JOURNAL OF GEOPHYSICAL RESEARCH, VOL. 103, NO. D20, PAGES 25,967–25,974, OCTOBER 27, 1998

First difference method: Maximizing station density for the calculation of long-term global temperature change

Thomas C. Peterson[1], Thomas R. Karl[1], Paul F. Jamason[2,3], Richard Knight[1], and David R. Easterling[1]

Abstract. The calculation of global land surface air temperature trends using the instrumental record has been based primarily upon two methods of maximizing the availability of station records. *Hansen and Lebedeff* [1987] developed a technique that is still used today, known as the reference station method; *Jones et al.* [1986a] popularized the climate anomaly method in their calculations of global temperature trends. In this paper we introduce yet another approach designed to maximize station records, referred to as the first difference method. To test the sensitivity of global temperature trend analysis to the method used, we calculate worldwide-averaged land surface mean temperature using each of these methods with an identical data base, the Global Historical Climatology Network. For further comparisons, a global climate model (GCM) transient model simulation is interpolated to the Global Historical Climatology Network station locations and the three techniques are then applied to data interpolated to the station locations from the model. The Intergovernmental Panel on Climate Change (IPCC); [*Nicholls et al.*, 1996] estimated a global land and ocean temperature change of 0.45°C ± 0.15°C since the 19th century. Their assessment of the uncertainty associated with this temperature trend did not specifically address the differences that the method of calculating a global temperature time series might produce. Our results indicate that the differences in 1880-1990 trends produced by these three different methods are only a few hundredths of a degree centigrade per 100 years on trends of approximately 0.5°C/100 years. This is quite small compared to the 0.15°C/100 years uncertainty associated with the IPCC global land and ocean assessment which included factors such as data homogeneity which are not addressed here. Indeed, our results indicate that the source of differences in trends is more likely to be the method used to calculate a linear trend from a global temperature time series than the method used to create the global temperature time series. The modeled results confirm this finding but highlight other important characteristics: the reference station method has uncharacteristically low interannual variance, more similar to time series from the entire globe (land and ocean) than the global land area from which the data were observed. This lower variance can impact the statistical significance associated with linear trends.

1. Background

The data and techniques developed and applied by *Jones et al.* [1986a, b] and *Jones* [1988, 1994] to estimate long-term changes of worldwide land surface air temperature (hereinafter simply global temperature) have been the primary source of data and information on land surface air temperature change used by the IPCC [*Folland et al.*, 1990, 1992; *Nicholls et al.*, 1996]. Other estimates of global temperatures include those provided by *Hansen and Lebedeff* [1987, 1988] and *Vinnikov et al.* [1990]. IPCC [*Folland et al.*, 1990; *Nicholls et al.*, 1996] discussed and identified a large number of uncertainties with calculations of global temperature change. As a result, an uncertainty of ± 0.15°C was assigned to the global land and ocean warming rate of 0.45°C since the late 19th century [*Nicholls et al.*, 1996], although the standard error for individual years is in the 0.05°–0.09°C range [*Jones et al.*, 1997] and decadal trends have a substantially larger uncertainty than century trends [*Karl et al.*, 1994].

The calculation of large-area average changes of temperature and of other geophysical quantities using long-term in situ observations is hampered by a variety of problems. The homogeneity of long-term observations is one of the problems which has received a great deal of well-deserved attention [e.g., *Jones et al.*, 1986a; *Easterling and Peterson*, 1995]. Another common problem is the inadequate spatial coverage [*Karl et al.*, 1994; *Madden et al.*, 1993; *Parker et al.*, 1994]. *Karl et al.* [1994] have assessed the magnitude of the errors and biases relative to decadal-to-century scale trends of temperature associated with incomplete worldwide coverage. In that study an assumption was made that the data from every land surface observation site in (the archives of) the Global

[1]National Climatic Data Center / NESDIS / NOAA, Asheville North Carolina

[2]DynTel Inc., National Climatic Data Center, Asheville, NC 28801

[3]Now at Cecil H. and Ida M. Green Institute of Geophysics and Planetary Physics, Scripps Institution of Oceanography, La Jolla, California.

Paper number 98JD01168.
0148-0227/98/98JD-01168$09.00

Historical Climate Network (GHCN); [*Peterson and Vose*, 1997] database could be used to estimate global temperature changes. Such an assumption however, is overly optimistic, as many stations may not have an adequate period of record to be used with the Climate Anomaly or Reference Station Methods.

The methods used by *Hansen and Lebedeff* [1987; hereinafter referred to as H&L], *Jones* [1988, 1994], *Vinnikov et al.* [1990], and IPCC [*Folland et al.*, 1990, 1992; *Nicholls et al.*, 1996] do not use all of the available data for a variety of reasons. Some long-term stations do not have enough observations available during an overlap with other stations as required in the H&L method. Some stations have too many missing data during a common reference period (often 30 years long) to allow reliable calculation of anomalies for the climate anomaly method (CAM) as in the work of *Jones* [1994; hereinafter referred to as J]. Alternately, *Vinnikov et al.* [1990; hereinafter referred to as V] simply tried to get a homogeneous spatial coverage of stations with complete records. V identified about 300 such stations that operated through the 1980s. The calculation of temperature anomalies is fundamental to the J and V methods because when stations temporarily, or permanently, drop out of the network, the drastically different mean values from observing sites on mountaintops versus low-lying valleys, coastal versus inland locations, etc., prohibits the calculation of climate changes and variations by averaging actual values, even within relatively small grid cells. Therefore anomalies, which are more coherent geographically, have been used.

