Monitoring Global Monthly Mean Surface Temperatures

KEVIN E. TRENBERTH

National Center for Atmospheric Research, Boulder, Colorado*

JOHN R. CHRISTY

Atmospheric Science Program, University of Alabama in Huntsville, Huntsville, Alabama

JAMES W. HURRELL

National Center for Atmospheric Research, Boulder, Colorado

(Manuscript received 31 May 1991, in final form 3 February 1992)

ABSTRACT

An assessment is made of how well the monthly mean surface temperatures for the decade of the 1980s are known. The sources of noise in the data, the numbers of observations, and the spatial coverage are appraised for comparison with the climate signal, and different analyzed results are compared to see how reproducible they are. The data are further evaluated by comparing anomalies of near-global monthly mean surface temperatures with those of global satellite channel 2 microwave sounding unit (MSU) temperatures for 144 months from 1979 to 1990. Very distinctive patterns are seen in the correlation coefficients, which range from high (>0.8) over the extratropical continents of the Northern Hemisphere, to moderate (~0.5) over tropical and subtropical land areas, to very low over the southern oceans and tropical western Pacific. The physical difference between the two temperature measurements is one factor in these patterns. The correlation coefficient is a measure of the signal-to-noise ratio, and largest values are found where the climate signal is largest, but the spatial variation in the inherent noise in the surface observations over the oceans is the other major factor in accounting for the pattern.

Over the oceans, sea surface temperatures (SSTs) are used in the surface dataset in place of surface air temperature and the Comprehensive Ocean–Atmosphere Data Set (COADS) has been used to show that 80% of the monthly mean air temperature variance is accounted for in regions of good data coverage. A detailed analysis of the sources of errors in in situ SSTs and an overall estimate of the noise are obtained from the COADS by assessing the variability within 2° longitude by 2° latitude boxes within each month for 1979. In regions of small spatial gradient of mean SST, individual SST measurements are representative of the monthly mean in a 2° box to within a standard error of 1.0°C in the tropics and 1.2° to 1.4°C in the extratropics. The standard error is larger in the North Pacific than in the North Atlantic and much larger in regions of strong SST gradient, such as within the vicinity of the Gulf Stream, because both within-month temporal variability and within-2° box spatial variability are enhanced. The total standard error of the monthly mean in each box is reduced approximately by the square root of the number of observations available. The overall noise in SSTs ranges from less than 0.1°C over the North Atlantic to over 0.5°C over the oceans south of about 35°S. Greater daily variability in surface marine air temperatures than in SSTs means that two to three times as many observations are needed per month to reduce the noise in the monthly mean air temperature to the same level as for SST. Tests of the reproducibility of SSTs in analyses from the U.K. Meteorological Office (UKMO) and the U.S. Climate Analysis Center (CAC) and from COADS reveal monthly anomaly correlations on a 5° grid exceeding 0.9 over the northern oceans but less than 0.6 in the central tropical Pacific and south of about 35°S. Root-mean-square differences between CAC and UKMO monthly SST anomalies exceed 0.6°C in the regions where the correlation is lower than about 0.6.

With the marked exception of the eastern tropical Pacific, where the large El Niño signal is easily detected, there are insufficient numbers of SST observations to reliably define SST or surface air temperature monthly mean anomalies over most of the oceans south of about 10°N. The use of seasons rather than months can improve the signal-to-noise ratio if careful treatment of the annual cycle is included. For seasonal means, SST anomalies cannot be reliably defined south of 20°S in the eastern Pacific and south of ~35°S elsewhere except near New Zealand. In light of the noise estimates and the much fewer numbers of observations in the past, difficulties in establishing temperatures from the historical record are discussed.

Corresponding author address: Dr. Kevin E. Trenberth, National Center for Atmospheric Research, P.O. Box 3000, Boulder, CO 80307-3000.

* The National Center for Atmospheric Research is sponsored by the National Science Foundation.

1. Introduction

How well do we know individual monthly mean surface temperatures over the entire globe? How well can we determine the surface conditions for monitoring the earth's climate and, in particular, climate change, such as that expected from increases in greenhouse gases? Monitoring changes is actually easier than determining absolute values because systematic errors, such as those arising from biases in methods of measurement or spatial coverage, are largely removed. As an alternative question, we can ask whether there is some other tropospheric temperature, such as from satellite measurements, that might be suitable for monitoring global temperature changes. These are the questions addressed in this paper.

Several prodigious efforts have been made to reconstruct the surface temperature record for the past hundred years or more from instrumental observations. Fairly complete observations of surface temperature over land now exist, although their continuity back in time is an issue. Over and near Antarctica, useful climatological records are virtually nonexistent prior to about the International Geophysical Year (IGY: 1957), except at Orcadas. Similarly, the Greenland Ice Cap is not monitored. Several compilations of surface air temperature variations for the land areas of the two hemispheres are available (Jones et al. 1986a,b; Jones 1988; Hansen and Lebedeff 1987, 1988; Vinnikov et al. 1990). All agree that, in global terms, a warming of ~0.5°C has occurred since 1880. Substantial corrections need to be applied to the marine surface sea and air temperatures (Folland et al. 1984; Jones et al. 1991; Bottomley et al. 1990; IPCC 1990). Recent decades appear to have been somewhat warmer than any other time in the past 120 years or so, while the period from about 1890 to the 1920s was distinctly colder than average. After 1910, these fluctuations are quite similar to those found from land data alone.

A primary objective of this paper is to assess how well the global surface air temperatures are known for a particular month over the past decade or so, essentially using all of the in situ observations available today. Satellite estimates of sea surface temperatures (SSTs) have not been included in compiling the historical record, although recent Climate Analysis Center (CAC) global SST analyses have used information from satellites on SST patterns and gradients (Reynolds 1988). While we are less concerned in this paper about the temperature record over the past century or so, our results do have important implications for this longer time scale, as discussed in section 5.

One motivation for this study stems from some apparently puzzling results of a comparison of a new dataset of global atmospheric temperatures from satellite microwave sounding unit (MSU) measurements (Spencer and Christy 1990; Spencer et al. 1990) with global surface datasets (denoted SFC here) used by the IPCC in their assessments. The surface data are a combination of Jones land data (Jones et al. 1986a,b; Jones 1988) and the United Kingdom Meteorological Office (UKMO) SST data (Bottomley et al. 1990).[1] Correlation coefficients between monthly mean anomalies for the period 1979 to 1990 for 5° latitude by 5° longitude boxes reveal very distinctive spatial patterns with values ranging from less than zero to over 0.9. The physical differences between the two quantities measured in each case is one factor whose importance needs to be assessed (see section 2). Otherwise, the working hypothesis was that the patterns could be explained by properly accounting for the size of the climate signal, as measured by the actual temperature variations during this period, versus the noise in the data. Both the signal and noise characteristics are quite different in the two datasets.

The satellite data have fairly uniform data coverage and good regular sampling, while the expected standard error of the monthly means for SST and land temperatures varies spatially according to the distribution and numbers of observations available. The latter statement assumes that there is a significant random error component to the in situ observations which will decrease for the monthly mean roughly with the square root of the number of observations available (Parker 1984). Accordingly, this has led us to examine in detail the kinds of errors that are expected, their magnitude and distribution, and the numbers of observations available.

In section 2 we describe the satellite dataset. Section 3 presents the results of the comparison between the MSU and surface data. Section 4 delves into the characteristics of the surface dataset in detail and discusses the nature of the systematic and random errors that should be expected. It goes on to examine the Comprehensive Ocean–Atmosphere Data Set (COADS) (Slutz et al. 1985) to see how the expected errors are manifest in practice over the oceans for the year 1979. In this way, an estimate of the noise can be made and, along with measures of the signal, an attempt is made to explain the results of section 3. SSTs from the CAC (Reynolds 1988; see Shea et al. 1990) are compared with those from the UKMO and with MSU anomalies to determine the reproducibility of the anomalies. In several regions, including the North Pacific and North Atlantic, more extensive comparisons are made. The discussion in section 5 considers the implications of these results for the climate record and section 6 provides some recommendations.

---

[1] We began this study using the IPCC (1990) dataset but later learned that Antarctic air temperature anomalies had been erroneously reduced to one-tenth of their true value in that dataset (Folland, personal communication). This has been rectified. However, in the interim, the UKMO SST dataset has been updated with data from the Comprehensive Ocean–Atmosphere Data Set to enhance coverage, and with minor corrections implemented. This revised dataset is used, perhaps with further changes, in the forthcoming IPCC 1992 supplement report.

