GEOPHYSICAL RESEARCH LETTERS, VOL. 35, LXXXXX, doi:10.1029/2008GL036228, 2008

How unusual is the recent series of warm years?

E. Zorita,[1] T. F. Stocker,[2] and H. von Storch[1]

Received 5 October 2008; accepted 18 November 2008; published XX Month 2008.

[1] Previous statistical detection methods based partially on climate model simulations indicate that, globally, the observed warming lies very probably outside the natural variations. We use a more simple approach to assess recent warming at different spatial scales without making explicit use of climate simulations. It considers the likelihood that the observed recent clustering of warm record-breaking mean temperatures at global, regional and local scales may occur by chance in a stationary climate. Under two statistical null-hypotheses, autoregressive and *long-memory*, this probability turns to be very low: for the global records lower than $p = 0.001$, and even lower for some regional records. The picture for the individual long station records is not as clear, as the number of recent record years is not as large as for the spatially averaged temperatures.
Citation: Zorita, E., T. F. Stocker, and H. von Storch (2008), How unusual is the recent series of warm years?, *Geophys. Res. Lett.*, *35*, LXXXXX, doi:10.1029/2008GL036228.

1. Introduction

[2] Global mean surface air temperature shows a positive trend in the 20th century [*Brohan et al.*, 2006; *Hansen et al.*, 2006; *Smith and Reynolds*, 2005]. After an initial "flat" development at a pre-industrial level, a clear warming took place from the 1910s to the 1940s, leveling off in the 1970s. Since mid 1970s temperatures show a positive trend of about 0.18°C per decade. This temperature increase and its geographical patterns have been subject to detection studies [*Hegerl et al.*, 1996; *Barnett et al.*, 2005], which are partially based on sophisticated statistical analysis of observations and climate simulations. Here, we pursue a simpler and intuitive idea, more accessible to the non-expert, to estimate the likelihood that the recent observed warming is consistent with the natural variability.

[3] The increasing anthropogenic greenhouse forcing would, according to simple physical reasoning [*Arrhenius*, 1896], cause a clustering of record warm years at the end of the observed record, and in fact, the 13 warmest years in 1880–2006 have all occurred in or after 1990. A clustering of record years at the beginning of the record would be in contradiction to anthropogenic greenhouse forcing. The probability of this clustering occurring by chance at the end of the record can be estimated under different null hypotheses of the statistical characteristics of natural variability. The most simple of those is that the annual values of mean global temperatures are independent of one another. In this case, the probability p of the event E of finding at least 13 of the largest values of a sequence of 127 independent random numbers on the last 17 places (year 1990 to 2006) is $p(E) = (114!17!)/(127!4!) = 1.25 \times 10^{-14}$. Such clustering appears as an extremely improbable random event in a stationary climate. However, the annual surface air-temperatures displays a serial correlation, even in a stationary natural climate, due to processes occurring on the land surface, ocean, and cryosphere. We conceptualize this natural memory with two statistical models, namely "short term" and "long term" memory. Within the former, annual mean temperature is assumed to be the result of an autoregressive process, which displays an exponentially decaying auto-covariance function. In the latter, it is described by *long-term persistence* process with a power-law decay of the auto-covariance function. Some processes that may be relevant in the climate context have been shown to display such type of behavior [*Bunde et al.*, 2005; *Rybski et al.*, 2006]. Although it is quite difficult to ascertain whether a short time series such as the observed global mean annual temperature obeys this type of behavior, the present approach can be readily applied to other more complex null models. It also offers the advantage that it can be also extended to those spatial scales using some of the longest individual station records.

2. Data and Methods

[4] We have analyzed global, regional and long station temperature records: global annual mean temperature in the period 1880–2006 from three global temperature data sets - Hadley Centre-CRU [*Brohan et al.*, 2006], NASA GISS [*Hansen et al.*, 2006] and NOAA NCDC [*Smith and Reynolds*, 2005]; the regional annual temperature means in the period 1850–2006 from a spatial average of those grid-cells of the Hadley-Centre-CRU data in 26 geographical regions [*Giorgi and Bi*, 2005]; and eleven long station records constructed by blending original time series kindly provided by *Jones and Moberg* [2003], and which end in year 2000, with the station data from year 2001 onwards provided by NASA-GISS. Details are given in the auxiliary material.[1]

[5] The parameters of the autoregressive models were estimated from the observed records in the period up to 1960 to limit the influence of the anthropogenic forcing. For all but two regional records (South Australia and Southern South America) an autoregressive model of order one (AR-1) would be adequate. For these two regional records, the Durbin-Watson test indicated the presence of autocorrelated residuals in a AR-1 processes. However, this may be due to chance, as they appear in 2 out of 26 regional records. Estimation of confidence intervals for lag-1 autocorrelation

[1]Institute for Coastal Research, GKSS Research Centre, Geesthacht, Germany.