Our aim is to identify the magnitude of the errors associated with the existing methods of calculating worldwide changes in temperature and to describe and evaluate an alternative technique. The focus is on large-area (worldwide) changes of temperature, but the paper may also be relevant to other long-term in situ geophysical measurements, such as precipitation, sea level pressure, and tide gauge measurements. Owing to our focus on worldwide changes in land surface air temperature, we do not discuss other techniques appropriate for sparse and irregularly distributed data such as sea surface temperatures [*Smith et al.*, 1994, 1998].

2. Techniques and Rationale

2.1 Climate Anomaly Method (CAM)

Anomalies calculated from a common reference period have been used routinely in synoptic climatology to obtain geographical fields of temperature departures from a mean, and this usage has been carried over into global-scale time series analysis of climate-related variables. For temperature, a common practice is to use a 30-year reference period. Stations that have missing data during this common reference period often have their full reference period values estimated through a variety of objective and subjective methods [*Jones*, 1994]. Recently, J updated his normal reference period from 1951-1970 to 1961-1990 and went to great lengths to ensure that a maximum number of stations are used within the reference period [*Jones*, 1994]. To ensure a maximum possible use of stations, J developed a procedure for estimating the 1961-1990 means by comparison with neighboring stations and the 1951-1970 grid point analyses and also using some of the World Meteorological Organization's normals data for the 1961-1990 means. His method did not, however, make use of stations that closed prior to 1961. While it is common practice to use 30-year normals for estimating anomalies of temperature [*World Meteorological Organization* (WMO), 1996], the impact of using too few years has not been well studied. In the lower limit, however, the use of 1 year as a reference period clearly would introduce large biases because in any given year some stations would have experienced a cold year and others a warm year, thereby deflating or inflating the reference temperature for that station relative to other stations. Using short reference periods would introduce biases as individual stations enter or drop out of the network. For details of the gridding and areal-averaging in CAM, see *Jones* [1994].

2.2 Reference Station Method (RSM)

H&L first described an alternative to the common anomaly reference period method of calculating changes of global temperature. They developed a reference station method (RSM) whereby a single reference station, having the longest record, is selected within a grid cell. Their grid cell size was approximately 20° x 30° in the midlatitudes. Within the cell, successively shorter station's temperature records were adjusted so that their time average equaled that of a composite of stations already processed, including the reference station, during a common period of operation. Then, for a given year and month, distance-weighted grid cell averages were calculated (see *Hansen and Lebedeff* [1987] for details). Their method effectively enabled use of stations without requiring that they all have the same common, 30-year reference period of operation as required by CAM. Still, if there were too few years of overlap with the reference station, data from the additional station could not be used. Although this method makes use of more data than the anomaly method, one potential problem relates to the quality of the data from the reference station. The RSM is very dependent on the fidelity to which the reference station actually represents changes in climate as opposed to time-varying biases caused by minor observation site relocations, instrument changes, changes in calibration, and other observing procedures. If the reference station is affected by nonclimatic time-varying biases, not only will its time-varying temperature be inhomogeneous but also all data within the grid cell will be adversely affected. The extent to which this is a significant concern is dependent on the quality of the reference station.

2.3 First Difference Method (FDM)

An alternate method for constructing area-averaged time series is introduced here. First, for each station, we calculate a series of calendar-month first differences, δ_t, using

$$\delta_t = x_t - x_{t-1} \qquad (1)$$

where x_t is the value of geophysical quantity x at time t. For example, when creating a station's first difference series for mean February temperature, we subtract the station's February 1989 temperature from the station's February 1990 temperature. While differencing works well for temperature and pressure, a ratio or normalized difference might work better for some other variables such as precipitation. Figure 1a depicts a hypothetical temperature time series and Figure 1b its first difference series; δ_t represents the time rate of change

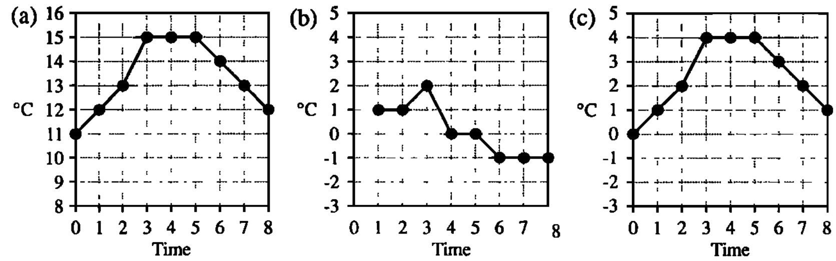


Figure 1. An illustrative example of the first difference method the first difference method (FDM): (a) original time series, (b) first differences, and (c) cumulative sum of first differences.

between times t-1 and t and is the property of fundamental interest in climate change analysis. Second, we carry out all area averaging on the quantity δ, not x. We average, unweighted, into 5°x5° grid cells, then area-average these cells globally. The first difference series can be integrated or area-averaged across observing sites because each value of δ for all observing sites is referenced to the same year. Finally, the cumulative sum of the area averaged first difference series is calculated. An example is depicted in Figure 1c. Note that the character of the time series in Figures 1a and 1c (the area integration step is omitted for clarity) is unchanged by the differencing and cumulative sum operations, but the level is changed such that the value at each time t represents the difference from the original time series at time t=0. The initial level is conveniently set to 0.

For the calculation of long-term changes, the first difference method (FDM) has the advantage of making use of all available data without referencing each observing site to a single reference site or period. Difficulties occur when there are frequent gaps in the data requiring estimates of changes before and after missing values, though it is possible to span missing values using the average δ_t for the missing years. It may also be more sensitive to undetected outliers at the end or beginning of the record than either the RSM or the CAM, so caution should be exercised when applying FDM to geophysical measurements such as precipitation which contain infrequent extremely high or low values. For example, if a large fraction of the stations/grid boxes happen to have their data start with very cold years and end with extremely warm years, these endpoints would add an artificial warming to the final area-averaged time series because they would include the warming first difference part of these extreme events but not the cooling part (i.e., the return to normality after an extremely warm event or the departure from normality to the extremely cold event). This can be especially relevant when the data for a large fraction of the stations/grid boxes ends in a particular year. The impact of this endpoint-induced random walk problem can be decreased by not using the endpoint of an original or first difference time series if it exceeds some threshold measure such as 1σ.