2. The MSU and SFC datasets

The surface dataset is based upon station temperatures over land surfaces combined with in situ SSTs in the form of anomalies, relative to 1951–80. Near islands and in coastal regions, the merging procedure averaged the land and ocean anomalies over each 5° square according to the proportion of land or ocean in the square. Large differences occur between the surface air temperature and SST in some places and at certain times of the year, notably off the east coast of Asia and North America and off Antarctica in winter, where mean differences exceed 4°C, but it is generally expected that *anomalies* of air and sea surface temperatures will go hand in hand. Because sampling is a major problem over most of the ocean and SSTs have much greater persistence, it has generally been preferred to use SSTs for monitoring surface temperatures. The differences between SST and air temperature anomalies over the North Atlantic and North Pacific are quantified in section 4b.

A second dataset is the channel 2 microwave data at 53.74 GHz from the MSUs that have been on NOAA satellites since late 1978 (Spencer et al. 1990). The MSUs sample globally twice daily from each of two satellites with different equator crossing times, although there are gaps in the coverage. However, even with only one satellite, the ability to obtain many observations each month globally is a major asset. Data were available from both a morning and an afternoon satellite during about half of the 12-year record. The main intervals when only one satellite was operating are February 1980–January 1981 (NOAA 6), April 1983–December 1984 (NOAA 7), and April 1987–September 1988 (NOAA 10). Limited numbers of retrievals in April and August 1981 have led us to exclude them from the comparison. Channel 2 is sensitive to thermal emission from molecular oxygen in the middle troposphere and is relatively insensitive to water vapor and the earth's surface, thereby providing the excellent long-term stability desired for atmospheric monitoring. Moreover, the excellent agreement between two MSUs on different satellites (NOAA-6 and NOAA-7), where the monthly mean hemispheric temperatures are reproduced to within about 0.01°C, confirms the suitability for this channel as a monitoring tool. Locally, monthly mean temperatures are reproduced by the two satellites to within a standard deviation of ~0.1°C, with slightly higher values at high latitudes and over land areas (Spencer and Christy 1992a), which provides a direct measure of noise in the MSU data.

The main loss of data from the MSU arises from the need to filter out the effects of precipitation-sized ice in thunderstorms. In mountain regions the surface emissivity has a noticeable effect but because it is systematic, the interference can be mostly eliminated when the mean annual cycle is removed. Because of "limb darkening," scans well off nadir are excluded. The data are binned into 2.5° grid squares from 88.75°S to 88.75°N. Over the course of a month, individual grid squares in the tropics receive about 40 observations (double with two satellites operating), while regions poleward of 45° have well over 100 observations because of the convergence of orbital paths at high latitudes. Once outliers (such as from precipitation contamination) are removed, the MSU values are averaged in each grid square. The anomalies in the MSU dataset are constructed from a base annual cycle that is itself dependent on the equator crossing time of the satellite. In this way, the systematic diurnal effects are effectively removed. A two-dimensional 3×3 binomial filter is then applied to the anomalies to allow for slight smoothing.

Spencer et al. (1990) presented the temperature weighting function for channel 2 as having a broad peak between 500 and 1000 mb and dropping to half the peak near 300 mb. More recent evidence indicates that the true weighting function peaks near 500 mb at nadir and at even higher altitudes off nadir (Spencer and Christy 1992a). There is a nontrivial influence from the stratosphere, an assessment of which and a means to remove it are discussed in Spencer and Christy (1992b). Essentially, the off-nadir data, which have a somewhat different vertical weighting function, can be used to remove the stratospheric influence and thus provide an adjusted vertical weighting function (called channel 2R). Correlations between the SFC and MSU 2R data are slightly higher than with MSU 2 in both hemispheres but, as the reproducibility of results with MSU 2 is about a factor of 3 better than with MSU 2R, we have preferred to use the original channel 2 data.

The MSU is monitoring a different physical quantity than the SFC, and the vertical structure of the temperature anomalies is one major factor in expecting differences to occur between the climate signals in the two datasets. Because the largest temperature variance is found over land (Gutzler et al. 1988, and see Fig. 6), it is clear that the surface has a profound influence on tropospheric temperatures. The best relationship, however, might be expected to occur between the MSU data and radiosonde information such as the mean temperature of a layer. This comparison has been done by Spencer and Christy (1992a) and results show best agreement for the 1000–200-mb layer, with correlation coefficients typically exceeding 0.95. The overall standard error of the monthly mean differences between the MSU and the radiosonde data for that layer was ~0.2°C. If the actual weighting function for the MSU channel 2 is applied to the radiosonde data, the results are slightly better and the standard error of the difference with individual stations ranges from 0.14° to 0.25°C (Spencer and Christy 1992a). If half the error variance is assigned to the radiosonde data, then the error (or noise) in the MSU data should be a factor of $2^{1/2}$ less. Accordingly, the radiosonde comparison

reinforces the results from the comparison of the MSU data between different satellites that the noise level of the MSU data is assessed as between 0.10° and 0.15°C.

Physically we would anticipate that differences between the MSU and SFC data would most likely be found where there is some degree of decoupling in the vertical between the surface and middle troposphere. The main sign of a disconnection is probably a strong temperature inversion. A pervasive inversion occurs over the tropics as the trade-wind inversion, decoupling the surface boundary layer from the free atmosphere. Spencer and Christy (1992a) find that monthly mean temperatures for the layer from 1000 to 700 mb were correlated with MSU values less than ~0.4 at Hawaii and Guam in the tropical Pacific, indicating that there clearly is a real physical difference in the tropics between MSU and SFC data.

Shallow temperature inversions are also commonly found over land in winter, especially in high latitudes. Over most of the Northern Hemisphere (NH) in winter, monthly circulation anomalies are found to have an equivalent barotropic structure so that the vertical anomalies are highly coherent (Blackmon et al. 1979). The main exceptions are over the interior of continents where mountains appear to play some role in creating a more baroclinic structure. Most of the lower-troposphere (850 mb) wintertime temperature anomalies appear to be associated with changes in large-scale advection in the atmosphere (Gutzler et al. 1988) rather than surface fluxes, although this is less true over Asia, where Gutzler et al. suggest that snow cover may contribute significantly through changes in diabatic heating in the heat balance.

Hurrell and Trenberth (1992) have compared the MSU data with European Centre for Medium-Range Weather Forecasts (ECMWF) temperatures at each level and with the ECMWF data weighted by the MSU weighting function. They further reveal aspects of the vertical structure sensed by the MSU data, and the highest correlations are found at 300 mb. However, results mostly reveal shortcomings in the ECMWF data when used for climate monitoring and the ECMWF 1000-mb temperature fields, in particular, have experienced major problems (e.g., see Table 4 presented later).

3. Comparison of the MSU and SFC data

The MSU data have been compared with the surface temperature dataset over the 12 years 1979 to 1990 (142 monthly values; 2 months missing) after being averaged into 5° boxes to coincide with those in the surface dataset. The mean annual cycle for 1979–90 was first subtracted from each dataset, with 9 out of 12 years required to define the annual cycle. Otherwise, the values for the nonqualifying grid points were not used in this intercomparison. If missing data occurred in one dataset, then it was also set missing in the other to make sure that only the common data were compared. The limiting factor on missing data was mostly the surface temperature dataset. (Because of the different treatment of missing data, the resulting time series are not the same as those prepared by the UKMO.)