[2]Physics Institute and Oeschger Centre for Climate Change Research, University of Bern, Bern, Switzerland.

0094-8276/08/2008GL036228$05.00

[1]Auxiliary materials are available in the HTML. doi:10.1029/2008GL036228.

is based on bootstrap methods [*Effron and Tibshirani*, 1993]; details are given in the auxiliary material.

[6] The second null-hypothesis is that the temperature records are realizations of a 'long-range autocorrelation' process [*Cohn and Lins*, 2005]. The power-law decay of the autocorrelation is characterized by $C(k) \sim k^{-\gamma}$ where k is the time lag. The value of γ is related to the fractional differencing parameter d by $\gamma = 1 - 2d$. For the process to be stationary d must lie between 0 and 0.5. The Whittle method [*Shimotsu and Phillips*, 2005] was used here to estimate its value. Different statistical test of the stationarity of the global mean temperature have yielded conflicting results [*Stern and Kauffman*, 2000]. The Whittle method gives values slightly larger than 0.5, even disregarding the period from 1960 onwards, for all three global records. Proxy-based reconstructions of the Northern Hemisphere annual temperature in the past millennium, with variations less affected by anthropogenic greenhouse forcing than in the 20th century, yield values for d in the range 0.32 to 0.54 [*Rybski et al.*, 2006]. Considering these possible uncertainties, it will be assumed here that d is smaller, but very close, to 0.5. It is noted that other more complex null models are possible, e.g., where both types of short-term and long-term persistence are present [*Cohn and Lins*, 2005]. The simultaneous estimation of both parameters becomes, however, is more problematic. This will be pursued in further analysis.

3. Global Mean Temperature

[7] We focus on the number of warm record years in the last 17 years of the record, as illustrated by *Meehl et al.* [2007]. However, this choice is not critical, as the conclusions remain robust when considering the last 10 to last 20 years of the record. Figure 1 shows the probability p, as a function of the lag-1 autocorrelation α or of the fractional differencing parameter d, that the clustering of high values fulfills the criterion that at least m of the largest values occupy the last 17 places. $Log(p)$ under the autoregressive null-hypothesis exhibits a remarkably linear dependence on α up to very high values. It is not straightforward to determine which value of α would better represent the natural persistence of the global temperatures, i.e., without the effect of anthropogenic forcing. The last decades of the 20st century are probably too strongly affected. Temperature records in the decade 1940–1950 have been found to be distorted by changing in the measuring devices of sea-surface temperatures [*Thompson et al.*, 2008] and temperature data in the late 19th century are burdened by higher uncertainties [*Brohan et al.*, 2006]. To be conservative, i.e., risking an overestimation, a value of α in the range of 0.75 to 0.85 (auxiliary material) can be assumed. Within this range the probability of a random occurrence of event E is about 10^{-5} to 10^{-3}, a higher likelihood than under the null-hypothesis of white noise, but still an extremely rare event. The dependence of $log(p)$ on the value of the fractional differencing parameter d is not so steep in the range of 0 to 0.45, but $p(E)$ is still quite low for values of d close to 0.5, yielding a statistical significance for event E of about 10^{-3} as well. The results for the last n = 17 years can be considered typical. Figure 1 includes the range of probabilities obtained when considering the range from n = 10 to n = 20, for the cases $\alpha = 0.85$ and $d = .45$ (auxiliary material).