There are some practical advantages of FDM relative to the CAM or the RSM, and to better understand these advantages, it is important to understand the practical limitations of implementing the CAM and RSM. These limitations include (1) many observing sites have significant amounts of data but end prior to the normal reference period or lack adequate overlap with the nearest reference station and therefore either cannot be used in CAM or RSM or are only able to be used by estimating many years of missing data. This will become more relevant with the continuing digitization of early data [e.g., *Peterson and Griffiths*, 1997]. (2) Using the FDM, all changes in the geophysical quantity x are due to actual changes in the time series, not because of a station starting or ending. Figure 2 depicts an example of what could occur in the CAM and the RSM when calculating area-averaged changes. Discontinuities are introduced into the composite time series when a station/grid box starts or stops. This is especially true for a grid box whose trend is quite different from that of the composite time series whether due to inhomogeneities in the data or differing regional trends. Also, the influence of observing sites with shorter records may not impact the overall ΔT of the CAM or RSM composite time series (see Figure 2). Within a grid box, this problem may be less important because homogeneity assessments are often more robust on longer time series. However, for larger areas, one can have whole regions where the data start later and end earlier and therefore have no impact on the total calculated change in temperature.

Because of the limitations of a reference period or an overlap with a reference station, CAM and to a lesser extent RSM could be confined to using substantially fewer stations than FDM. In the CAM presented here, stations were required

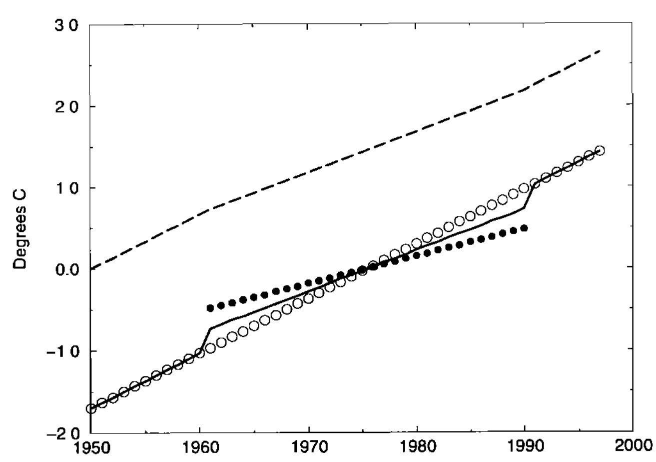


Figure 2. An illustrative example of discontinuities introduced into a composite climate anomaly method (CAM) or reference station method (RSM) time series by an observing site entering and leaving the network. Station data are indicated as open and solid circles. Area-averaged time series created by CAM or RSM is represented by a solid line and FDM by dashed. Since the shorter period of record station has data from 1961 to 1990 and each station's mean for the period 1961-1990 is zero, CAM and RSM produce identical time series. Note also that the total 1950 to 1997 change in temperature for CAM and RSM is not affected by the addition of the shorter station's data, while the FDM's ΔT is.

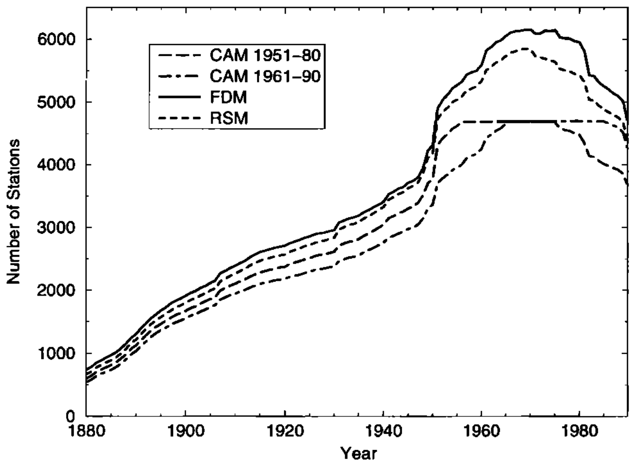


Figure 3. Number of Global Historical Climate Network (GHCN) stations used in CAM, RSM, and FDM by year.

to have at least 25 years of data during the 30-year base period, which is consistent with the recommendations of the World Meteorological Organization of the creation of 30-year normals [*WMO*, 1989]. For the FDM to use a given year's value for a station, only the station's previous year's value is required. In order to use a station's data in the RSM there must be an overlap period between that station and the grid box reference station. For this study we required a combined 20-year overlap of data between that station and the relevant reference station or other station within 1200 km, as per *Hansen and Lebedeff* [1987].

FDM used a total of 7227 stations versus 5134 for RSM and ~4700 for CAM (depending on the base period selected). This numerical advantage for the FDM is present for all years (Figures 3). For example, in 1880 the FDM could only use 740 stations but that is still using 23% more stations than the CAM with the 1951-1980 base period. The area of the globe represented by the time series also varied with time and technique (Figure 4). While a good geographic distribution of the stations can be more important than the actual number of stations, it is clear from Figure 3 that the different number of stations available for use by a given technique impacts the area represented.

Also the size of the grid boxes impacts the results. Using 80 equal-area grid boxes allows greater spatial representation from a limited number of stations than 2592 5°x5° grid boxes. In addition to using 80 equal-area grid boxes, H&L's RSM interpolates station data into grid boxes and subgrid boxes within a 1200 km area of influence. This combination of large grid boxes and interpolation of data allows the RSM approach to take GHCN continental, island, and stationary ship stations and obtain global coverage (see Figure 5 for station locations) but also frequently requires that a single station get processed many times for many different subgrid boxes.

