INTERNATIONAL JOURNAL OF CLIMATOLOGY, VOL. 15, 369 377 (1995) 551.583.14:551.501.45

A NEW METHOD FOR DETECTING UNDOCUMENTED DISCONTINUITIES IN CLIMATOLOGICAL TIME SERIES

DAVID R. EASTERLING AND THOMAS C. PETERSON

Global Climate Laboratory, NOAA/NESDIS, National Climatic Data Center, Federal Building, Asheville, NC 28801, USA

Received 7 January 1994
Accepted 12 July 1994

ABSTRACT

The development of homogeneous climatological time series is a crucial step in examining climate fluctuations and change. We review and test methods that have been proposed previously for detecting inhomogeneities, and introduce a new method we have developed. This method is based on a combination of regression analysis and non-parametric statistics. After evaluation against other techniques, using both simulated and observed data, our technique appears to have the best overall performance.

KEY WORDS Temperature Time series Homogeneity

INTRODUCTION

The importance of using homogeneous climatological time series in climate research has received much attention recently (e.g. Bradley and Jones, 1985; Folland *et al.*, 1992). A homogeneous time series is defined as one where variations are caused only by variations in weather and climate (Conrad and Pollak, 1962), and using climatological time series containing non-climatic induced variations can lead to inconsistent conclusions.

Inhomogeneities in time series occur either as a gradual trend, as in urban warming, or as an abrupt discontinuity (jump) in the times series. Discontinuities occur for a number of reasons: from station moves and instrument changes to changes in methods for calculating time-averaged values. For example, Quayle *et al.* (1991) demonstrated that changing from liquid-in-glass (LIG) thermometers to the thermistor-based maximum–minimum temperature system (MMTS) in the USA Cooperative Network caused an instrument-induced decrease in monthly maximum temperatures (−0·4°C) and increase in the monthly minimum temperatures (+0·3°C), resulting in an artificial narrowing of the monthly mean diurnal temperature range. Chenoweth (1992) examined thermometer exposure changes from a north wall to Cotton Region Shelter in the USA around the turn of the century, which appeared to cause a cooling in temperature readings leading him to suggest that late nineteenth century temperatures generally were too high. Discontinuities can also occur due to changes in averaging methods for time-averaged quantities and changes in observing times (Bradley and Jones, 1985).

Trend inhomogeneities are generally more difficult to detect, because they may be superimposed on a true climate trend. Karl *et al.* (1988) examined the problem of detecting and accounting for urban warming in temperature time series. They proposed a regression-based approach using urban population as a predictor variable to adjust for urban-warming effects. However, other trends, such as instrument drift, or changes in shelter characteristics (Quayle *et al.*, 1991) are even more difficult to detect.

It is impossible to verify that a climatological time series is truly homogeneous, and it is very rare to find a long climatological time series that is even homogeneous relative to surrounding stations. Furthermore, even if there are accompanying metadata, in the form of station histories, that are used to make data adjustments for homogenization purposes, the potential still exists that some undocumented station change

CCC 0899-8418/95/040369-09

has affected the statistical characteristics of the time series. Station history files are often incomplete and may not routinely document changes that can cause a discontinuity. For example, Easterling and Peterson (1992) examined a number of techniques for detecting discontinuities, and using data from the United States Historical Climatology Network (USHCN, Karl *et al.*, 1990) demonstrated that discontinuities can be identified for such instances as replacing a thermometer with a new one of the same model, or simply calibrating the instrument. The purpose of this paper, therefore, is to describe a new technique that we have developed for detecting undocumented discontinuities and compare its performance to others we reviewed in an earlier paper (Easterling and Peterson, 1992).

REVIEW OF PREVIOUS TECHNIQUES

Most of the techniques we review and describe in this paper are more suited to detecting and providing adjustments for abrupt discontinuities, generally defined as a change in the mean of a time series. However, in certain instances they also may be effective for removing trend inhomogeneities. We have also restricted the selection of techniques to those that are objective and are easily automated. This restriction is imposed because our goal is to develop a robust methodology for detecting and adjusting for discontinuities that is objective, reproducible, and can be applied to large data sets.