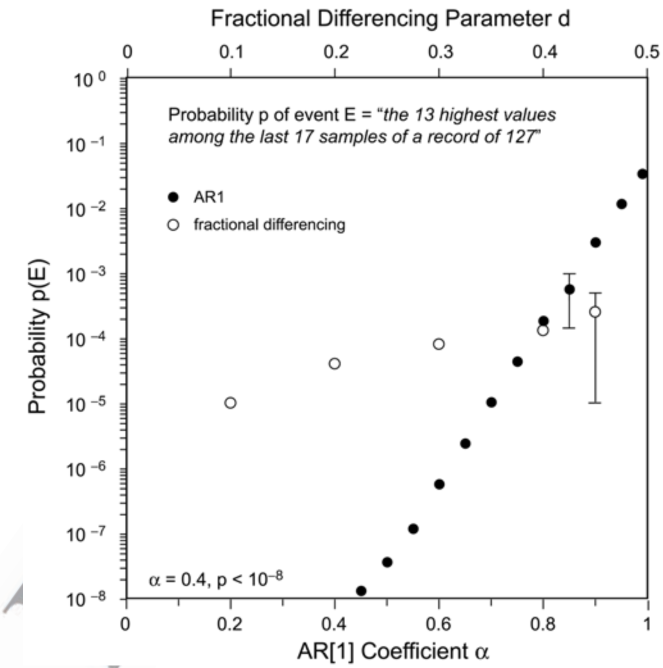

Figure 1. Probability that the 13 largest values in synthetic time series of 127 values occupy the last 17 places: closed circles, autoregressive process with different lag-1 autocorrelation α (bottom axis); open circles, long-range-correlation process for different values of the fractional differencing parameter d (top axis). Vertical lines around the probability at $\alpha = 0.85$ and $d = 0.45$ display the range of probabilities obtained when considering, instead of 17 years, the last 10 to 20 years in the record and, instead of the 13 largest values, their corresponding number of record years as observed in the global mean temperature (see also auxiliary material).

4. Regional Temperature Records

[8] Series of record warm years can be also found in regional temperatures. Here, we consider broad geographical regions as defined previously [*Giorgi and Bi*, 2005] and discussed by *Meehl et al.* [2007]. The number of warm record years in the last 17 years is smaller in the regional series than in the global mean because amplitudes of natural variability at regional scales are larger. However, the persistence of the regional temperature series is also smaller (smaller α and d), and therefore the likelihood of the clustering of a smaller number of record warm years could still be theoretically smaller than for the global mean temperature, thus making such events even more rare than for the global mean temperature.

[9] Some statistical characteristics of the regional records are given in the auxiliary material. Figure 2 displays the value of the lag-1 autoregressive parameter α and the log-probabilities of event E under the autoregressive null-hypothesis (numerical values are enclosed in the auxiliary material). The log-probabilities of event E depend on α, on the number of missing values present in the observed