3. Methodology

The Global Historical Climatology Network (GHCN, version 2); [*Peterson and Vose*, 1997] is a monthly temperature, precipitation, and pressure database with over 7000 land surface temperature stations. All GHCN stations have at least 10 years of data, and ~1000 of the stations have over 100 years of data. Area-averaged global temperature time series from 1880 to 1990 were calculated from GHCN station data using the three different methods: FDM, CAM, and RSM. CAM was used with two different commonly used base periods, 1951-1980 and 1961-1990. FDM and CAM used J's 5°x5° grid boxes, while RSM used H&L's 80 equal area grid boxes. The linear trend was calculated on the resultant time series in three different ways: the common least squared deviation (LSD), a least absolute deviation (LAD); [*Mielke*, 1984], and a resistant trend estimate (RTE); [*Emerson and Hoaglin*, 1983]. The LSD approach, because it minimizes the square of the distance to the regression line, is significantly impacted by extreme values, particularly at end points of the time series. Both the LAD approach and the RTE are more robust to outliers.

Additional comparisons were made using modeled data to ensure that our results were not impacted by undetected inhomogeneities in some of the observed data and to provide verification independent of observed station locations. Annually averaged surface temperature output from a Geophysical Fluid Dynamics Laboratory GCM transient run with CO_2 and sulfate concentration corresponding to the observations during the period 1880-1990 [*Haywood et al.*, 1997] was interpolated to GHCN station coordinates to create modeled station data. It should be noted that the simulation has a trend not too different from the observed temperature record [*Haywood et al.*, 1997]. The simulated GHCN stations were ascribed the same period of record and missing data points as their respective observational records. As with the observed data, the three techniques were used to create global mean temperature time series from the modeled GHCN station data. For more direct comparisons, the FDM was also applied to model data from those stations and grid boxes used in the CAM and RSM because the three techniques have different criteria in selecting which stations to use. Therefore the number of GHCN stations that could be used varied with the technique.

4. Results

The linear trends of 1880-1990 global temperature time series created by FDM, CAM (two different base periods), and RSM are shown in Figure 6. Considering the differences in grid box characteristics, the differences in the number of stations used in the analyses and the differences in the methods, the close agreement in the trends is rather

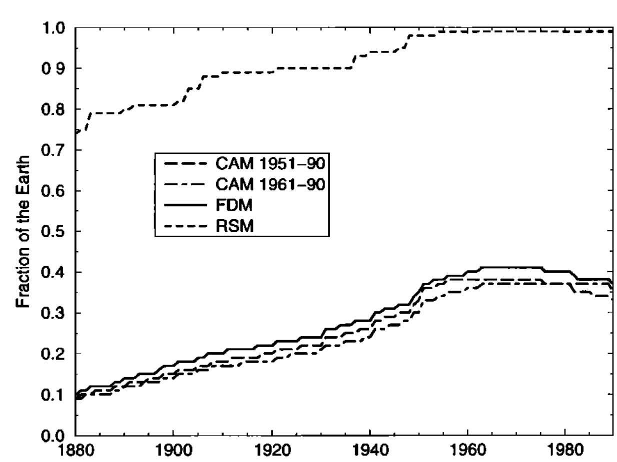


Figure 4. Fraction of the Earth represented by the different time series creation techniques.

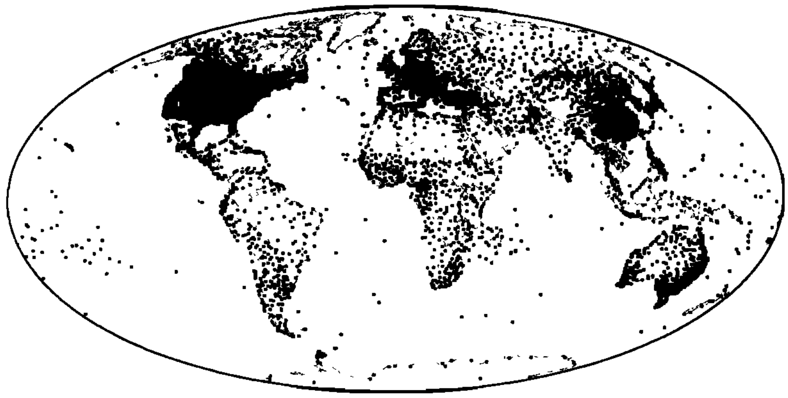
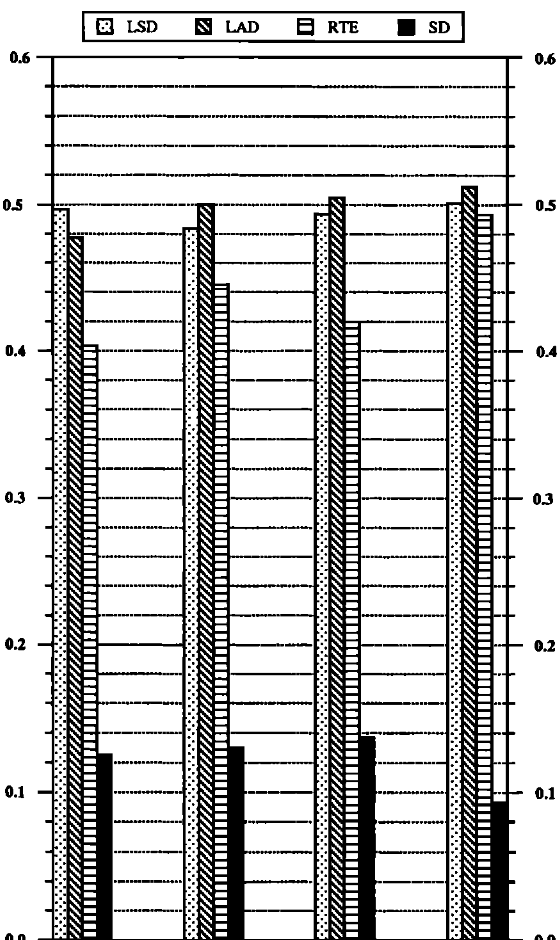