The determination of homogeneity must be made by comparison of time series from the station of interest (candidate) to those from nearby, closely related (highly correlated) reference stations. Implicit in this methodology is the assumption that any regional climate change or fluctuations experienced by the candidate station will show up in the nearby reference stations as well. In this sense, we define homogeneity in terms of what Conrad and Pollak (1962) have defined as relative homogeneity.

One of the first techniques proposed for determining homogeneity was double-mass analysis (Kohler, 1949), which is performed by direct comparison of one station against another. However, the major drawback of this approach is that it is impossible to determine which station contains the discontinuity. To circumvent this problem, a number of reference stations may be used, with each plot examined for discontinuities near the same point in time, as in parallel CUSUMS (Rhoades and Salinger, 1993). However, this method is subjective, is a tedious task for a large data set, and has the disadvantage of not immediately providing the means for calculating an adjustment factor.

Several techniques, such as Alexandersson's (1986) and Potter's (1981) use a single reference series created from a number of nearby stations. The use of multiple reference stations in constructing a single reference series helps minimize the effects of a discontinuity in one of the reference stations' time series. The analysis is performed on a difference series created by subtracting the candidate station's time series from the reference time series created for that station. In a companion paper (Peterson and Easterling, 1994) we have described our methodology for constructing homogeneous reference series. The methodology selects five reference stations by examining the correlation of the year-to-year change in temperature (dT/dt) series for the candidate station and each reference station. The use of the dT/dt series confines the effects of any one discontinuity to only 1 year in the correlation calculation. A single composite dT/dt series is then constructed from the five reference dT/dt series, then the composite series is reconstituted into a temperature series in order to serve as the reference series for a candidate station.

The regression-based method proposed by Gullet *et al.* (1991) is based on the idea of multiphase regression fits similar to a technique introduced by Solow (1987). The method uses a series of multiple regression fits to subsections of the candidate time series using reference stations as the predictor variables. When the best fit is determined by the decrease in the sum of squares of the residuals, the end points of each subsection then indicate potential discontinuities. Using simulated time series, we tested this approach and a variation, and found the variation to be more effective. Instead of using the candidate as the predictand and the reference stations as predictors, our variation used the difference series as the predictand and time as the predictor.

APPLICATION STRATEGIES AND TECHNIQUE DESCRIPTION

The purpose in developing this methodology is to have an automated, objective, and reproducible discontinuity identification and adjustment routine. The methodology must be able to identify the place in the time series of the discontinuity, and to calculate the amount of bias introduced so that an adjustment can be made starting at the discontinuity. Furthermore, the adjustments must be applied to bring the entire time series into agreement with the last (most recent) homogeneous segment of the series in order to allow data to be added to the end of the series. There are a number of strategies that can be used to apply the homogeneity test and calculate bias values, some of which are described below.

The first application strategy, which we used in Easterling and Peterson (1992), is to search the entire difference series for the most significant discontinuity, calculate the bias values as the difference in the means of the difference series before and after the discontinuity and adjust the first part of the series using this bias. The method continues to search and adjust the entire series until no significant discontinuity is found. However, there are drawbacks to this application strategy. Because the bias value is always the difference between means of the difference series using the entire length of the series, other discontinuities may cause the initial bias values to be over- or underestimated. Often this causes the same discontinuity to appear in a subsequent pass through the series, with the sum of the two bias values roughly equivalent to the proper value. Furthermore, if the initial bias value is less than the true bias value, and the time series is adjusted, it may cause the test not to identify the discontinuity as statistically significant on a subsequent pass, even though the discontinuity still exists in a diminished form due to a misspecified adjustment factor. A second problem with this strategy is that in the initial pass through the series the test may not identify any discontinuities, because multiple discontinuities, particularly those of opposite sign, may mask one another.

Another application strategy we evaluated is to use a windowing technique, where each year (place in the difference series) is tested for a discontinuity using only a specified number of the years (e.g. 15) on either side. However, we found this to be an unsatisfactory strategy, particularly when discontinuities were close together in time.

