

COMPARISONS OF OBSERVED NORTHERN HEMISPHERE SURFACE AIR TEMPERATURE RECORDS

J. B. Elsner

Department of Meteorology, Florida State University, Tallahassee, FL

A. A. Tsonis

Department of Geosciences, University of Wisconsin-Milwaukee, Milwaukee, WI

Abstract. Comparisons are made between three climatic data sets consisting of Northern Hemispheric annual surface air temperature departures over the past 100 years. It is shown that despite the large values of linear cross-correlation between the data records, the long-term trends are significantly different according to the bootstrapping procedure. The confidence of relying on any particular data set for assessing global climate change statistics is thus questioned.

Introduction

Over the past several years, data sets of long-term monthly mean surface air temperature anomalies over the globe have been made available to the scientific community from different sources. As a result, numerous studies concerning the possibility of global climate change have recently been performed utilizing these data [e.g. Jones et al., 1986, Hansen and Lebedeff, 1987, Gruza et al., 1988, Tsonis and Elsner, 1989, Lozowski et al., 1989]. The issue whether the climate over the past century has indeed changed, possibly as a result of growth in the concentrations of greenhouse gases in the atmosphere, or whether the observed trends are natural fluctuations on a longer time scale is still being debated. For this reason there exists great interest in this type of data.

The three surface air temperature data sets which have received the most attention thus far include the global set developed by Hansen and Lebedeff at the Goddard Institute for Space Studies in the USA (data set H), the global set developed by Jones et al. at the Climate Research Unit of the University of East Anglia in the UK (data set J), and the northern hemispheric set developed by Gruza et al. at the Hydrometeorological Scientific Research Center in the USSR (data set G). The data sets have been constructed independently with each group using a different averaging technique and a different observational data base. For example, data set H contains only observations taken from stations located on land whereas data set J contains observations taken from both land stations and aboard ships. Data set H is constructed using averages of stations in equal-area boxes over the globe whereas data set G is constructed using visual interpolation of anomalies from sea level temperature analyses [Gruza and Ran'kova, 1980].

Empirical studies of global climate change have generally relied on a particular data set with comparisons between other similar sets limited to a mention of linear cross-correlations [Hansen and Lebedeff, 1987, Gruza et al., 1989, Jones and Wigley, 1990]. The implicit assumption is that the chosen data set is a representative sample from the population defined as the true northern hemisphere annual surface air temperature record. Figure

1 is the time series of northern hemisphere annual surface temperature departures for the period 1891-1987 for each of the three data sets. The temperature departures are based on the 30 year period from 1951-1980. The linear cross-correlation between data set H and data set J is 0.86, the cross-correlation between data set H and data set G is 0.83; and the cross-correlation between data set J and data set G is 0.79. With reference to these correlations, it is usually stated that the data are in good agreement [e.g. Hansen and Lebedeff, 1987, Jones and Wigley, 1990]. However, since there exists an overall trend in the record the linear cross-correlation coefficient as a measure of covariability can be misleading; and with regard to studies of climate change, it is the trend in these data which has been the focus of widespread concern.

In this paper we use a statistical procedure to compare the three data records and to show that despite the relatively high correlations between them significant differences can be revealed. In particular we address the question of whether or not the three data sets can be considered drawn from the same population. Because data set G contains only northern hemisphere temperature departures and since we are interested in comparing records from all three sources, in the following analysis we restrict our attention to the northern hemispheric portions of data sets H and J.

Method

Two questions are addressed in this study. The first of them concerns whether or not the records are significantly different from one another; while the second concerns whether the records reveal the same linear trend or not. Each of the three records used for comparison consists of 97 northern hemispheric annual temperature departures given in time series (Figure 1). Due to the heavy autocorrelation that exists in the departure records [Kuo, 1990] we choose to work with difference records instead. A difference record is constructed by subtracting annual departures of one data set from annual departures of another data set. For example, we obtain difference record H-J by subtracting the annual departures of data set J from the annual departures of data set H. Likewise we obtain difference record H-G by subtracting annual departures of data set G from annual departures of data set H and difference record J-G by subtracting departures of data set G from departures of data set J. The three difference records constructed in this manner are shown in Figure 2.