The correlation coefficient between the two datasets on a 5° box-by-box basis (Fig. 1) reveals the highest correlations of over 0.85 across North America. Values

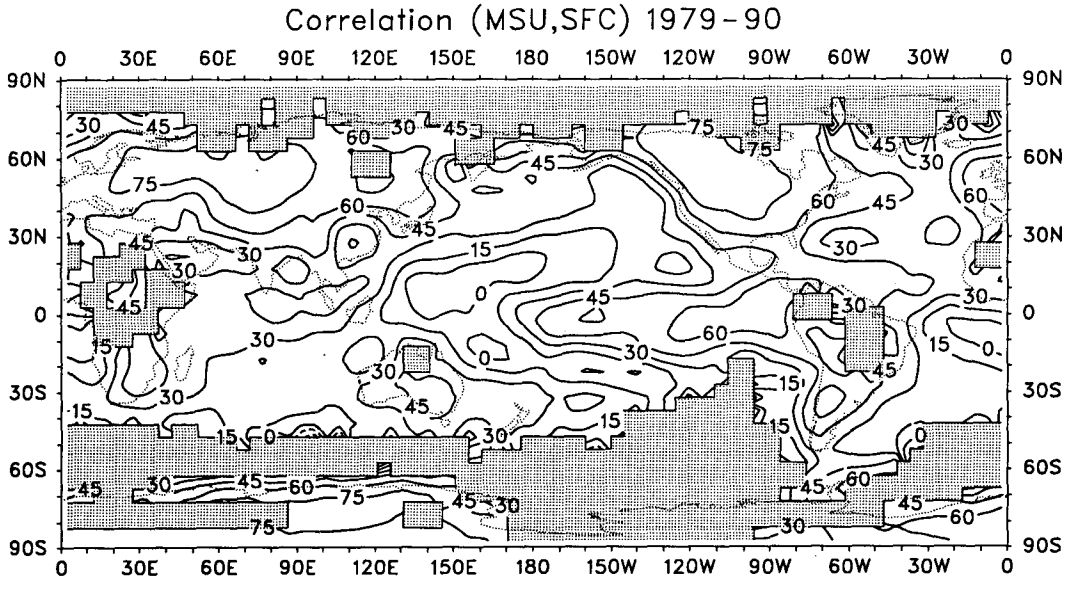


FIG. 1. Correlation coefficient over 144 months from 1979 to 1990 of the MSU and SFC data. Values are set to missing (indicated by stippling), if there were missing data in either set (see text). Values have been multiplied by 100 and the contour interval is 15.

TABLE 1. For the areas as labeled, given are the correlation r between MSU and SFC data for up to 13 months 1979–90, the standard deviation of each series s_{MSU} and s_{SFC} in degrees Celsius, and the percentage of the area with observations.

Area	Latitude	Longitude	r	s_{MSU}	s_{SFC}	Percent area
Globe	90°N–90°S		.62	0.20	0.15	81
Northern Hemisphere	0°–90°N		.58	0.21	0.21	91
Southern Hemisphere	0°–90°S		.48	0.21	0.15	75
North America	30°–70°N	75°–125°W	.88	0.69	1.28	100
Eurasia	40°–70°N	0°–140°E	.81	0.45	1.03	96
Australia	15°–40°S	115°–155°E	.51	0.47	0.43	96
North Atlantic	20°–50°N	20°–60°W	.47	0.36	0.22	100
South Atlantic	0°–55°S	10°E–35°W	.25	0.26	0.23	85
North Pacific	30°–50°N	150°E–130°W	.47	0.44	0.37	100
Indian Ocean	0°–40°S	45°–110°E	.57	0.28	0.22	99
Eastern Tropical Pacific	10°N–20°S	85°–150°W	.82	0.47	0.57	90
Western Tropical Pacific	15°N–15°S	120°W–180°	.15	0.25	0.17	99

exceeding 0.8 are widespread over Asia and Europe but the only oceanic area with such high correlations is in the tropical east Pacific. Correlations are much less ($r \approx 0.5$) over the tropical continents. Generally correlations are much less over the oceans although values ~0.5 occur in parts of the North Atlantic, tropical Atlantic north of the equator, and in the tropical Indian Ocean. Values are distinctly lower (~0.4) in the North Pacific and over the Southern Hemisphere (SH) and are lowest (~0.0) in the western tropical Pacific. Low correlations ($r < 0.2$) are also found in the South Atlantic. The pattern of correlations is quite distinctive and a central purpose of this paper is to provide an explanation for these results.

Given two positively correlated (by r) time series (e.g., MSU and SFC), then after the series are normalized to unit variance, an approximate measure of the signal-to-noise ratio, as given by the variances (see Appendix), is

$$\frac{\text{signal}}{\text{noise}} = \frac{r}{1 - r} \qquad (1)$$

which ranges from 0 for $r = 0$, to 1 for $r = 0.5$, to 9 for $r = 0.9$. This provides an alternative interpretation of Fig. 1, although it should not be taken too literally, since it is limited by the physical differences between the MSU and SFC datasets.

It is often argued for climate purposes that temperature anomalies are large in scale so that averaging over larger areas better serves to define the anomalies while reducing sampling noise. In the following, area averages over a number of broad regions have been taken to emphasize the regional variations and to see the extent to which the correlations improve. In fact, experience shows that one must be cautious in adopting this approach. For example, when the variance changes across an area, the time series of the area average is dominated by the sector of largest variance, which is often not the same in the two datasets, so that area averaging actually reduces the correlation coefficient.

Some 2-panel figures are presented that show (i) the two time series normalized by their standard deviations, and (ii) the difference between the normalized MSU–SFC values. The fraction of each subarea included in the calculation is given in Table 1, along with the correlation and standard deviations of each series. The standard deviation is a measure of the signal plus noise in each dataset.

Figure 2 shows the global values and Figs. 3 and 4 values for the NH and SH. Globally, coverage is about 81% with the correlation coefficient of 0.62. Largest differences occur in individual months rather than over extended periods, and it is immediately apparent that most of the large differences arise from month-to-month variability in the SFC dataset, indicating less continuity in time. Apparently, some of this variability arises from changes in data coverage (C. Folland, personal communication). There is a slight upward trend of 0.14°C per decade for the SFC data relative to the MSU data.

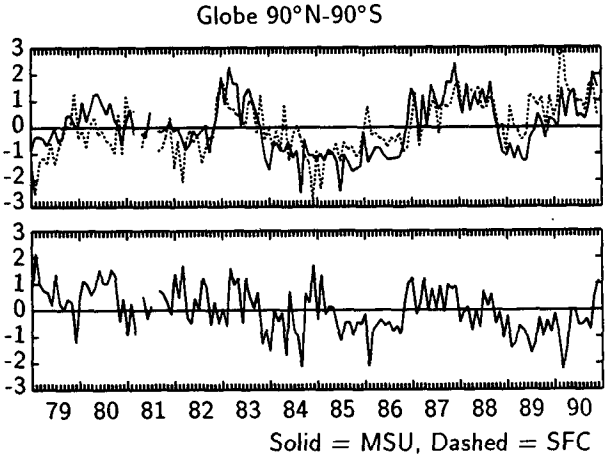


FIG. 2. Time series of monthly values of temperature from MSU (solid) and SFC (dashed) for the entire globe. Both time series are normalized by the standard deviation given in Table 1 and their difference is plotted in the lower panel. The vertical scale is in units of standard deviation and ranges over ±3 standard deviations.

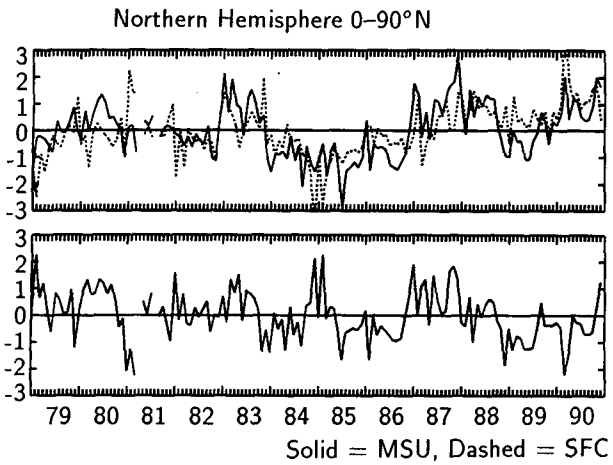


FIG. 3. As in Fig. 2 but for the NH.

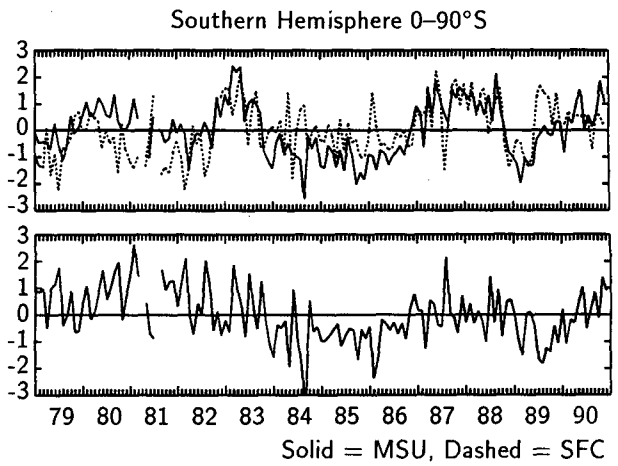


FIG. 4. As in Fig. 2 but for the SH.