The strategy that we found most effective is integral to the technique we developed. We use a regression approach with the difference series as the predictand and time as the predictor. First, a simple linear regression is fitted to the entire difference series, and the residual sum of squares (RSS_1) is calculated. Next, a two-phase regression approach is taken, however, the two regression lines are not constrained to meet as in Solow's (1987) technique. For each year in the difference series, a simple linear regression is fitted to the difference series before the year being tested, and another simple linear regression is fitted to the part of the difference series after that year (see Figure 1). The RSS_2 is the sum of the RSSs calculated from each of the two regressions. Hence, for each place (year) in the difference series a two-phase regression is fitted and the RSS_2 calculated. The place with the minimum in RSS_2 is flagged as the location of a potential discontinuity.

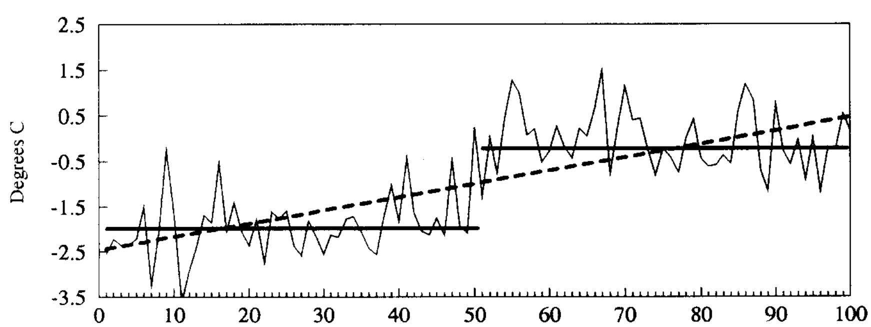

Figure 1. Example of a difference series between a simulated reference–candidate pair with one discontinuity of 2·0 with simple linear and two-phase regression lines

To test the significance of the two-phase fit we use the following likelihood ratio statistic from Solow (1987):

$$U = [(RSS_1 - RSS_2)/3]/[RSS_2/(n-4)]$$

This test statistic is F-distributed with 3 and $n-4$ degrees of freedom. We also test the difference in the means of the difference series before and after the potential discontinuity using Student's *t*-test. The *t*-test ensures that a difference series with a trend is found to be inhomogeneous, even when the U value is not significant.

In our current approach, instead of immediately identifying the bias value and adjusting the time series, the year is recorded as a potential discontinuity and the difference series is subdivided into two parts, before and after the previously identified potential discontinuity. Each subsection is then tested in the same manner as the first step, with any potential discontinuities being recorded and used to subdivide the difference series for further testing, until either the sections are found to be homogeneous, or they become too small (<10 places) to test. The smallest subsection that can be tested is 10 places, allowing only the fifth place (year) to be tested for a discontinuity, because the regression calculations are limited to a minimum of five places before and after the year being tested.

After we have identified all potential discontinuities in the candidate time series the statistical significance of each discontinuity is tested using multiresponse permutation procedures (MRPP, Mielke *et al.*, 1981). The MRPP test is non-parametric and compares the Euclidean distances between members within each group with the distances between all members from both groups, to return a probability that two groups more different could occur by random chance alone. In this case the two groups are comprised of the difference series values before (D_b) and after (D_a) the potential discontinuity and the distances are the difference between any two of the difference series values. The average distance for each of D_b and D_a is computed and used to compute the weighted average of the two groups (weighted by n for each group). This weighted average is then compared with the weighted averages for all possible partitionings of D_b and D_a, and if it is found to be small then clustering is indicated. For a more complete description see Mielke *et al.* (1981).

We employ MRPP using a windowing technique similar to Karl and Williams (1987) by taking 12 years (places) from the difference series on either side of a potential discontinuity and test the significance of the difference at the 0·05 probability (95 per cent significance) level. If another identified potential discontinuity is within 12 places of the current discontinuity, the window is truncated at the second potential discontinuity. For example, if a second discontinuity is only 8 years (places) earlier, the window runs from −8 years to +12 years of the discontinuity being tested. We use a window with this technique because we found that undetected discontinuities, although usually small (<0·5°C), can cause significant discontinuities to appear statistically insignificant. Using a smaller window reduces the effects of undetected discontinuities in most cases, however, there is a trade-off between window size and statistical significance. The larger the window that contains no undetected discontinuities, the more likely smaller discontinuities will be found significant, however, the larger the window the more likely it will contain undetected discontinuities.