The effect of subtracting the records is apparent by comparing the autocorrelation function of data set H with the autocorrelation function of difference record H-J (Figure 3). Clearly the serial correlation is substantially less for the difference record compared to the original departure record. We therefore replace the original heavily autocorrelated time series with difference records each consisting of 97 uncorrelated values. We note here, however, that although the serial correlation in the difference records is small it is not zero and therefore the

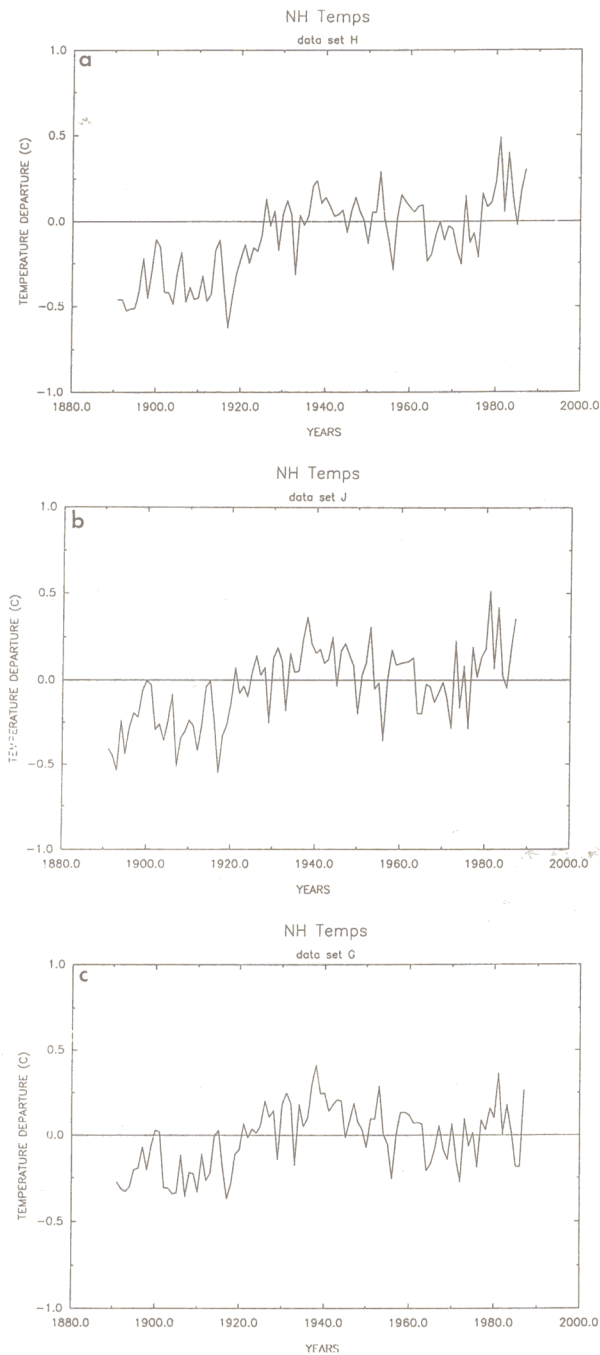


Fig. 1. Northern hemisphere annual surface air temperature departures in $^{\circ}\text{C}$ from 1891-1987 based on the period 1951-1980 from (a) Hansen and Lebedeff, 1987 (data set H), (b) Jones et al., 1986 (data set J), and (c) Gruza et al., 1988 (data set G). Each data set represents a different data base and different averaging techniques. An overall trend is evident in all three records.

statistical results which follow may be somewhat biased.

During the past two decades numerous new statistical techniques were developed to take advantage of the increasing computational efficiencies of computers. One such method, developed recently is called the bootstrap method [Efron, 1982, Diaconis and Efron, 1983]. Like many of the new computationally intensive statistical techniques the bootstrap method is simple in principle and has the meritorious advantage that it does not require the

data to have a gaussian distribution. Another advantage of the new methods, including the bootstrap, is that they allow for investigations of properties of a statistical sample which cannot easily be manipulated analytically. Further, and perhaps most important is that these methods can be applied to small data sets while still providing a reliable evaluation of confidence limits on the statistic of interest. These methods are therefore natural candidates for use in studying climate change problems where the data samples are generally of relatively small size. In this study we apply the bootstrap technique to investigate the sample properties of the difference records.