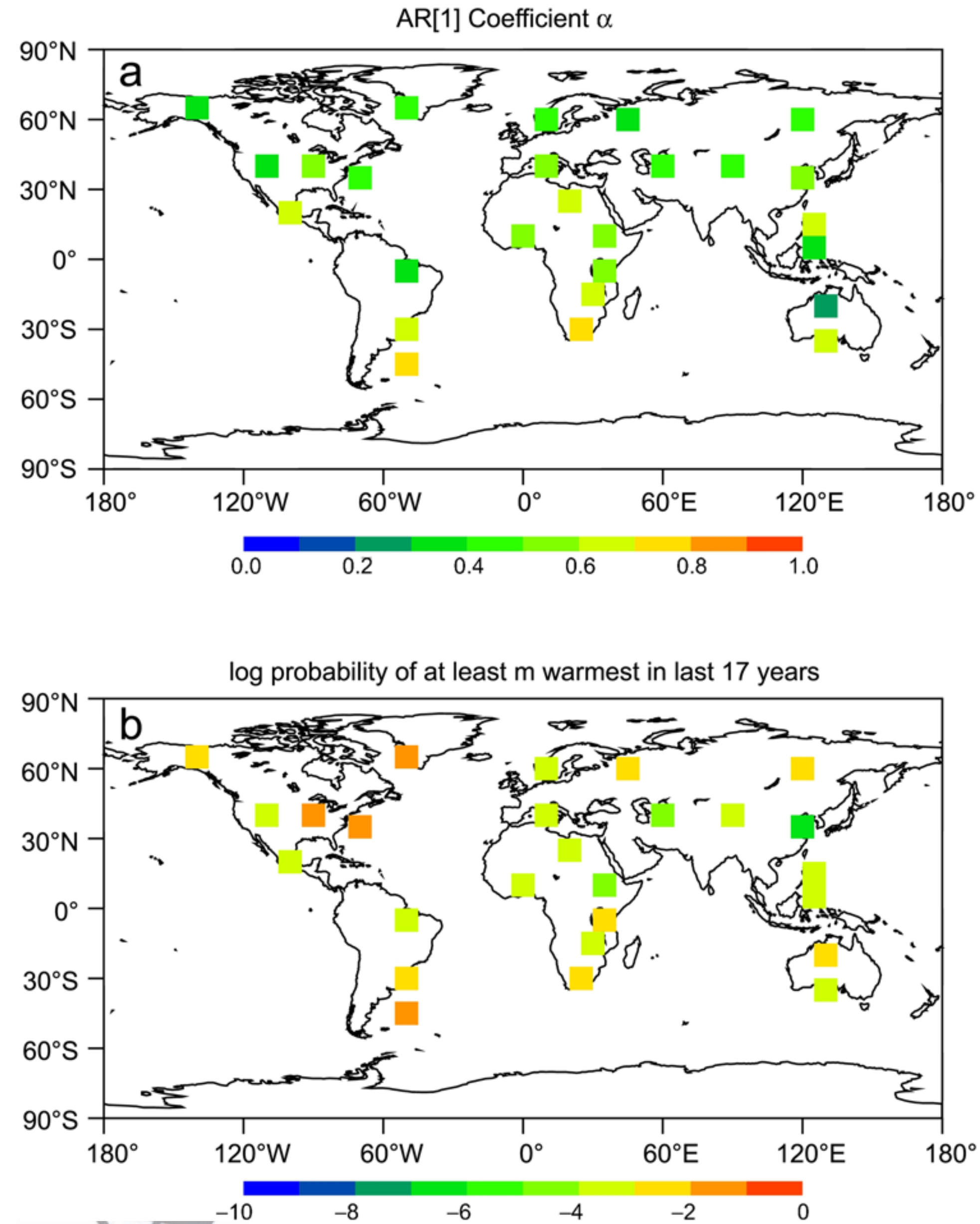

Figure 2. (a) Estimated values of the lag-1 autocorrelation parameter α for the annual mean temperature in each Giorgi region in the period 1850–2007. (b) Log-probability of the event E that the m largest values of 157 values occupy the last 17 places in lag-1 autoregressive synthetic time series. For the observed values of m see the auxiliary material.

records and on the number of warm record years observed in the last 17 years. The value of α shows a slight but perceptible meridional gradient with lower values at high latitudes in the Northern hemisphere and higher values in the regions adjacent to the Southern Ocean. The probabilities of event *E* occurring by chance under the autoregressive null-hypothesis, on the other hand, do not show a clear geographical pattern. For many of the 26 regions the likelihood of event E under the autoregressive null-hypothesis is quite low, especially in Western Europe, Africa and Central Asia, in the range 10^{-5} to 10^{-4}, which is even lower than for the global mean temperature.

[10] The estimated fractional differencing parameter *d* for the regional records also show a clear geographical pattern, with a tendency for higher values in the Southern Hemisphere (Figure 3). The interaction of this gradient with the geographical variation of the number of record-years in the last 17 years yields no clear pattern. The probability of event E occurring by chance under the long-memory null-hypothesis is in general higher than in the autoregressive case. Nevertheless, in several regions in Africa and Eurasia, this level of significance lies between 10^{-2} and 10^{-3}.

5. Individual Long Station Records

[11] The individual stations are geographically clustered in Europe. The results are summarized in Figure 4 and in the auxiliary material. Due to the local temperature noise the