Once all the significant discontinuities have been identified, the bias values are calculated from the difference series, again using a window similar to the significance testing. For example, if two discontinuities were found in a 100-year time series, say at years 31 and 61, the first correction would be to find the bias value for the most recent discontinuity. This would be the difference in the mean of years 49–60 in the difference series and the mean of years 61–72. This correction would be applied from the beginning of the time series (year 1) through to the year prior to the discontinuity (year 60), then the second bias (for the discontinuity at year 30) would be found as the difference between the mean of years 19–30 and 31–42 of the difference series.

The entire test is performed again on the adjusted candidate series in order to determine if any hidden discontinuities, which may have been masked by the larger discontinuities, are present. If no further discontinuities are found the candidate series is considered to be as homogeneous as possible. If additional discontinuities are found, they are tested for significance, and the additional adjustments applied. This process continues until no further discontinuities are found.

During this process all verified discontinuities are held for a final processing pass. In other words, the adjusted series from earlier passes are discarded, and all verified discontinuities from all passes through the time series are used with the original, unadjusted candidate time series to produce the final adjusted time series. Furthermore, we restrict discontinuities to those that are further than 5 years (places) apart.

EVALUATION OF TECHNIQUES

In a previous paper, Easterling and Peterson (1992) evaluated a number of techniques that can be used to detect and adjust climatic time series. Following is a brief review of the results of that work. The methodology that was originally introduced in Easterling and Peterson (1992) to evaluate the various tests is also applied here. This is a Monte Carlo-type approach where 1000 candidate–reference time series pairs each of which represent 100 year annual temperature series are simulated using a random number generator. Each time series is forced to be autocorrelated using an AR(1) process, then each candidate–reference pair is forced to be correlated by taking the z score value for a given year in the reference series, multiplying it by a constant ranging from 0·5 to 2·0 (depending on the level of correlation desired), and adding this value to the corresponding candidate series z-score value. Once this is done, the candidate series is restandardized to a mean of zero and standard deviation of one, then both the candidate and reference series are transformed into simulated temperature time series with a mean of 23·0 for each candidate, and 25·0 for each reference. The use of simulated as opposed to real climatic time series, allows us to begin with homogeneous time series and impose discontinuities of any magnitude, at any place in the series. Real climatic time series, even those with accompanying station history files, always have the possibility that undocumented discontinuities exist that may hinder effective technique evaluation.

For the initial test in Easterling and Peterson (1992) we imposed a single discontinuity on each of the 1000 candidates by adding a specified value (0·5–2·0) starting at year (place) 65. The discontinuities, therefore, corresponded to half to twice the standard deviation of each series. The results are shown in Table I. The other three tests in this table are Student's t-test, and two regression-based methods referred to as Reg Y-hat and Rev avg Y-hat. These are similar to the two-phase approach described in this paper, but use confidence intervals to determine the significance of discontinuities. The results clearly indicated that the Alexandersson (1986) and Potter (1981) methods, both variations of the standard normal test, are best at identifying the smaller discontinuities. However, when the discontinuity became large each method was able to correctly identify the discontinuity. Further evaluation indicated that the Alexandersson (1986) method was slightly more sensitive than the Potter (1981) method, however, we discovered in subsequent analyses that in certain instances the standard normal method did not perform well. This led to the development of our technique.

As a first step in the evaluation process we subjected our regression approach to the same elimination process that we used to identify the Alexandersson technique as the best of the existing techniques. This process used 1000 simulated candidate–reference pairs, as described above, not only with one small + 0·5 (half the standard deviation) discontinuity imposed beginning at year 50 but also multiple discontinuities.

Table I. Percentage of simulated time series with the date of the inhomogeneity identified correctly to within 1 year for various techniques, by the magnitude of the discontinuity (from Easterling and Peterson 1992)

	Discontinuity			
	0·5	1·0	1·5	2·0
Alexandersson	50·8	87·6	97·2	99·8
Potter	47·3	84·1	96·3	99·4
t-test	45·7	86·1	95·9	99·0
Reg Y-hat	24·7	75·9	94·5	98·9
Reg avg Y-hat	31·2	74·4	92·6	97·6
Double-mass	30·1	73·8	92·3	98·0
CUSUM	30·7	73·0	92·0	97·0

The candidate–reference correlation for the 1000 pairs was typically 0·75. We then applied our regression approach and the Alexandersson test implemented in a similar algorithm (subdividing the difference series as described above) to the simulated data set.