Coverage for the NH is ~91% versus ~75% for the SH,[2] and the correlations are lower than the global in both cases, 0.58 for the NH and 0.48 for the SH. Again, several spiky features in the SFC dataset in the NH also show up in the difference field, notably January, February, and March 1981; January 1982; November 1983; December 1984; February 1985; and March 1990, all winter months. These may be associated with the common wintertime surface temperature inversions over land. There are also some features that arise from the MSU data (July 1985 and December 1987 in the NH and September 1984 and February 1987 in the SH).

Time series for a few other regions are presented in Fig. 5. A summary of the correlations, standard deviations and coverage for several areas is given in Table 1. By far the largest correlation (0.88) occurs over North America, which also has the largest standard deviation in both time series. Biggest discrepancies are consistently found in winter months (Fig. 5). In addition, the SFC data contain a linear trend of 0.2°C per decade relative to the MSU data for North America. Over Eurasia (not shown) the fairly high correlation is lowered by differences during the winters of 1981/82, 1982/83, and 1983/84. Over Australia (not shown), the overall correlation is only 0.52 although the agreement is better after mid 1983. Over Antarctica, south of 65°S (not shown), correlations are surprisingly high at 0.75, with standard deviations of 0.79° and 1.23°C for the MSU and SFC, respectively; however, coverage is mostly a meager 53%. [Antarctic coverage drops to zero in 1990 in our dataset, but this is shortly to be remedied (Folland, personal communication).]

Over the oceans, for the North Atlantic (Fig. 5), there is a multiyear signal, although modest in size, that is captured by both series with temperatures below the mean from 1984 to 1986 and above in 1987 and 1988. The agreement deteriorates in 1989, however, lowering the correlation by 10%. In contrast in the South Atlantic (not shown), agreement is reasonable only in 1987 and there is a major difference in the decadal time-scale fluctuations. The correlation in the northern extratropical Pacific (0.47), where data coverage is good, is enhanced by the warmth of 1989 and 1990 relative to other years. By far the best ocean area correlation occurs in the eastern tropical Pacific (Fig. 5) and is associated with the very large climate signal

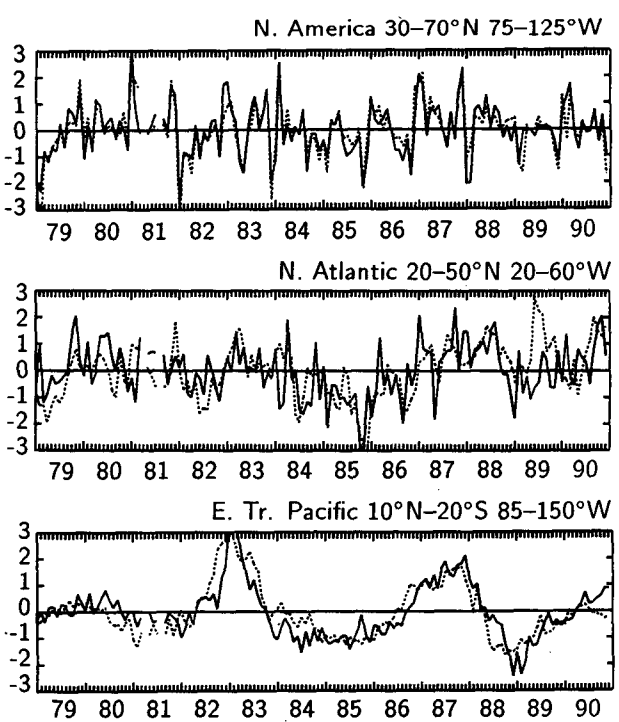


FIG. 5. Time series of monthly values of temperature from MSU (solid) and SFC (dashed) for several areas of the globe. Both time series are normalized by the standard deviation given in Table 1.

[2] For the IPCC (1990) dataset SH coverage was 68%.

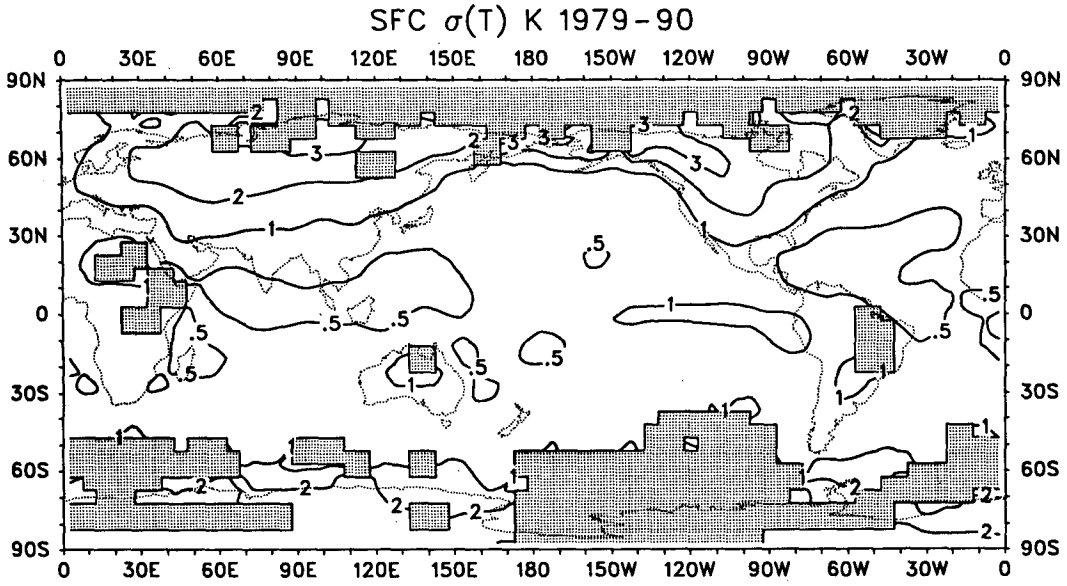


FIG. 6. Standard deviation over 144 months from 1979 to 1990 of SFC temperature anomalies. Contours are 0.5, 1, 2, and 3 K. Areas of missing data are indicated by stippling.

of the 1982–83 and 1986–87 El Niño events, and the cold 1988 La Niña. These events are also present in the Indian Ocean series (not shown), although with smaller amplitude. In the western tropical Pacific (not shown), however, warmings in 1982–83 and 1987 in the MSU data are not seen in the SFC dataset (see further discussion in section 4d).

Clearly, one reason for the differences in correlation relates to the size and persistence of the signal, as can be seen from the standard deviations in Table 1. In fact, this quantity squared depicts the sum of the actual signal plus the noise variance. A more complete view of the signal can be seen from the map of standard deviations of monthly mean anomalies from the full SFC dataset (Fig. 6). Over continents the signal is large and hard to miss and the large signal of El Niño is also easily captured in both datasets in its gross character, although details differ. The lowest correlations in Table 1 coincide with the smallest standard deviations, implying that the signal is small and the main variance may be noise in at least one dataset.

4. Characteristics of the surface dataset

The lower correlations over the oceans (Fig. 1) are thought to arise in part from problems in the SFC dataset because the coverage and sampling are fairly uniform in the MSU data. The signal is smaller over the oceans, but we need to quantify the level of noise in the data. We therefore examine expected and actual error sources in the SSTs in detail. In this section, the focus is on in situ observations over the past decade or so.

a. Noise and errors in monthly mean SSTs

The sources of errors in estimating monthly mean SSTs from ship data can be divided up into (i) errors in making individual observations, (ii) incomplete sampling of the diurnal cycle, (iii) incomplete sampling of within-month variance other than the diurnal and seasonal cycles, (iv) within-month mean variance due to the seasonal cycle, and (v) incomplete sampling of the spatial gradients within a grid square.

(i) Individual observations are not particularly accurate. Saur (1963) compared routine injection with high-quality bucket temperatures taken from the same ship for several different ships. The quality of the injection temperatures varied considerably; there was a warm bias overall of 0.7°C and a standard deviation between the two measurements of 0.9°C. For different ships, the bias ranged from −0.3° to 1.7°C. Saur noted that the seawater intake varied from 3 to 7 m below the surface and the pipes ranged from 10 to 51 cm in diameter. These factors and the thermometer calibration and exposure in the engine room account for part of the differences and biases. More generally, the bias of intake relative to bucket measurements appears to be 0.3°C (James and Fox 1972; Bottomley et al. 1990). Even today there is a mix of bucket and intake SSTs in the observations. Tabata (1978) has compared observations from merchant ships with those from ships or buoys on station in the northeast Pacific. Overall, the standard deviation of the differences was 1.5°C. Bernstein and Chelton (1985), in their Fig. 1, compare ship–ship observations made within 100 km and 6 h,

with the standard deviation of differences being 1.5°C, so that, assuming the error comes equally from each ship, there is a 1.1°C error on average from each ship. Based on these studies, it seems reasonable to conclude that we should expect a standard error for individual observations of ~1°C.