Figure 5. GHCN version 2 mean temperature station locations.

remarkable. a testament to the robustness of the signal. The range (maximum slope minus minimum slope) in LSD slopes is 0.017°C/100 years, LAD 0.035°C/100 years, and RTE 0.090°C/100 years on trends of ~0.5°C/100 years. There is often a larger difference among the LSD, LAD, and RTE linear trends on the same time series than, for example, the LSD trends on the different global temperature time series. The results of modeled data comparisons are shown in Figure 7 which includes trends from three additional runs of the FDM using the same stations and grid boxes used by the CAM and RSM. Figure 7 also includes three different verifications: complete global land grid boxes, global land but excluding Antarctica, and the entire global land and ocean. These time series were calculated from gridded model output directly, not from interpolated station values.

Clearly, the FDM provides a reasonable assessment of global temperature trends. This is also true for RSM and CAM, and the justification has been documented in other articles [e.g., *Hansen and Lebedeff*, 1987; *Jones et al.*, 1986a, b]. One interesting difference in trends between the various results was the modeled data CAM 1961-1990 trends and the FDM run with the same modeled stations. Since these two methods used exactly the same grid boxes, exactly the same within grid box averaging technique (all stations within a grid box are equally weighted), and exactly the same stations. these differences are most likely due to the disparate effects of beginning and ending of shorter-period stations, as illustrated in Figure 2.

Perhaps the most significant difference between the results of the three techniques is the variance. As expected with all three techniques, the variance within a grid box decreases as more stations become available within that grid box, and the variance of large-area averages decreases as more grid boxes with data become available. However, as shown in Figures 6 and 7, the standard deviation for RSM is smaller than the standard deviation for either the FDM or the CAM time series. Because trends in a time series greatly impact the calculation



Technique	FDM	CAM 51-80	CAM 61-90	RSM
Grid box	5° x 5°	5° x 5°	5° x 5°	80
No. of Sta.	7,227	4,681	4,695	5,134

Figure 6. Estimates of trends and standard deviations in observed global temperature time series over the period 1880 to 1990 derived from the three different approaches, the first difference method (FDM), the climate anomaly method (CAM), and the reference station method (RSM). CAM was applied using two different base periods, 1951-1980 and 1961-1990. The grid boxes used was either 2592 5° x 5° or 80 equal area grid boxes. Three estimates for trends, least squares deviation (LSD), least absolute deviation (LAD), and resistant trend estimator (RTE), in degrees centigrade per 100 years are presented. The standard deviation (SD), in degrees centigrade, of a detrended area-averaged time series, is also shown.

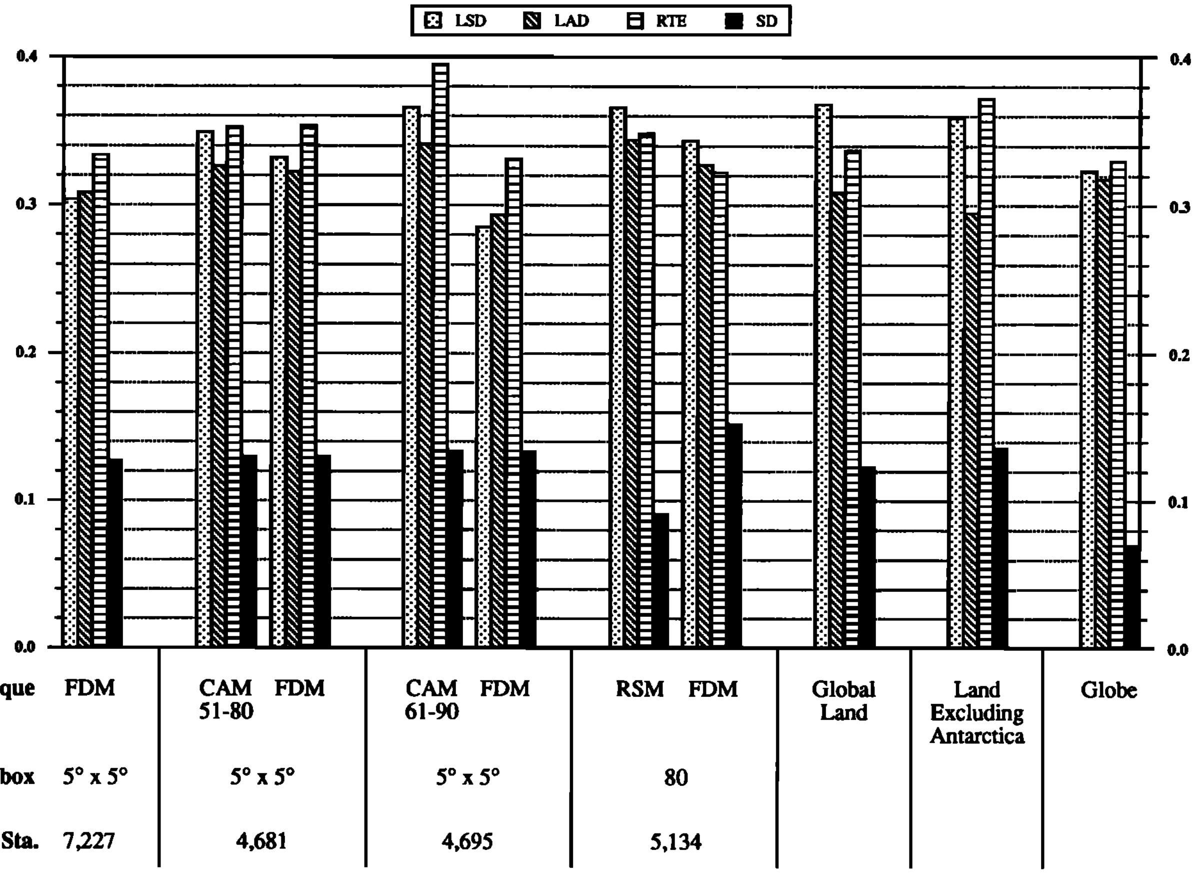


Figure 7. Estimates of global temperature trends and standard deviations as in Figure 6 but based on modeled GHCN station data. Additionally, the FDM was applied to the same stations and grid boxes as the CAM and RSM. The final three sets of graphs are three different versions of model verification: the global land areas, global land excluding Antarctica, and the full globe including land and ocean.