The results (Table II) indicate that the Alexandersson test implemented in our application strategy is somewhat better than our regression approach at identifying one small discontinuity, with the Alexandersson algorithm identifying over 60 per cent of the discontinuities within ± 2 years and our approach identifying only 40 per cent. Clearly the Alexandersson approach is more robust at finding one small discontinuity. However, a much more telling evaluation occurred when we simulated a data set of 1000 candidate–reference pairs with two discontinuities relatively close in time. For this data set we imposed two discontinuities on each of the candidate time series, $-1{\cdot}0$ at year 45, and $+1{\cdot}0$ at year 55. The results of this test (second row, Table II) shows our technique to be clearly superior, with the Alexandersson approach finding only 298 within ± 2 years of a possible 2000 discontinuities, and our approach finding 1737. Furthermore, the error rate for the Alexandersson method was 41 per cent, and ours was 31 per cent. (The percent correct is the number identified correctly (within 2 years) out of the total number of imposed discontinuities. However, the error rate is defined as the percentage wrong out of the total found, therefore the total number found can exceed the number of imposed discontinuities in the simulated data set.) Because it is essentially a weighted difference of means test, it appears that the Alexandersson test has problems with offsetting discontinuities, particularly those close together in time.

Table II also contains the results of applying the Alexandersson and our tests to other simulated data sets. The results from these additional tests confirm that the two tests perform similarly until the number of discontinuities becomes greater than four. At this point, the performance of our technique becomes better, as measured both by the number of discontinuities found and the error rate.

Given the above results, it remains for us to apply our approach to observed time series. To do this we chose annual average temperature time series for seven stations from the USHCN, a network of 1200 high quality stations selected from the USA Cooperative Network. Each time series was complete (no missing years) and unadjusted for approximately 100 years. A reference series was developed for each station using the methodology described in Peterson and Easterling (1994) with the reference stations drawn from the remaining stations in the USHCN. As described earlier, the method we use to calculate adjustment factors is to take the 12 years (places) on either side of a discontinuity, take the mean of the difference series for each set of 12 years, then take the difference in the two means.

The results for two of these stations are shown in Figure 2. The objective of the methodology, in essence, is to adjust the data such that the resulting difference series after adjustment contains no significant trend,

Table II. Results of comparison with 1000 simulated candidate–reference pairs

Discontinuities—magnitude at year	Alexandersson (standard normal) (per cent)[a]		Easterling–Peterson (per cent)[a]	
	Found	Error	Found	Error
+0·5 at 50	63	51	40	63
1000 total	(634)	(652)	(400)	(702)
−1 at 45, +1 at 55	15	41	87	31
2000 total	(298)	(205)	(1737)	(783)
+1 at 25, −1 at 50, +1 at 75	90	26	90	24
3000 total	(2709)	(933)	(2688)	(839)
−0·5 at 25, +1 at 35,	78	27	75	23
−1·5 at 65, −0·75 at 80				
4000 total	(3133)	(1174)	(2977)	(904)
−0·5 at 15, +0·5 at 25,	49	26	61	23
−1·5 at 60, +0·5 at 90	(2457)	(872)	(3025)	(916)
5000 total				

[a] Actual numbers given in parenthesis.