(ii) The diurnal cycle is not insignificant everywhere. Cornillon and Stramma (1985) show that at times when winds are light, diurnal heating can exceed 2°C, but mostly it is less than 0.5°C. North of Australia, Morrissey (1990, see his Fig. 10) finds a mean amplitude of ~0.15°C. The diurnal cycle is very small, but noticeable, at OWS (Ocean Weather Station) P (Fissel et al. 1976) in the northeast Pacific—maximum in spring with an amplitude of 0.095°C. Amplitudes of 0.3° and 0.2°C have been reported at other midlatitude areas by Ostapoff and Worthem (1974) and Price et al. (1986).

(iii) The within-month variance sometimes includes the diurnal component. Fissel et al. (1976) show that the annual and semiannual component is dominant at OWS P and the mean within-season variance (excluding the seasonal cycle) is only 0.18 $(°C)^2$. From the spectra they present it appears that about 40% of the within-season variance is at periods of less than two months, that is, the within-month variance is ~0.08 $(°C)^2$. A more complete analysis at OWSs A, B, C, D, E, I, J, K, and M, all in the North Atlantic, can be gleaned from Kraus and Morrison (1966). From their Table 4 and tables in the Appendix to the paper, the within-month variance in degrees Celsius squared is 0.045, 0.060, 0.044, 0.113, 0.054, 0.025, 0.022, 0.052, and 0.039, respectively, that is, ranging from 0.02 to 0.11 $(°C)^2$. The distribution of high-frequency variance probably varies considerably spatially, however, and appears to be much greater in regions of strong baroclinicity such as in western boundary currents and over the southern oceans (Fu et al. 1988).

(iv) The annual cycle has typical amplitudes for SST of 3°C in the SH and 5°C in NH but only 1°C in the tropics (Shea et al. 1990). In some months, the seasonal trend within a month can be large—for example, see July–June (Fig. 7), where it is over 3°C in the North Pacific. A fairly conservative analysis, assuming only the 12-month (first) harmonic of amplitude A, reveals that the average contribution to within-month variance is $0.0112A^2$. This assumes a linear trend in each month for 1/12 of a cycle. If the sampling is uniform throughout the month, this component does not contribute. But if sampling is random, its variance is reduced roughly as $1/N$ (see Kidson and Trenberth 1988), where N is the number of observations.

(v) The problem of SST gradients across a grid box is mainly of importance in regions of strong gradients such as in the boundary currents off the east coast of Asia and in the Gulf Stream off the east coast of North America. For instance, at 60°W from 42° to 44°N the gradient in winter is 9°C over 2° latitude. Any minor change in ship tracks across that region can lead to substantial apparent anomalies in SST locally, although for large area averages the effect should be fairly random resulting in cancellation of errors. For a uniform gradient of $2a$°C across 2° latitude the expected variance within a 2° box would be $\frac{1}{3}a^2$ $(°C)^2$, if well sampled.

These problems are all present in COADS. However, for the UKMO, Bottomley et al. (1990) attempted to

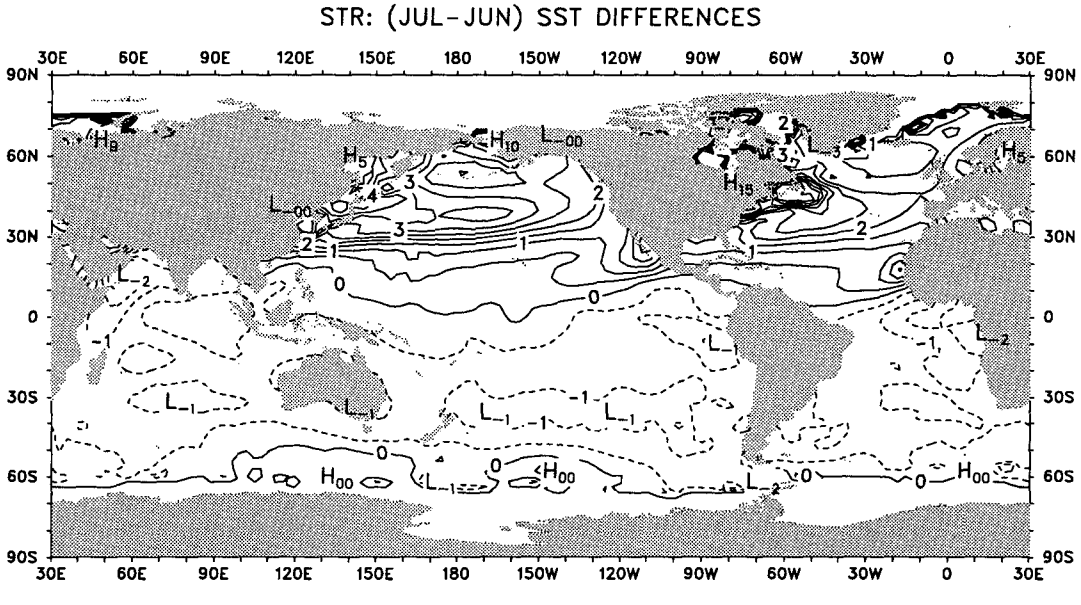


FIG. 7. Differences in mean SST July minus June from the Shea et al. (1990) climatology. Contours are every 0.5°C.

minimize the effects of some of these potential biases by first generating a climatological background field smoothed in space and time. The individual monthly means were then created by analyzing anomalies that were subsequently added to the background field. To reduce the effects of spatial gradients within 5° squares and temporal variations from the annual cycle, a working grid of 1° boxes was used with a temporal resolution of 5 days (pentads). Further, a process called "winsorization" was applied over the 5° boxes for each month in which all anomalies in the first or fourth quarter (1° box by box and pentad by pentad) were replaced with second and third quarter values, respectively, to remove the effects of outliers. The 5° averaged windsorized anomalies were added to the monthly background field to give the 5° monthly means. The success of these procedures is clearly a function of the number of observations available.

All of the preceding five items are sampling issues. As a guide, if we estimate the sampling variance from the first four items above as 1.0, 0.01, 0.1, and 0.28 $(°C)^2$ (corresponding to $A = 5°C$), respectively, the total error variance would be $1.39°C^2$. Expressed as an anticipated standard error for a monthly mean, this would be 1.2°C for a single observation. However, to a first approximation, the error in the mean has a variance that decreases by $1/N$, where N is the number of observations, so that the standard error of the mean decreases by a factor of $N^{1/2}$.

To get a better estimate of how the various sources of error are manifest in practice, the COADS have been used to estimate the total noise from all sources. To do this, the standard deviation of the SST observations within each box for each month of 1979 has been examined. The grid squares are all 2° × 2° (not 5° × 5°). This can be combined with the values of N to give a total estimate of noise variance that includes all the factors above.

Rather than use the actual standard deviation of observations in each box, which may be contaminated by outliers, all observations from COADS were sorted within each box and month into sextiles with a slight modification to the first and fifth sextiles such that, for a Gaussian distribution, the difference between the first and fifth sextile values is equal to 2 times the standard deviation. Accordingly, a robust estimate of the standard deviation can be obtained (called e in COADS documentation).

In an exploratory analysis of e, we found that the values were very much as expected from the considerations noted previously, provided there were a reasonable number of observations present and the box was not located in a region of strong SST gradient. In the vicinity of the Gulf Stream, for instance, e was consistently inflated relative to other areas, with values in excess of 3°C. It was clear that spatial gradients of SST within a box were insufficient to explain the observed increase in variance in these regions, which implies that the temporal variability is also greatly enhanced in such regions and contributes significantly to the within-box/month variance.

