Lund R, Seymour, L, Kafadar K (2001) Temperature trends in the United States. Environmetrics 12:673–690
Meehl GA, Tebaldi C, Walton G, Easterling D, McDaniel L (2009) Relative increase of record high maximum temperatures compared to record low minimum temperatures in the U.S. Geophys Res Lett 36:L23701. doi:10.1029/2009GL040736
Munasinghe L, O'Flaherty B, Danninger S (2001) Globalization and the rate of technological progress: what track and field records show. J Polit Econ 109:1132–1149
Nevzorov VB (1985) Record and interrecord times for sequences of nonidentically distributed random variables. J Math Sci 36:510–516. doi:10.1007/BF01663462
Nogaj M, Yiou P, Parey S, Malek F, Naveau P (2006) Amplitude and frequency of temperature extremes over the North Atlantic region. Geophys Res Lett 33:L10801. doi:10.1029/2005GL024251
Parmesan C, Yohe G (2003) A globally coherent fingerprint of climate change impacts across natural systems. Nature 421:37–42
Peterson TC (2005) Climate change indices. WMO Bulletin 54:83–86
Rényi A (1962) Theorie des elements saillants d'une suite d'observations. Colloquium on Combinatorial Methods in Probability Theory, pp 104–117
Rittenhouse CD et al (2010) Avifauna response to hurricanes: regional changes in community similarity. Glob Change Biol 16:905–917
Rosenzweig C et al (2008) Attributing physical and biological impacts to anthropogenic climate change. Nature 453:353–357
Schäl C et al (2004) The role of increasing temperature variability in European summer heatwaves. Nature 427:332–336
Thibault KM, Brown JH (2008) Impact of an extreme climatic event on community assembly. Proc Natl Acad Sci USA 105:3410–3415
Trenberth KE et al (2007) Observations: surface and atmospheric climate change. In: Solomon et al (eds) Climate change 2007: the physical science. Cambridge University Press, New York

Walther GR et al (2002) Ecological responses to recent climate change. Nature 416:389–395
Wang XL, Swail VR (2001) Changes of extreme wave heights in northern hemisphere oceans and related atmospheric circulation regimes. J Climate 14:2204–2221
Weber RO, Talkner P, Stefanicki G (1994) Asymmetric diurnal temperature change in the Alpine Region. Geophys Res Lett 21:673–676
Willmott CJ, Matsuura K (1995) Smart interpolation of annually averaged air temperature in the United States. J Appl Meteorol 34:2577–2586
Yan Z (2002) Trends of extreme temperatures in Europe and China based on daily observations. Clim Change 53:355–392
Yang MCK (1975) On the distribution of the inter-record times in an increasing population. J Appl Probab 12:148–154
Yiou P, Goubanova K, Li ZX, Nogaj M (2008) Weather regime dependence of extreme value statistics for summer temperature and precipitation. Nonlinear Process Geophys 15:365–378
Zhang X, Vincent LA, Hogg WD, Niitsoo A (2000) Temperature and precipitation trends in canada during the 20th century. Atmos Ocean 38:395–429
Zhang X, Zwiers FW (2004) Comment on "Applicability of prewhitening to eliminate the influence of serial correlation on the Mann-Kendall test" by Yue S, Wang CY. Water Resour Res 40. doi:10.1029/2003WR002073
Zwiers FW, Kharin VV (1998) Changes in the extremes of the climate simulated by CCC GCM2 under CO2 doubling. J Climate 11:2200–2222
Zwiers FW, Zhang X, Feng J (2011) Anthropogenic influence on long return period daily temperature extremes at regional scales. J Climate 24:881–892