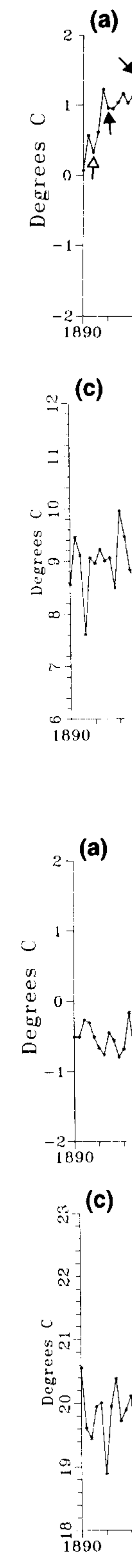

Figure 2
Spokan
Easterl

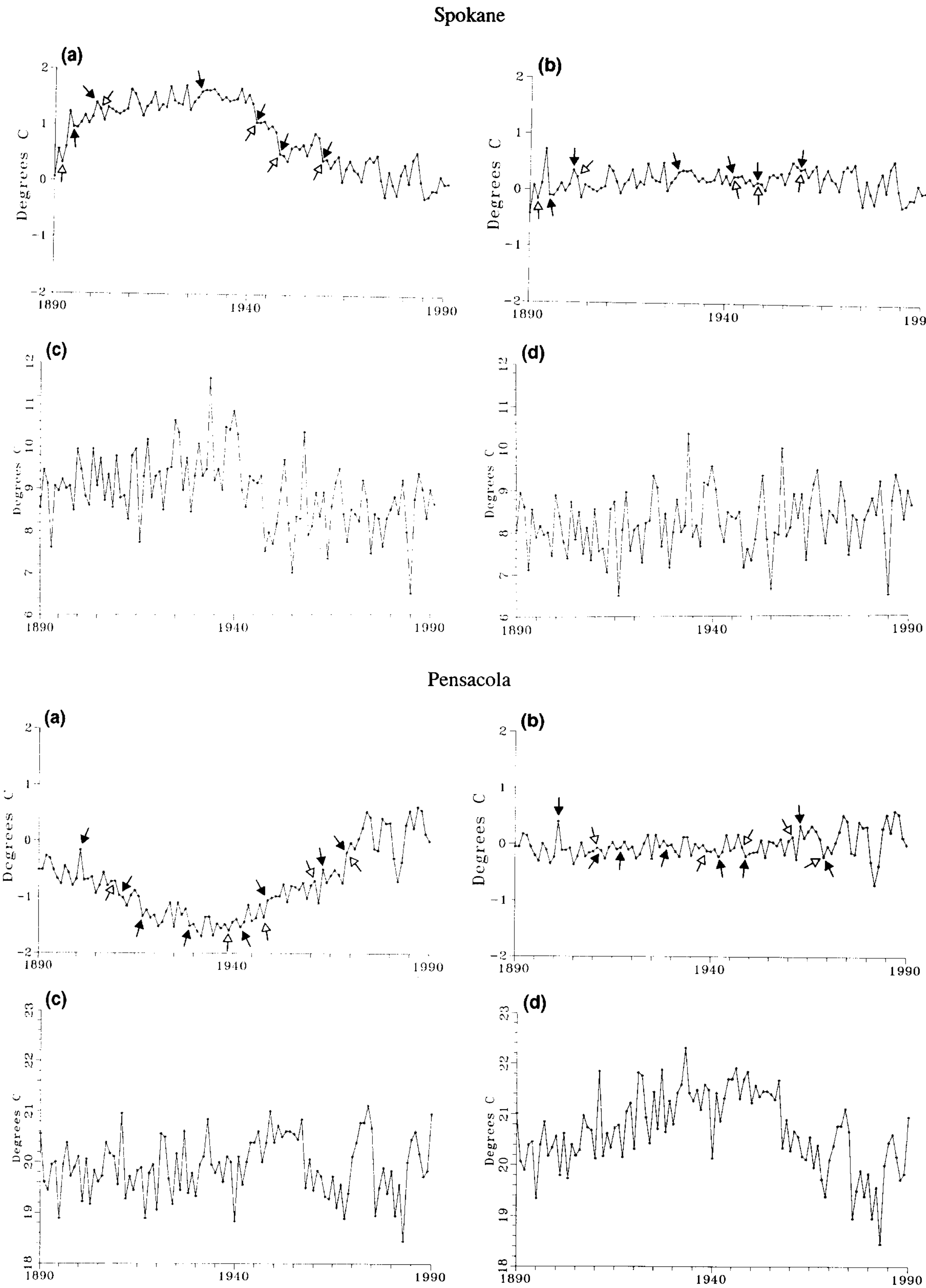

Figure 2. Unadjusted (a) and adjusted (b) difference, and unadjusted (c) and adjusted (d) annual temperature time series for: Spokane, Washington (top), and Pensacola, Florida (bottom). Black arrows indicate discontinuities identified by the Easterling–Peterson technique, open arrows indicate potential discontinuities identified from station histories (see Table III)

with the remaining variance being white noise. The before and after adjustment difference series for each of the stations shows that the method does a good job of producing a difference series where the variance appears to be white noise. Furthermore, the resulting adjusted temperature time series for the stations correspond well with the regional trends for each area. For example, the unadjusted time series for Pensacola, Florida shows a slight increase over time, which is uncharacteristic of the regional trend for the south-eastern USA. However, the adjusted time series reflects more closely the regional trend of some warming to the late 1940s, then cooling afterward. The reverse is true of the temperature plots for Spokane, Washington, with the unadjusted time series showing cooling after the 1940s due to a series of discontinuities, but the adjusted time series showing some overall warming.