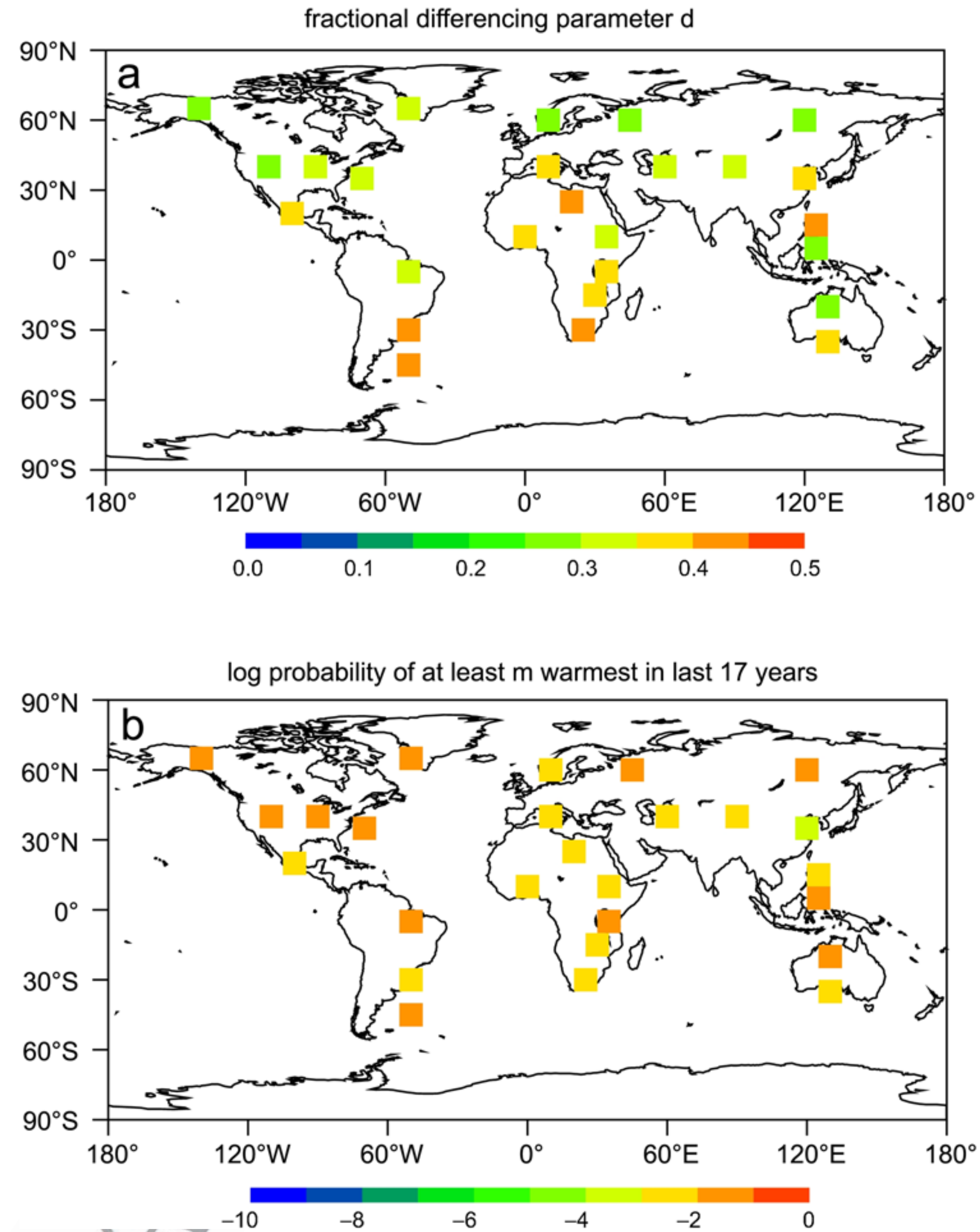

Figure 3. (a) Estimated values of the fractional differencing parameter d for the annual mean temperature in each Giorgi region in the period 1850–2007. (b), Log-probability of the event E that the *m* largest values of 157 values occupy the last 17 places in long-term autocorrelation synthetic series. For the observed values of m see auxiliary material.

values of α are lower than for the regional or global mean but the number of recent record years is also not as large. Both factors tend to balance, yielding probabilities in the same range as for the regional records (Figure 4b). In general, the value of d for the stations and regional records lies within the range 0 to 0.3 (Figure 4c), in accordance to independent estimations [*Bunde et al.*, 2005]. The results for the long-term-persistence null-hypothesis are qualitatively similar to the autoregressive case (Figure 4d).