of variance, these standard deviations (s) were calculated from the residuals after the global temperature time series had been detrended by applying a 17-point binomial high-pass filter.

The 5° x 5° FDM and CAM time series had interannual standard deviations ranging from 0.125°C to 0.133°C, while the RSM had an interannual standard deviation (s) of 0.094°C (observed) and 0.091°C (modeled). The modeled data FDM run using the RSM stations and grid boxes had s of 0.152°C, indicating that the lower s for the RSM is not due to the stations selected or the larger grid boxes but rather to the area averaging and interpolation technique. The modeled global land area had an s of 0.123°C and 0.136°C if Antarctica was excluded. This is in close agreement with the FDM and CAM. When the oceans are included as well, s drops to 0.070°C.

Therefore the RSM's variance is closer to the full globe's (land and ocean) than global land area's. This, perhaps, should be expected since RSM interpolates land air temperature stations' data in a manner that gives more weight to coastal stations and simulates full global coverage, as indicated by Figure 4. However, one effect of the decreased variance using the RSM is that the estimated significance of identical linear regression slopes from all three techniques will be different. Generally, the regression from a time series with lower variance would be expected to have greater significance.

Examination of the modeled area-averaged time series in Figure 8a reveals that the RSM has a lower variance throughout the time series. Figure 8b shows the difference among the modeled FDM, CAM, and RSM and global land area (excluding Antarctica) verification time series. It is interesting to note that after the number of stations and the global distribution of these stations dramatically rises in the 1940s (see Figures 3 and 4), the FDM and CAM time series more closely reflects the verification, while the RSM's does not. Indeed, the standard error of temperature for RSM (see Figure 9) when compared to land area verification is greater after 1950 than before 1950. This is due entirely to the fact that, as shown in Figure 4, though the RSM uses land stations, it attempts to produce a full global (including oceans) time series. When its verification is changed to the full globe (land plus oceans), RSM's standard errors of temperature are very similar to FDM's and CAM's using global land as verification. The CAM pre-1950 standard errors shown in Figure 9 are very similar to the global interannual time series standard error calculated by *Jones et al.* [1997] using a GFDL model simulation, but the post1950 errors are smaller.

Examination of Figure 8b reveals that during the 1940s the plotted FDM residuals changed from being generally positive to generally negative. As shown in Figure 8a, the strong model warming during the 1940s in the verification time series is not well represented by any of the three techniques, so this sudden change in model climate must have been largely driven by regions that were not well represented by stations at that

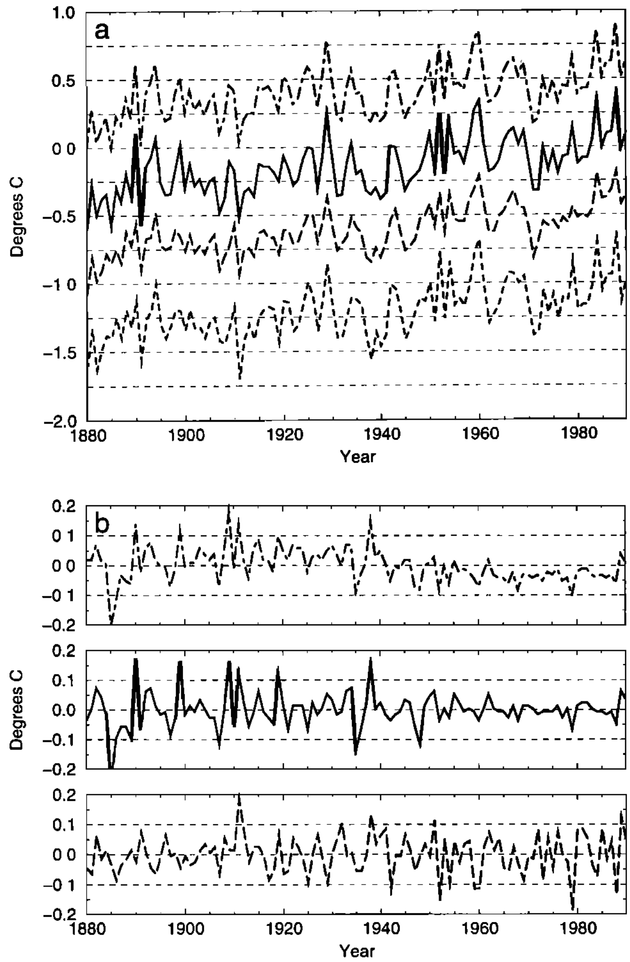


Figure 8. (a) Four time series of modeled annual global temperature: from top to bottom, FDM (dashed-dotted line), CAM 1961-1990 (solid line), RSM (long dashed), and verification (global land areas excluding Antarctica, short dashed). The time series have been offset for easier comparison. Note that the RSM has less year-to-year variability than the others. (b) Annual errors: the FDM (top), CAM (middle), and RSM (bottom) time series shown in Figure 8a minus the verification time series also shown in Figure 8a, adjusted so the mean of each equals zero. Note that as the number of stations and area represented increase (Figures 3 and 4), the FDM and CAM better represent the verification time series.

time. Since all changes in the FDM time series are due to observed changes in climate, the effect of these missed observations remains in the time series.