The question remains as to how valid the discontinuities and adjustments are for these stations. To evaluate this, metadata files for the USHCN were consulted. The USHCN metadata files contain extensive station histories, which document most station changes that can cause discontinuities. The plots of the unadjusted difference series have the discontinuities found by our technique (solid arrows), and the potential discontinuities identified from the metadata (open arrows). The plots show that nearly all of the station changes indicated in the metadata are found by our technique as significant discontinuities (within a couple of years). Furthermore, the offset values given in Table III show that the offsets are of the proper direction, when the direction of the offset is able to be theoretically specified (as in the height above the ground changes).

In addition, a few discontinuities were found by our method at places where the metadata indicates that no significant station change occurred. These 'undocumented' discontinuities provide additional support for the view that metadata, although as complete as possible for the USHCN, may not contain documentation of all changes that can cause a discontinuity. For example, a discontinuity can occur when a broken thermometer that had column separations is replaced with one that has no column separations, or is calibrated differently. Many of these types of changes are not documented routinely, yet they can cause significant discontinuities.

Table III. Results of comparison between discontinuities found by the Easterling–Peterson technique and potential discontinuities indicated by metadata files for Spokane, Washington, and Pensacola, Florida.

Easterling–Peterson technique		Metadata	
Year	Offset	Year	Cause
Spokane, Washington			
1895	+0·87	1892	Move, and height above ground change
1901	+0·42	1902	Slight elevation change
1928	+0·18	–	–
1942	−0·92	1942	Height above ground change (−21·5 m)
1948	−0·80	1948	Move to airport and height above ground change to standard height
1959	−0·61	1959	Installation of HO60 hygrothermometer
Pensacola, Florida			
1901	−0·22	–	
1911	−0·59	1910	Move and height above ground increase (+5·5 m)
1917	−0·65	–	
1929	−0·35	–	
1942	+0·30	1939	Move and height above ground decrease (−13·2 m)
1949	+0·82	1948	Move to airport and change height above ground to standard (1·5 m)
1963	+0·49	1961	Instrument change from LIG to HO61 hygrothermometer
1969	+1·41	1969	FAA take over and change to HO63 hygrothermometer

CONCLUSIONS

We have presented a new technique for detecting and adjusting for undocumented discontinuities in climatic time series, and compared it to other methods. We feel our technique is more robust at identifying discontinuities than the other techniques we examined when applied to simulated data. In applications to observed time series our method was able to identify nearly all station changes indicated by metadata.

Although our evaluation indicates that our technique is more robust overall, these results must also be viewed in light of the original intent of the two techniques. Alexandersson (1986) clearly states in his assumptions that only one break point occurs, and that the time series would be irretrievably inhomogeneous if more than one break point is indicated. This assumption and the test's effectiveness given this assumption is best illustrated by its superior performance in the initial evaluation with one small discontinuity. However, this assumption, although valid perhaps for small networks of high-quality stations, as Alexandersson (1986) used, is not realistic for application to a global data set. With this in mind, our technique is designed for the assumption that multiple discontinuities are present in many if not all of the time series in the data set. Therefore, excluding all stations with more than one discontinuity would be unrealistic. Lastly, we intend to apply this technique, along with the reference series development methodology presented by Peterson and Easterling (1994) to the Global Historical Climatology Network (GHCN, Vose *et al.*, 1992) to produce a global data set of 'relatively homogeneous' temperature time series.

ACKNOWLEDGEMENTS

The comments of Dr Phil Jones substantially improved this paper. This work was supported by the National Climatic Data Center, NOAA's Climate and Global Change Program and the US Department of Energy through Interagency Agreements DE-AI05-90ER60952 and DE-AI05-900R21956.