To summarize the results from all areas and focus on the values of interest, we used the exploratory analysis to define criteria for sampling boxes. Boxes without strong SST gradients were chosen by excluding all boxes where the mean SST difference from any surrounding 2° box was greater than 3°C, so that the contribution from spatial variance within each box is less than 0.75 $(°C)^2$. All boxes with less than 16 observations were excluded. We then averaged over all qualifying boxes in large areas by using

$$\langle e \rangle = \left[\frac{\sum Ne^2}{\sum N}\right]^{1/2}$$

where the areas chosen are the North Pacific and North Atlantic north of 20°N, South Pacific and South Atlantic south of 20°S, tropical Pacific and tropical Atlantic 20°N to 20°S, and the Indian Ocean, which is dominated by observations in the tropics. Table 2 shows the results for each ocean basin for each month of 1979 and the annual mean, and also gives the mean number of observations used in each month. Largest standard deviations are found in the North Pacific where the annual mean $\langle e \rangle = 1.42°C$, while the value in the North Atlantic is $\langle e \rangle = 1.15°C$. In fact, values are consistently larger in the Pacific for all regions. In the tropics, $\langle e \rangle$ averages close to 1°C. At higher latitudes, modest SST gradients of 2°C across a box could increase that to 1.15°C, so it appears that there is also a small contribution in the extratropics from temporal variability within each month. Other studies by Chelton (see Bernstein and Chelton 1985) have also found larger ship-to-ship differences in the North Pacific than in the North Atlantic and it appears that this is probably real and arises from a greater diversity of ships from many countries with different calibration standards and quality control checks. The slightly lower density of observations may also be a factor.

A measure of noise from individual observations has been shown in Table 2. To estimate the total noise, we also need to know the total numbers of observations. We have therefore used the COADS to provide estimates of N over 5° by 5° boxes, the same size as used by Bottomley et al. (1990). Rather than show N, which for an individual month ranges from 0 to over 1000 observations, we present $N^{1/2}$ for January and July 1979 (Fig. 8). Areas in these plots where $N^{1/2}$ is less than about 3 correspond to the regions in Fig. 1 where there are often missing data and insufficient numbers to calculate correlations. In the North Pacific, $N^{1/2}$ is typically 11 to 13, compared with 15 to 19 in the North Atlantic. The exception is near the coasts of all the northern ocean basins, where $N^{1/2}$ is over 20. On the other hand, $N^{1/2}$ drops to less than 5 over the Pacific south of 10°N and south of the equator in the Atlantic.

TABLE 2. Robust estimates of the within 2° latitude × 2° longitude boxes and within-month SST standard deviations in degrees Celsius for the areas given and average number of observations per month used to make the estimate for the year 1979. Boxes were excluded if they contained fewer than 16 observations per month and if the gradient in SST between any adjacent box was greater than 3°C. Also given is the number of 2° boxes used and the average number of observations per box for those that qualified.

	Area						
Month	North Atlantic	Tropical Atlantic	South Atlantic	North Pacific	Tropical Pacific	South Pacific	Indian Ocean
1	1.14	0.95	0.97	1.34	1.08	1.26	0.81
2	1.10	0.91	0.85	1.31	0.97	1.34	0.82
3	1.03	1.04	0.98	1.28	1.06	1.28	0.91
4	1.05	0.98	1.01	1.34	1.10	1.29	0.93
5	1.27	0.90	0.91	1.45	0.98	1.23	0.79
6	1.29	0.96	1.26	1.59	0.94	1.10	1.03
7	1.21	0.95	1.17	1.51	1.05	1.17	1.03
8	1.03	0.94	1.00	1.35	1.05	0.95	1.05
9	1.05	0.97	1.13	1.34	1.02	1.05	0.96
10	1.23	0.94	0.98	1.56	1.15	1.29	0.96
11	1.13	0.86	1.04	1.48	1.06	1.32	0.89
12	1.19	0.91	1.08	1.41	1.12	1.36	0.77
Year	1.15	0.95	1.03	1.42	1.05	1.23	0.92
Obs/Mo	19 180	5513	460	19 384	2428	683	3255
Number of boxes	353	124	13	435	67	25	73
Obs/box	54	44	35	45	36	27	45

The only area well observed in the western tropical Pacific is the ship track near 150°E, which has been explored in detail by Morrissey (1990).

Combining values from Fig. 8 with Table 2 for areas without strong SST gradients, an overall estimate of the noise in monthly mean SSTs over 5° boxes can be made. Lowest values of ~0.06°C are found in the North Atlantic. However, because of the strong SST gradients near the Gulf Stream, values will be a factor of two or more larger. Noise estimates are ~0.1°C in the North Pacific and tropical Indian and Atlantic oceans, but lie between 0.2° and 0.3°C for the tropical and South Pacific and South Atlantic north of about 30°S. Farther south, except in ship tracks, the noise generally exceeds 0.5°C. Morrissey (1990) estimates for the tropical western Pacific that the standard error of the monthly mean drops to ~0.2°C for three or more observations per day whereas most of the area has fewer than ten observations per month.

Of course, averaging over seasons can reduce the noise further by roughly $\sqrt{3}$, [e.g., see the autocorrelations in Table 7 in Kraus and Morrison (1966) or in Bottomley et al. (1990)] but there is more signal occurring within the averaging period. Still, it is clear from many areas where the SST spectrum is very red (e.g., tropical eastern Pacific) that further time averaging can be helpful provided the interest is in the longer time scales.

It is apparent that part of the pattern in Fig. 1 is largely explained by the preceding arguments. In particular, the signal is much larger than the noise over land. Over the oceans, the signal-to-noise ratio is best over the eastern tropical Pacific and the North Atlantic.

b. *Physical differences between air and sea surface temperatures*

The other potentially important factor that must be assessed is the extent to which the surface air temperature anomalies over the oceans are represented by SST anomalies. It appears that there is a difference between the SST and the surface air temperature anomalies on monthly time scales and, as noted in section 2, there is a further physical difference between the surface air temperature and the lower-tropospheric temperatures sampled by the MSU. Because marine air temperatures have a larger diurnal cycle than SSTs, more observations should be needed to resolve the climate signal. Also, SSTs do not respond strongly to individual cold-air outbreaks or to low-frequency synoptic disturbances, such as atmospheric blocking, whereas substantial anomalies in monthly air temperature means are associated with such events. Large et al. (1986) discuss the SST response to individual storms during the Storm Transfer and Response Experiment (STREX), for instance.

For comparison, results from the same analysis of the COADS in Table 2, but with the surface marine air temperature instead of the SST, are given in Table 3. The same SST gradient criterion for acceptance of a box was used but now at least 16 air temperature observations are required. Generally, there are more air temperature than SST observations so more boxes qualify. However, as suspected, the within-month variance for air temperatures is substantially larger, indicating that there is a need for more observations to reduce the standard error of the monthly mean to ac-