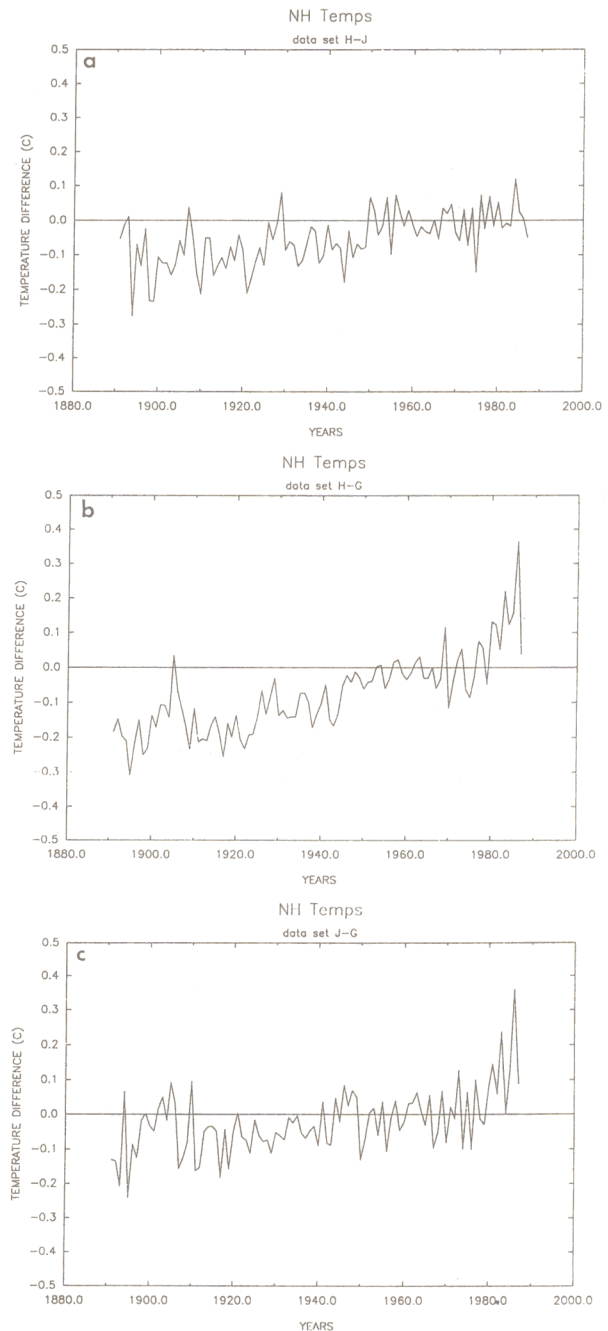


Fig. 2. Difference records in $^{\circ}\text{C}$ constructed by subtracting annual northern hemisphere surface temperature departures of (a) data set J from data set H (difference record H-J), (b) data set G from data set H (difference record H-G), and (c) data set G from data set J (difference record J-G).