REFERENCES

Alexandersson, H. 1986. 'A homogeneity test applied to precipitation data', *J. Climatol.*, **6**, 661–675.

Bradley, R. S. and Jones, P. D. 1985. 'Data bases for isolating the effects of increasing carbon dioxide concentration', in MacCracken, M. C. and Luther, F. M. (eds), *Detecting the Climatic Effects of Increasing Carbon Dioxide*, US Department of Energy, Carbon Dioxide Research Division, pp. 29–53.

Chenoweth, M. 1992. 'A possible discontinuity in the U.S. historical temperature record', *J. Climate*, **5**, 1172–1179.

Conrad, V. and Pollak, C. 1962. *Methods in Climatology*, Harvard University Press, Cambridge, MA, 459 pp.

Easterling, D. R. and Peterson, T. C. 1992. 'Techniques for detecting and adjusting for artificial discontinuities in climatological time series: a review', *Preprints: Fifth International Meeting on Statistical Climatology*, 22–26 June 1992, Toronto, pp J28–J32.

Folland, C. K., *et al.* 1992. 'Observed climate variability and change', in Houghton, J. T., Callander, B. A., and Varney, S. K., (eds), *Climate Change 1992, Intergovernmental Panel on Climate Change (IPCC)*, Cambridge University Press, Cambridge, pp. 135–170.

Gullet, D. W., Vincent, L. and Malone, L. 1991. *Homogeneity Testing of Monthly Temperature Series: Application of Multiple-phase Regression Models with Mathematical Changepoints*, Atmospheric Environment Service, Downsview, Ontario, 47 pp.

Karl, T. R. and Williams, C. N., Jr. 1987. 'An approach to adjusting climatological time series for discontinuous inhomogeneities', *J. Clim. Appl. Meteorol.*, **26**, 1744–1763.

Karl, T. R., Diaz, H. F., and Kukla, G. 1988. 'Urbanization: its detection and effect in the United States climate record', *J. Climate*, **1**, 1099–1123.

Karl, T. R., Williams, C. N., Jr., Quinlan, F. T. and Boden, T. A. 1990. *United States Historical Climatology Network (HCN) Serial Temperature and Precipitation Data*, Environmental Sciences Division, Publication No. 3404, Carbon Dioxide Information and Analysis Center, Oak Ridge National Laboratory, Oak Ridge, TN, 389 pp.

Kohler, M. A. 1949. 'Double-mass analysis for testing the consistency of records and for making adjustments', *Bull. Am. Meteorol. Soc.*, **30**, 188–189.

Mielke, P. W., Berry, K. J. and Brier, G. W. 1981. 'Application of multi-response permutation procedure for examining seasonal changes in monthly mean sea-level pressure patterns', *Mon. Wea. Rev.*, **109**, 120–126.

Peterson, T. C. and Easterling, D. R. 1994. 'Creation of homogeneous composite climatological reference series', *Int. J. Climatol.*, **14**, 671–680.

Potter, K. W. 1981. 'Illustration of a new test for detecting a shift in mean in precipitation series', *Mon. Wea. Rev.*, **109**, 2040–2045.

Quayle, R. G., Easterling, D. R., Karl, T. R. and Hughes, P. Y. 1991. 'Effects of recent thermometer changes in the cooperative station network', *Bull. Am. Meteorol. Soc.*, **72**, 1718–1724.

Rhoades, D. A. and Salinger, M. J. 1993. 'Adjustment of temperature and rainfall records for site changes', *Int. J. Climatol.*, **13**, 899–913.

Solow, A. 1987. 'Testing for climatic change: an application of the two-phase regression model', *J. Clim. Appl. Meteorol.*, **26**, 1401–1405.

Vose, R. S., Schmoyer, R. L., Steurer, P. M., Peterson, T. C., Heim, R. R., Karl, T. R. and Eischeid, J. K. 1992. *The Global Historical Climatology Network: Long-term Monthly Temperature, Precipitation, Sea-level, and Station Pressure Data*, Environmental Sciences Division, Publication No. 3912, Carbon Dioxide Information and Analysis Center, Oak Ridge National Laboratory, Oak Ridge, TN, 189 pp.