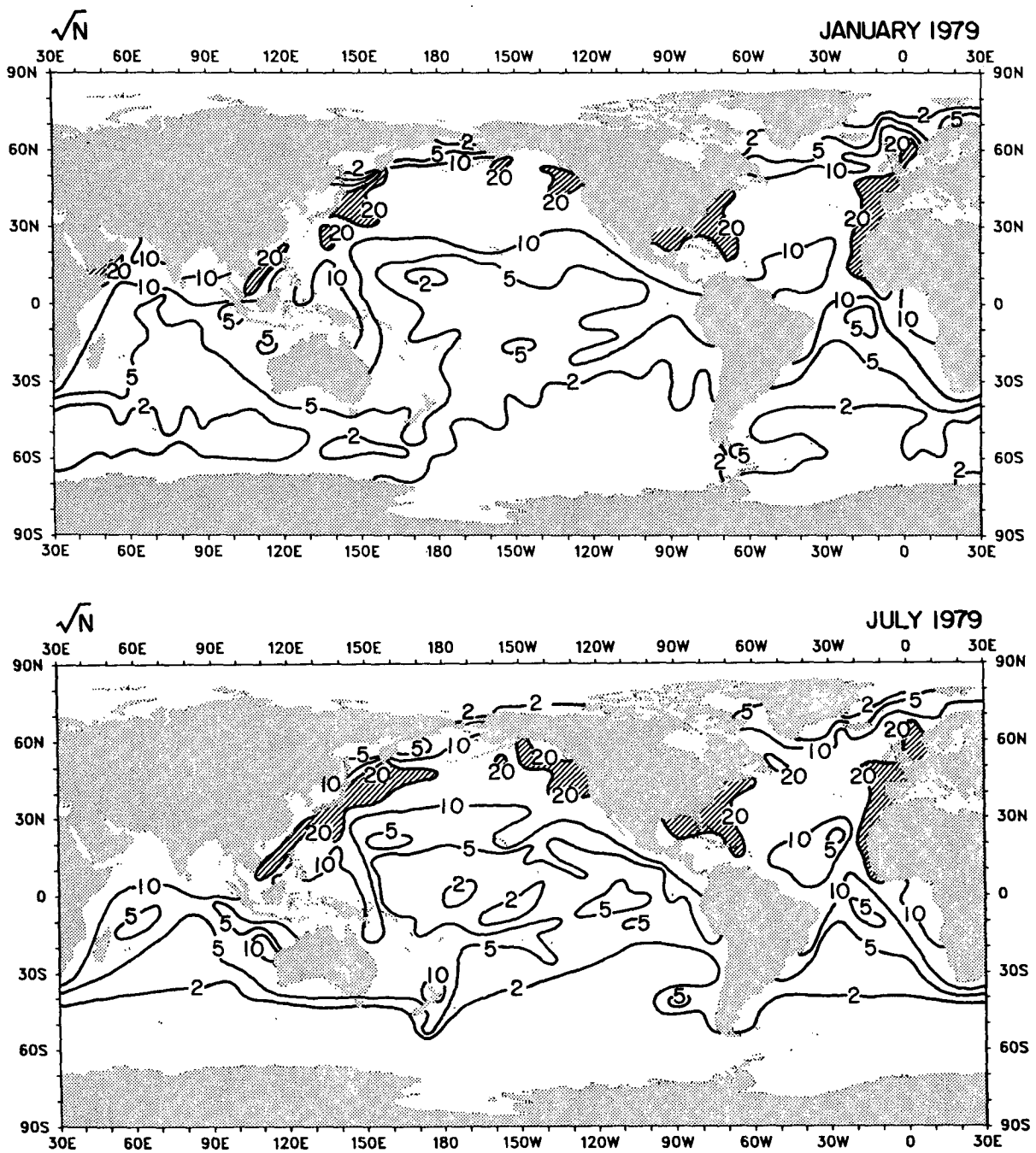


FIG. 8. Maps of the square root of the number of SST observations in each 5° by 5° box from COADS for the months of January and July 1979.

ceptable levels. The standard deviation of the within-month air temperature is 20% to 30% larger in the tropics and SH and 43% to 52% larger in the NH. There is also more seasonal variability (not shown) to the sampling problem of air temperature than for SST, with standard deviations in the North Atlantic averaging 1.46°C for July, August, and September versus 1.84°C for the rest of the year. In the North Atlantic on average, for example, a factor of 2.3 more observations of air temperature would be required to produce the same standard error of the monthly mean as for SST, and this factor rises to over 3 in February.

Bottomley et al. (1990) show that there does seem to be a good correspondence between anomalies of

TABLE 3. Robust estimates of the within 2° latitude × 2° longitude boxes and within-month marine air temperature standard deviations in degrees Celsius for the areas given and average number of observations per month used to make the estimate, for the year 1979. Boxes were excluded if they contained fewer than 16 observations per month and if the gradient in SST between any adjacent box was greater than 3°C. Also given is the number of 2° boxes used and the average number of observations per box for those that qualified.

	Area						
	North Atlantic	Tropical Atlantic	South Atlantic	North Pacific	Tropical Pacific	South Pacific	Indian Ocean
Year	1.75	1.16	1.27	2.03	1.36	1.58	1.17
Obs/Mo	21 284	6289	582	21 153	2782	1168	3682
No. boxes	369	136	16	461	74	40	83
Obs/box	58	46	36	46	38	29	44

marine air temperatures and SSTs when averaged over large areas on seasonal time scales. We have compared monthly air temperature with SST anomalies from COADS in two regions, the North Atlantic and the North Pacific, where sampling is not a problem. Included in the comparison are two other SST products to further assess how reproducible SSTs are, as well as the 1000-mb temperature from ECMWF. The other SST products are the SSTs from the UKMO (labeled UK) that were used in Fig. 1, and SST analyzed values from the Climate Analysis Center (labeled CAC). The area means for the COADS data are computed from the mean of anomalies from each 2° box with 3 or more observations, and because there are areas where the data are inadequate, the means should differ slightly from the results of other products that extend their analyses into regions of poor data. The correlations are restricted to 84 months from 1982 to 1988 because the 1979 through 1981 data from ECMWF clearly contained major errors of several degrees celsius (Hurrell and Trenberth 1992).

Correlation coefficients between all of the parameters for the North Atlantic and North Pacific are shown in Table 4. All of the SST products are very highly correlated ($r > 0.94$), showing that the SST analyses are very robust and should be considered well defined for these areas. Correlations of SST with surface air temperature anomalies are $r \approx 0.89$, showing that there is a distinct difference between SST and surface air temperature but 80% of the variance of one is captured by the other. Correlations of all the SST products and surface air temperature with ECMWF 1000-mb temperatures are modest in the North Pacific but nonexistent in the North Atlantic (see Hurrell and Trenberth 1992). Correlations with the MSU data differ slightly from those in Table 1, mainly because 1989 and 1990 data are not included (a small effect arises also from the different treatment of missing data). In particular, the North Atlantic correlation between SST and MSU is lower in Table 1 because of poor coherence during 1989; see Fig. 5. In the North Pacific, the MSU correlation with surface marine air temperature is higher (0.4) than with any SST product (~0.2).

c. *Comparison of UKMO and CAC SSTs*

As a further test of the validity of the preceding inferences concerning the noise in SST data, the reproducibility of SST anomalies between the CAC and UKMO SST analyzed values has been checked. Because there are many missing data in the UKMO set, both datasets were screened to be consistent in the way the mean annual cycle was removed and the anomalies calculated. The CAC analyses are available from 1982 and, because they use satellite data, contain no missing data. The CAC data are therefore included only if the UKMO data were not missing. Separate annual cycles for the same period were removed using the monthly means, thereby removing possible systematic biases. The CAC anomalies, originally on a 2° grid, were av-

TABLE 4. Correlation coefficients among a number of different estimates of temperature for the North Pacific (30°–50°N, 150°E–130°W) above the diagonal and for the North Atlantic (20°–60°N, 60°–20°W) below the diagonal. The temperature anomalies are 1) mean surface air temperature from COADS = T_a; 2) mean SST from COADS = T_s; 3) mean SST from the U.K. Meteorological Office = T_{UK}; 4) mean SST from the Climate Analysis Center = T_{CAC}; 5) mean 1000-mb air temperature from ECMWF analyses = T_{1000}; 6) mean channel 2 MSU temperature = T_{MSU}.

	T_a	T_s	T_{UK}	T_{CAC}	T_{1000}	T_{MSU}
T_a	1.0	0.89	0.89	0.89	0.59	0.40
T_s	0.89	1.0	0.96	0.98	0.61	0.19
T_{UK}	0.84	0.94	1.0	0.96	0.59	0.24
T_{CAC}	0.82	0.96	0.95	1.0	0.58	0.22
T_{1000}	−0.0	−0.11	−0.13	−0.15	1.0	0.01
T_{MSU}	0.58	0.57	0.60	0.58	−0.11	1.0

eraged onto the same 5° grid of the UKMO. Area averaging, rather than interpolating, is compatible with the way the UKMO data were computed (Bottomley et al. 1990). The comparison covered 108 months from 1982 to 1990.

The results (Fig. 9) reveal correlations over 0.9 for most of the northern oceans and in the eastern tropical Pacific. Values drop to less than 0.75 south of about 10°N elsewhere and to much less than 0.5 in the central tropical Pacific and over the southern oceans, all regions where the numbers of months of data drop off. Correlations are also low and rms differences high near the coastal regions, evidently because of the land anomaly influence in the SFC dataset. Elsewhere, the rms differences between the analyses (Fig. 9) increase from less than 0.2°C over the central North Atlantic to over 0.6°C in the central tropical Pacific, in the eastern Pacific south of 10°S, and generally south of 35°S except near New Zealand: all areas where the correlation drops to less than[3] ~0.6. While there are also missing months in these regions (Fig. 9), SST values were assigned by both centers. In fact, because essentially the same in situ observations are used by both centers, the correlations give an optimistic view of the knowledge of the true anomalies.

The results are remarkably compatible with the conclusions drawn in section 4a. Equation (1) can be readily applied to the two SST records as they are both ostensibly measuring the identical quantity. For the areas where $r = 0.6$, for example, the standard deviation is ~0.7 K so the total variance is ~0.5 K^2, of which the signal is $r \times 0.5 = 0.3\ K^2$, the noise is $(1 - r) \times 0.5 = 0.2\ K^2$, and the latter is compatible with a noise computed from $N \approx 25$ observations and an rms difference of $(2(1 - r) \times 0.5)^{1/2} \approx 0.6$ K, as found in Fig. 9.

d. *Comparison of MSU with CAC SSTs*

Although the results of this comparison are implicit in Figs. 1 and 9, to complete the triad of intercomparisons, Fig. 10 presents the correlation between the CAC SSTs and the MSU data over the oceans. The former has an advantage over the UKMO SSTs because of the use of satellite data and sea-ice information to enhance the coverage, especially over the southern oceans. The resulting pattern bears a strong resemblance to that in Fig. 1 although the extremes are all larger, as should be expected for fewer degrees of freedom (108 months in Fig. 10 versus 142 in Fig. 1). Correlations in Fig. 10 in the immediate vicinity of the major landmasses are lower, apparently reflecting the dominant influence of the land temperature variations and the land dataset included in Fig. 1.