6. Conclusions

[12] Our assessment of the rarity of the event E depends on the underlying null hypothesis. The risk of erroneously rejecting the null hypothesis strongly depends on the assumption about the character of the memory, as is demonstrated for long-term trends by *Cohn and Lins* [2005] and for the occurrence of extremes by *Bunde et al.* [2005]. This significance can change by orders of magnitude. But, nevertheless, the clustering of warm years at the end of the observed record would appear to be as an extremely rare event under either of the two assumptions about the memory that considered here. For the global mean temperature, conservative values of the lag-1 autocorrelation for global temperature yield very low likelihoods for the recent clustering of record warmth and under the long-term-persistence model this probability is even lower. The analysis of the regional records indicates that for some this likelihood is lower than for the global mean. This may seem surprising, as global records should inherently contain less internal

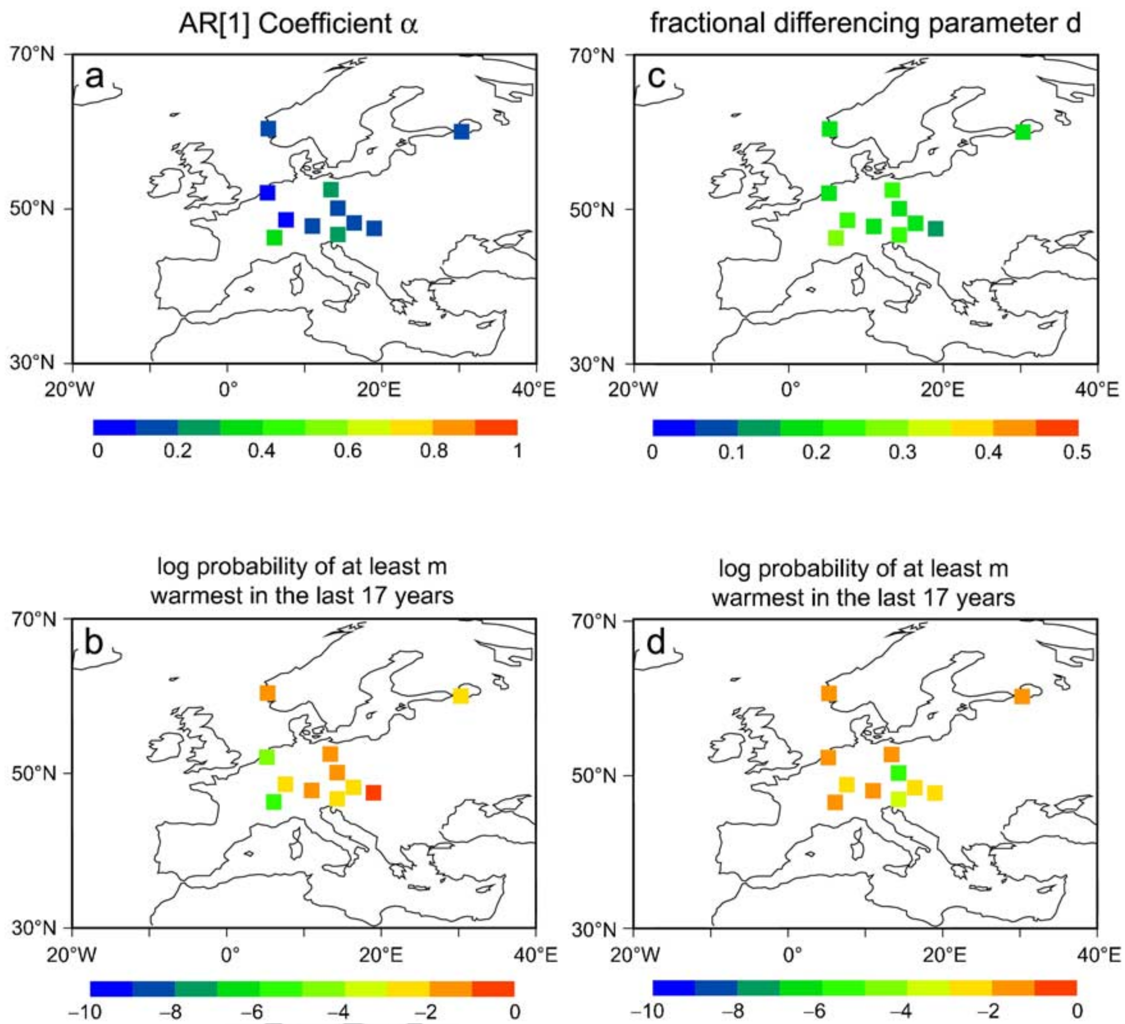

Figure 4. (a and b) Same as Figure 2 but for long individual stations records. (c and d) Same as Figure 3 but for long individual stations records.

noise and thus a higher signal-to-noise ratio. Two factors are regionally at play: a smaller number of record years and a smaller statistical persistence, rendering lower likelihoods of random clustering. This conclusion also holds for the individual station records. For these, the number of recent record years and the persistence of the record tend to be smaller than for the spatially aggregated records. It should be noted that the estimation of α and d from the observed records could be arguably biased towards larger values (e.g., larger persistence) by the effects of anthropogenic greenhouse forcing, and therefore the estimated probabilities of occurrence of event E, under these two null hypotheses, could be also biased high.