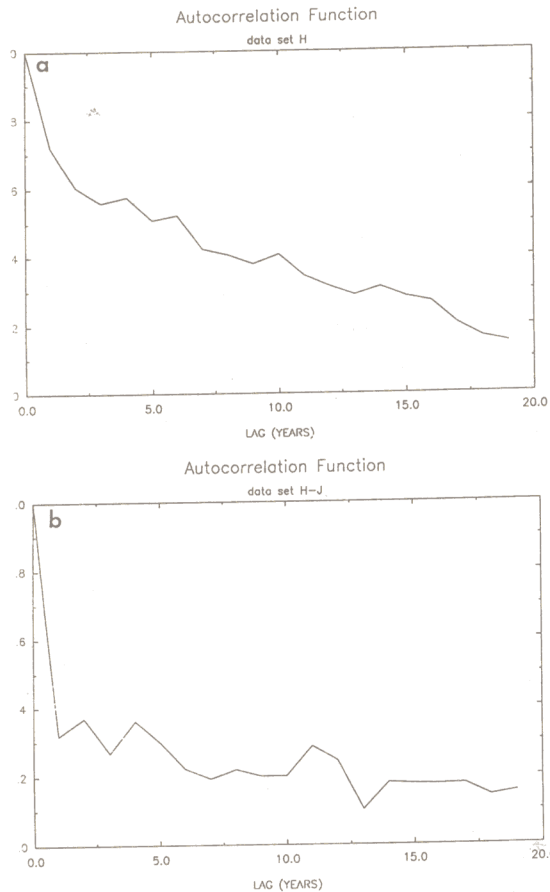


Fig. 3. Autocorrelation functions plotted over the first 20 lags (years) of (a) northern hemisphere surface air temperature departures from data set H and (b) difference record H-J (data set H minus data set J).

The bootstrap procedure is a way of estimating the statistical accuracy of a sample statistic by generating artificial samples from the original sample. For this study artificial samples (called bootstrap samples) are generated by placing all the years with their corresponding temperature difference in a bin and then randomly drawing, with replacements, samples of size 97. In this way a large number of bootstrap samples can be obtained. A frequency distribution of the sample statistic is then plotted based on the number of samples generated. The distribution gives empirical confidence intervals on the statistic that was calculated from the original sample.

One of the problems in climate change research is the limited length of the available data sets which makes it difficult to evaluate the confidence of any computed statistic. The bootstrap technique provides estimates of uncertainty by generating artificial samples from the original sample and evaluating the resulting frequency distribution.

Results and Conclusions

The average difference over all 97 years for difference record H-J relative to the 1951-1980 period is -0.05°C indicating that the hemispheric temperatures of Jones et al. are slightly warmer than those of Hansen and Lebedeff. Of course the natural question is whether or not the difference is significant from zero. To address this question we generate bootstrap sample means as described in the previous section. The frequency distribution of the mean

resulting from 10^4 bootstrap samples is shown in Figure 4a. The distribution does not overlap the zero-difference mark which allows us to conclude that data set H is significantly distinct from data set J in terms of their means. Repeating the procedure for the means of difference records H-G and J-G (Figure 4b,c) we conclude that all three hemispheric surface temperature records exhibit significantly distinct mean temperatures (i.e. nonzero differences) with the distinction between data sets J and G being a minimum and the distinction between data sets H and G being a maximum. We note that the distribution of the sample mean for difference record J-G does overlap the