5. Summary and Conclusions

The first difference method for calculating area-averaged time series of a geophysical variable was applied to observed global temperature data. The results from this method are very similar to two other methods already in use. Indeed, the type of regression used to determine the century-scale linear trend from the global temperature time series causes greater differences than the method used to create the time series. The FDM has some distinct advantages in its ability to optimize the data that can be used and have all the data impacting the final results. However, given the nature of the observed global temperature signal and the distribution and periods of records of the thousands of GHCN stations, these advantages do not appear to make any significant difference in the results. This is in agreement with the *Jones* [1995] determination that the century-scale global temperature signal could be adequately determined from an order of magnitude fewer stations if the stations were well distributed with long periods of record.

The FDM advantages may be important in regional analyses where far fewer stations are available, or for global precipitation analysis, since the soon-to-be-released GHCN version 2 precipitation data set will have over 20,000 stations (compared to ~5000 WMO normals stations), and the digital data for thousands of these stations ended in the 1970s. FDM may also be useful for analyzing some geophysical variables which may have been observed in ways that do not lend themselves to CAM or RSM. For example, most U.S. pan evaporation data come from stations with less than 15 years of evaporation data [*Peterson et al.*, 1995]. Of course, increasing

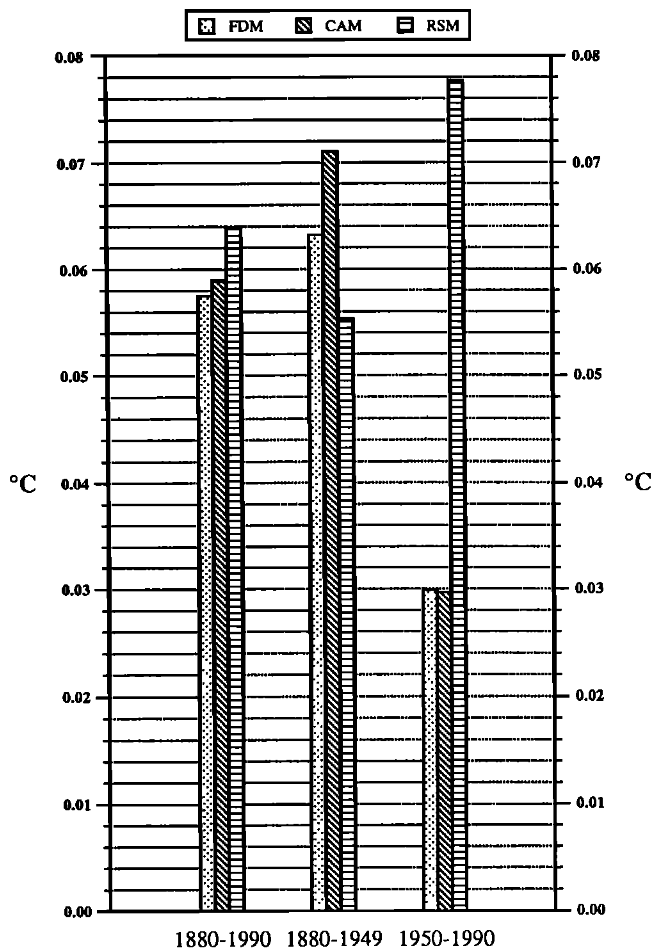


Figure 9. Standard error of modeled FDM, CAM, and RSM using global land areas (excluding Antarctica) for verification. The standard error is the standard deviation of the (calculated minus verification) time series shown in Figure 8b. As shown in Figures 3 and 4, the number of available stations and area represented rise dramatically just before 1950. In response, FDM's and CAM's 1950-1990 error is much lower than the error for the earlier period. While this does not occur with RSM in the analysis presented, it does if the verification used is the complete globe (land and ocean), in which case the RSM standard errors are very similar to FDM's and CAM's.

the number of stations available for any analysis could be expected to make trends more robust, especially when the spatial correlation is low (e.g., precipitation). Indeed, the distinct decrease in the magnitude of the FDM and CAM errors around 1950 shown in Figure 8b corresponds to the number of available stations increasing to ~4000 (Figure 3) and the global 5° x 5° grid box coverage increasing to ~35% (Figure 4) which includes most of the global land area.

On the basis of analyses using GCM temperature fields, all three global temperature time series creation techniques, the first difference, climate anomaly, and reference station methods, are valid approaches to determining global temperature trends. However, what the three methods measure is somewhat different. Rather than simply averaging the station values within each grid cell as the CAM and FDM do, each grid box and subgrid box used in the RSM have an interpolated value which results in reduced variance in the global-average time series. Also, the proportionally heavier weighting given to maritime stations in interpolation over open ocean is likely to be contributing to the smaller variance. The RSM's global temperature time series is closer to the full globe's (land and ocean) than the global land area's. Since the results of the different methods have markedly different variances, when assessing the statistical significance of results from a specific method, the effects that the method has on the derived time series' variance should be considered.

Acknowledgments. We'd like to thank R.J. Stouffer of GFDL for providing model data. This work has been supported by the National Climatic Data Center, NOAA Climate and Global Change Climate Change Data and Detection Program, and a Department of Energy Interagency Agreement. We'd also like to thank the reviewers whose careful reviews and insightful comments significantly improved this article.

References

Easterling, D.R. and T.C. Peterson, The effect of artificial discontinuities on recent trends in minimum and maximum temperatures, *Atmos. Res.*, *37*, 19-26, 1995.

Emerson, J.D., and D.C. Hoaglin, Resistant lines for *y* versus *x*, in *Understanding Robust and Exploratory Data Analysis*, edited by D.C. Hoaglin, F. Mosteller, and T.W. Tukey, pp. 129-165, John Wiley and Sons, New York, 1983.