[13] **Acknowledgments.** We thank Armin Bunde for his kind help with the long-term-persistence algorithms and A. Moberg and P. Jones for permission to use their data products.

References

Arrhenius, S. (1896), On the influence of carbonic acid in the air upon the temperature on the ground, *Philos. Mag.*, *41*, 237–276.

Barnett, T., et al. (2005), Detecting and attributing external influences on the climate system: A review of recent advances, *J. Clim.*, *18*, 1291–1314.

Brohan, P., J. J. Kennedy, I. Harris, S. F. B. Tett, and P. D. Jones (2006), Uncertainty estimates in regional and global observed temperature changes: A new data set from 1850, *J. Geophys. Res.*, *111*, D12106, doi:10.1029/2005JD006548.

Bunde, A., J. F. Eichner, J. W. Kantelhardt, and S. Havlin (2005), Long-term memory: A natural mechanism for the clustering of extreme events and anomalous residual times in climate records, *Phys. Rev. Lett.*, *94*, 048701, doi:10.1103/PhysRevLett.94.048701.

Cohn, T. A., and H. F. Lins (2005), Nature's style: Naturally trendy, *Geophys. Res. Lett.*, *32*, L23402, doi:10.1029/2005GL024476.

Effron, B., and R. J. Tibshirani (1993), *An Introduction to the Bootstrap*, *Monogr. Stat. Appl. Probab.*, vol. 57, Chapman and Hall, New York.

Giorgi, F., and X. Bi (2005), Updated regional precipitation and temperature changes for the 21st century from ensembles of recent AOGCM simulations, *Geophys. Res. Lett.*, *32*, L21715, doi:10.1029/2005GL024288.

Hansen, J., M. Sato, R. Ruedy, K. Lo, D. W. Lea, and M. Medina-Elizalde (2006), Global temperature change, *Proc. Natl. Acad. Sci. U. S. A.*, *103*, 14,288–14,293.

Hegerl, G. C., H. von Storch, K. Hasselmann, B. D. Santer, U. Cubasch, and P. D. Jones (1996), Detecting greenhouse-gas-induced climate change with an optimal fingerprint method, *J. Clim.*, *9*, 2281–2306.

Jones, P. D., and A. Moberg (2003), Hemispheric and large-scale surface air temperature variations: An extensive revision and update to 2001, *J. Clim.*, *16*, 206–223.

Meehl, G. A., et al. (2007), Global climate projections, in *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by S. Solomon et al., pp. 747–845, Cambridge Univ. Press, Cambridge, U. K.

Rybski, D., A. Bunde, S. Havlin, and H. von Storch (2006), Long-term persistence in climate and the detection problem, *Geophys. Res. Lett.*, *33*, L06718, doi:10.1029/2005GL025591.

Shimotsu, K., and P. Phillips (2005), Exact local Whittle estimation of fractional integration, *Ann. Stat.*, *33*, 1890–1933.

Smith, T. M., and R. W. Reynolds (2005), A global merged land air and sea surface temperature reconstruction based on historical observations (1880–1997), *J. Clim.*, *18*, 2021–2036.

Stern, D. I., and R. K. Kauffman (2000), Detecting a global warming signal in global temperature series: A structural time series analysis, *Clim. Change*, *47*, 411–438.

Thompson, D. W. J., K. K. Kennedy, J. M. Wallace, and P. D. Jones (2008), A large-discontinuity in the mid-twentieth century in observed global-mean surface temperature, *Nature*, *453*, 646–650.

T. Stocker, Physics Institute, University of Bern, Sidlerstrasse 5, CH-3012 Bern, Switzerland.

H. von Storch and E. Zorita, Institute for Coastal Research, GKSS Research Centre, Max-Planck-Straße 1, D-21502 Geesthacht, Germany. (eduardo.zorita@gkss.de)