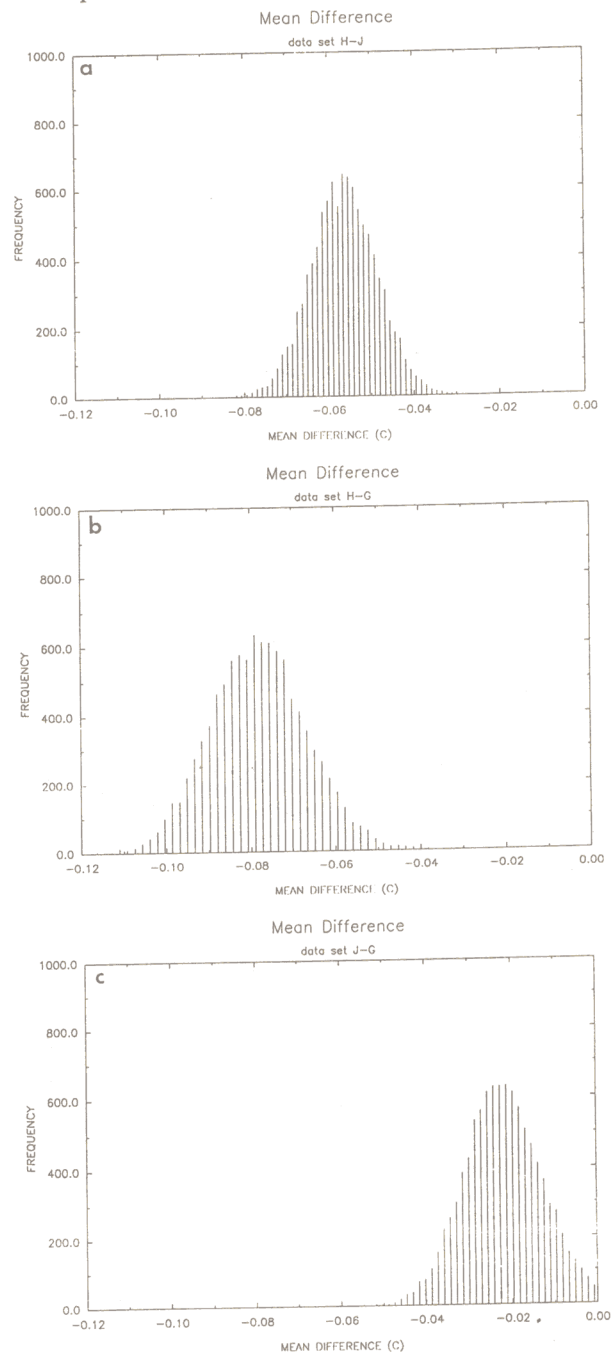


Fig. 4. Frequency distributions of the mean plotted for 10^4 bootstrap samples for (a) difference record H-J, (b) difference record H-G, and (c) difference record J-G. The ordinate scale is the number of times the bootstrap mean fell into a given interval. Note that all three distributions are located to the left of a zero mean difference.

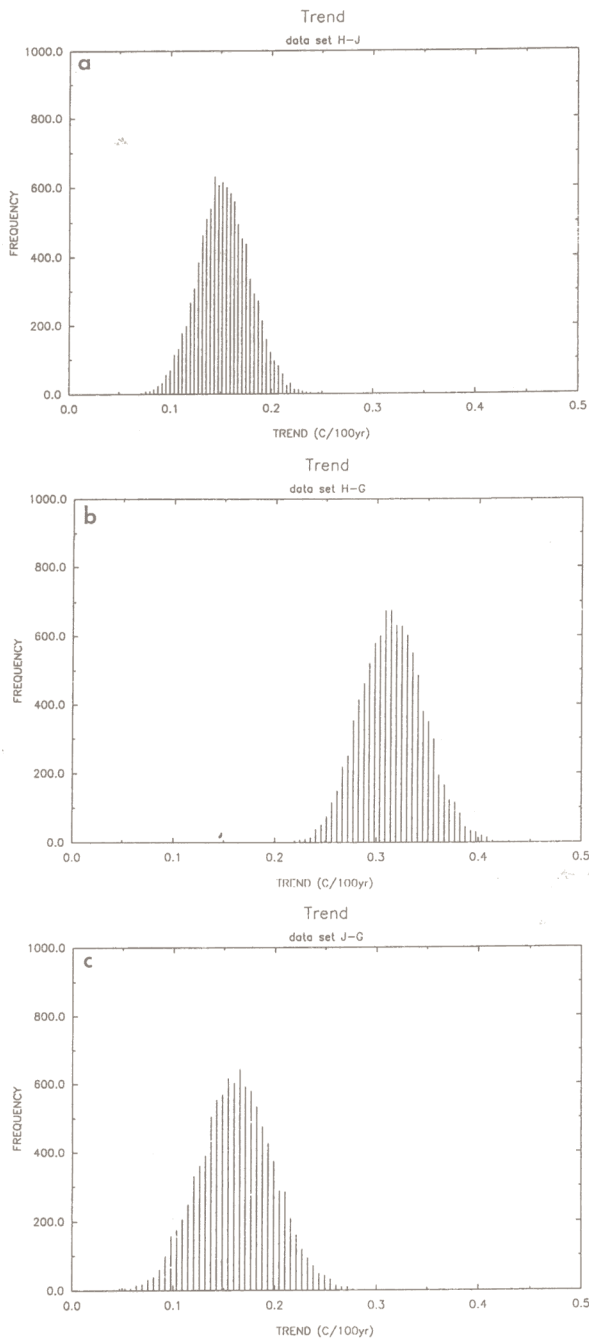


Fig. 5. Frequency distributions of the trend plotted for 10^4 bootstrap samples for (a) difference record H-J, (b) difference record H-G, and (c) difference record J-G. The ordinate scale is the number of times the bootstrap trend fell into a given interval. Note that all three distributions are separated from zero indicating significant differences between long-term surface temperature trends given by each of the three data sets.