Folland, C.K., T.R. Karl, and K.Ya. Vinnikov, Observed climate variations and change, in *Climate Change, The IPCC Scientific Assessment*, edited by J.T. Houghton, G.J. Jenkins, and J.J. Emphraums, pp. 195-238, Cambridge Univ. Press, New York, 1990.

Folland, C.K., T.R. Karl, N. Nicholls, B.S. Nyenzi, D.E. Parker, and K.Y. Vinnikov, Observed climate variability and change, in *Climate Change 1992 – The Supplementary Report to the IPCC Scientific Assessment*, edited by J. T. Houghton, B.A. Callendar, and S.K. Varney, pp. 135-170, Cambridge Univ. Press, New York, 1992.

Hansen, J., and S. Lebedeff, Global trends of measured surface air temperature, *J. Geophys. Res.*, *92*, 13, 345-13, 372, 1987.

Hansen, J.,and S. Lebedeff, Global surface air temperatures: Update through 1987, *Geophys. Res. Lett.*, *15*, 323-326, 1988.

Haywood, J., R. J. Stouffer, R. T. Wetherald, S. Manabe, and V. Ramaswamy, Transient response of a coupled model to estimated changes in greenhouse gas and sulfate concentrations, *Geophy. Res. Lett.*, *24*, 1335-1338, 1997.

Jones, P.D., Hemispheric surface air temperature variations: Recent trends update to 1987, *J. Climate*, *1*, 654-660, 1988.

Jones, P.D., Hemispheric surface air temperature variations: a reanalysis and an update to 1993, *J. Climate*, *7*, 1794-1802, 1994.

Jones, P. D., Land surface temperatures - is the network good enough?, *Clim. Change*, *31*, 545-558, 1995.

Jones, P.D., S.C.B. Raper, R.S. Bradley, H.F. Diaz, P.M. Kelly, and T.M.L. Wigley, Northern hemisphere surface air temperature variations, 1851-1984, *J. Clim. Appl. Meteorol.*, *25*, 161-179, 1986a.

Jones, P.D., S.C.B. Raper, R.S. Bradley, H.F. Diaz, P.M. Kelly, and T.M.L. Wigley, Southern Hemisphere surface air temperature variations, 1851-1984, *J. Clim. Appl. Meteor.*, *25*, 1213-1230, 1986b.

Jones, P.D., T.J. Osborn, and K.R. Briffa, Estimating sampling errors in large-scale temperature averages, *J. Clim.*, *10*, 2548-2568, 1997.

Karl, T.R., R.W. Knight, and J.R. Christy, Global and hemispheric temperature trends: Uncertainties related to inadequate spatial sampling, *J. Clim.*, *7*, 1144-1163, 1994.

Madden, R.A., D.J. Shea, G.W. Branstator, J.J. Tribbia, and R. Weber, The effects of imperfect spatial and temporal sampling on estimates of the global mean temperature: Experiments with model data, *J. Clim.*, *6*, 1057-1066, 1993.

Mielke, P.W., Meteorological applications of permutation techniques based on distance functions, in *Handbook of Statistics*, Vol. 4, edited by P. R. Krishnaiah and P. K. Sen, pp. 813-830, North Holland, Amsterdam, 1984.

Nicholls, N., G.V. Gruza, J. Jouzel, T.R. Karl, L.A. Ogallo, and D.E. Parker, Observed climate variability and change, *Climate Change 1995: The Science of Climate Change, Contribution of Working Group I to the Second Assessment Report of the Intergovernmental Panel on Climate Change*, edited by J.T. Houghton, L.G. Meira Filho, B.A. Callander, N. Harris, A. Kattenberg, and K. Maskell, pp. 137-181, Cambridge Univ. Press, New York, 1996.

Parker, D.E., P.D. Jones, C.K. Folland, and A. Bevan, Interdecadal changes of surface temperatures since the late 19th Century, *J. Geophys. Res.*, *99*, 14, 373-14,399, 1994.

Peterson, T.C. and J.F. Griffiths, Historical African data, *Bull. Am. Meteorol. Soc.*, *78*, 2869-2872, 1997.

Peterson, T.C., and R.S. Vose, An overview of the Global Historical Climatology Network temperature data base, *Bull. Am. Meteorol. Soc.*, *78*, 2837-2849, 1997.

Peterson, T. C., V. S. Golubev, and P.Ya. Groisman, Evaporation losing its strength, *Nature*, *377*, 687-688, 1995.

Smith, T. M., R. W. Reynolds, and C. F. Ropelewski, Optimal averaging of sea surface temperatures and associated confidence intervals (1860-1989), *J. Clim.*, *7*, 949-964, 1994.

Smith, T. M., R. E. Livezey, and S. S. Shen, An improved method for analyzing sparse and irregularly distributed SST data on a regular grid: The tropical Pacific Ocean, *J. Clim.*, in press, 1998.

Vinnikov, K., P. Y. Groisman, and K.M. Lugina, The empirical data on modern global climate changes (temperature and precipitation), *J. Clim.*, *3*, 662-677, 1990.

World Meteorological Organization (WMO), Calculation of monthly and annual 30-year normals, World Clim. Data Programme Rep. 10, 11 pp., Geneva, Switzerland, 1989.

World Meteorological Organization (WMO), Climatological normals (CLINO) for the period 1961-1990, *Pub. 847*, 768 pp., Geneva, 1996.

T. C. Peterson, Global Climate Laboratory, National Climatic Data Center, 151 Patton Avenue, Room 120, Asheville, NC 28801. (e-mail: Thomas.C. Peterson@noaa.gov)

(Received September 30, 1997; revised March 25, 1998; accepted March 27, 1998.)