zero-difference mark, however, the central significant region of the distribution is clearly offset from zero.

The more important question concerning these data is the significance of a long-term trend. It is therefore important to ascertain to what extent the trends in each of the records are similar. Again, since the original time series' contain serial correlations, we must rely on the difference records to answer this question. A nonzero trend in the difference record will indicate that the original records support distinct trends.

The trend in difference record H-J is $+0.15^\circ\text{C}/100\text{ yr}$ indicating that the temperature data of Hansen and Lebedeff support a larger long-term trend than do the data of Jones et al. Again, the natural question is whether or not the difference trend is significant from zero (identical trends). To answer this question we generate 10^4 bootstrap sample trends from the original difference record. The frequency distribution of the trend is shown in Figure 5a. Clearly the distribution is removed from zero allowing us to conclude that the long-term temperature trend in data set H is significantly distinct from the temperature trend in data set J. Repeating the procedure for the trends of difference records H-G and J-G (Fig. 5b,c) we conclude that all three northern hemispheric surface temperature records exhibit significantly distinct long-term trends.

Based on the above results we conclude that the differences in the observed surface temperature records are significant and that at least two of the three data sets do not represent the true population. Despite the large cross-correlations between the data sets the different construction and sampling techniques along with the different data bases result in statistically different estimates of the true temperature trend over the past nearly 100 years. Finally we comment that even though all three data sets show similar tendencies (positive long-term trends) this study raises the question on the confidence these data sets provide in arguing precise temperature trends over hemispheric or global scales.

References

- Diaconis, P. and B. Efron, Computer-intensive methods in statistics, *Sci. Amer.*, **248** (5), 116-130, 1983.
- Efron, B., *The jackknife, the bootstrap and other resampling plans*, SIAM, 92 pp., 1982.
- Gruza, G. V., and E. Ya. Ran'kova, Structure and variability of observed climate. Air temperature of the Northern Hemisphere, *Gidrometeoizdat*, Leningrad, 1980.
- Gruza, G. V., E. Ya. Ran'kova, and E. V. Rocheva, Analysis of global data variations in surface air temperature during instrument observation period, *Meteor. Gridr.*, **16**, 24, 1988.
- Hansen, J., and S. Lebedeff, Global trends of measured surface air temperature, *J. Geophys. Res.*, **92**, 13,345-13,372, 1987.
- Jones, P. D., T. M. L. Wigley, and P. B. Wright, Global temperature variations between 1861 and 1984, *Nature*, **322**, 430-434, 1986.
- Jones, P. D. and T. M. L. Wigley, Satellite data under scrutiny, *Nature*, **344**, 711, 1990.
- Kuo, C., C. Lindberg, and D. J. Thomson, Coherence established between atmospheric carbon dioxide and global temperature, *Nature*, **343**, 709-714, 1990.
- Lozowski, E. P., R. B. Charlton, C. D. Nguyen, and K. Szilder, Some aspects of the inter-annual variability and persistence of global atmospheric and oceanic surface temperatures, *Clim. Bull.*, **23**, 60-66, 1989.
- Tsonis, A. A. and J. B. Elsner, Testing the global warming hypothesis, *Geophys. Res. Lett.*, **16**, 795-797, 1989.

J. B. Elsner, Department of Meteorology B-161, Florida State University, Tallahassee, FL 32306.

A. A. Tsonis, Department of Geosciences, University of Wisconsin-Milwaukee, P.O. Box 413, Milwaukee, WI 53201

(Received: January 18, 1991;

Accepted: February 14, 1991)