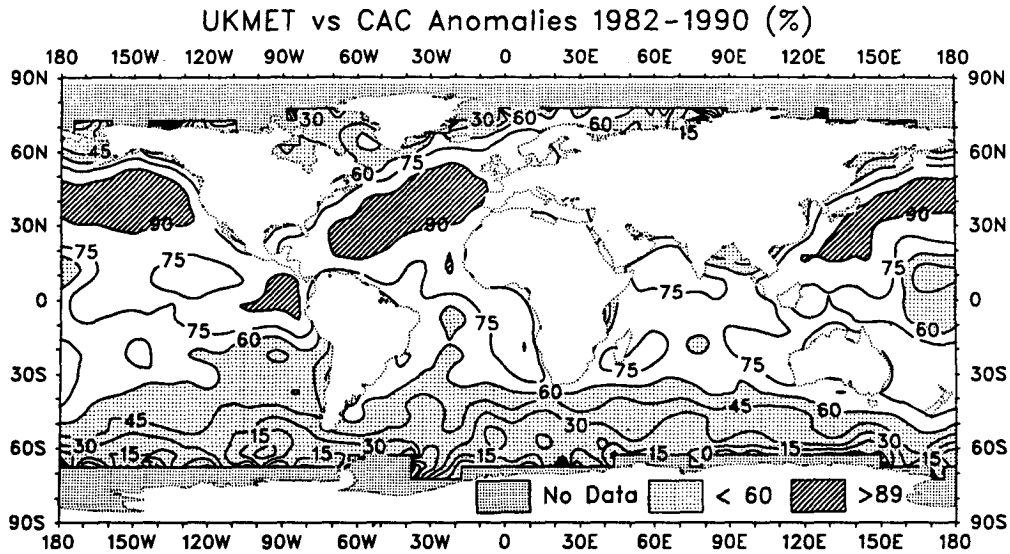
The general agreement between Figs. 1 and 10 suggests that part of the patterns is a result of the real physical differences between the MSU tropospheric temperatures and temperatures at the surface. A particular case in point is the region of negative correlations in the tropical western Pacific, which, as noted earlier, arose from an ENSO signal in the MSU data but little such signal in the SFC record. During ENSO events the entire tropical troposphere warms up a few months after the event (Newell and Weare 1976; Horel and Wallace 1981), following the SST signal in the central and eastern tropical Pacific. However, SSTs in the western tropical Pacific may actually have the reverse sign (e.g., see the composite SST anomaly for the mature phase of an El Niño in Rasmusson and Carpenter 1982), resulting in the correlation pattern seen in Fig. 10. In the tropics, therefore, the common pattern to Figs. 1 and 10 actually reveals the SST anomaly pattern associated with ENSO events, especially the large 1982–83 event.

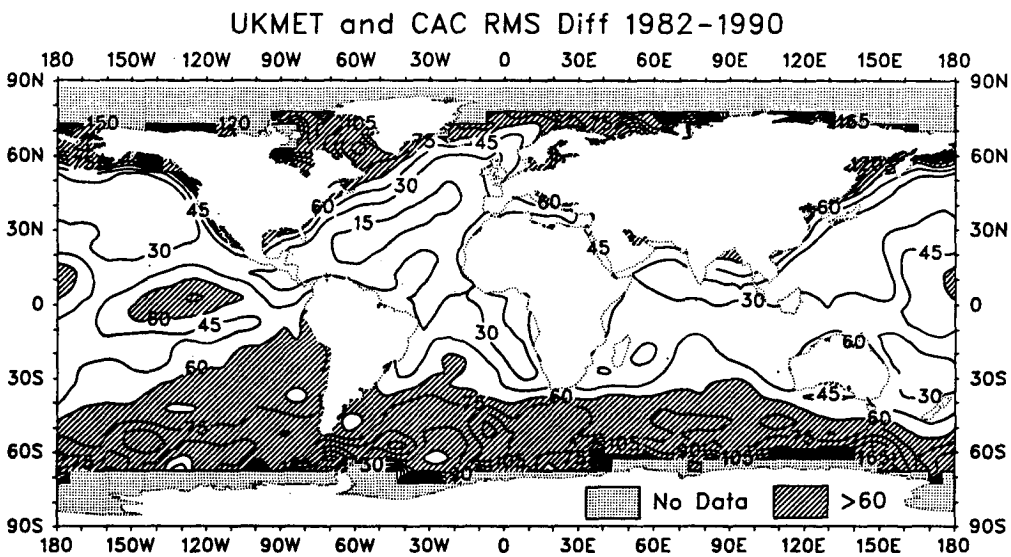
5. Implications for the climate record

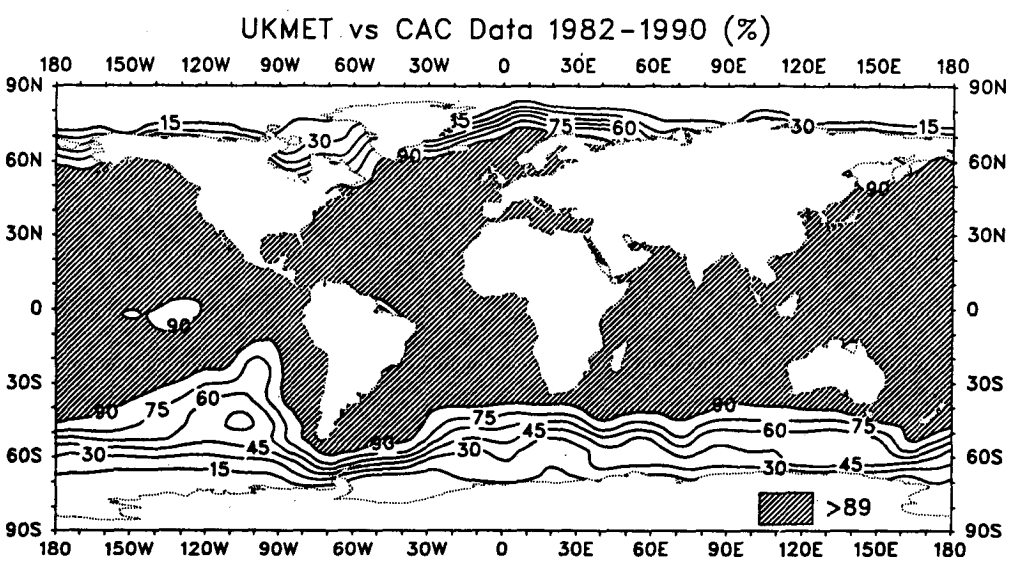
The results of previous sections have indicated that the inherent noise level in an SST observation is ~1.0°C and this is compounded when the observation is made in regions of large temperature gradient. If randomly distributed, N observations can reduce this noise level by about $N^{1/2}$. As seen from Fig. 7, spatial coverage is an issue for SSTs. Even today, there is no global in situ coverage. Changes in the spatial distribution and areal coverage of observations with time can therefore bias the best estimates of "global" trends. In the historical record, the number of observations typically decreases back in time, thereby increasing the noise level. In this section, we therefore consider the implications of our results concerning the noise level of the SST observations combined with the changes in spatial coverage with time for the climate record. A brief summary is given of other measurement problems in the past but it is beyond the scope of this paper to address those here.

Most of the marine observations have been taken by ships of opportunity and over three-fourths have been taken since World War II. The ships of opportunity follow preferred routes and consequently leave vast areas of the oceans inadequately sampled (Fig. 11). Over parts of the southern oceans, coverage was actually better last century, prior to the opening of the Suez and Panama canals. Vast areas of the southern oceans, especially south of about 40°S, are inadequately observed by conventional means to provide reliable analyses. Figure 11 shows coverage for the month of January, while from March to November there are

[3] It is worth noting that the added coverage in the updated UKMO dataset relative to the IPCC (1990) version has actually worsened the agreement with the CAC analyses over the southern oceans.


UKMET vs CAC Anomalies 1982-1990 (%)
No Data
< 60
>89


UKMET and CAC RMS Diff 1982-1990
No Data
>60


UKMET vs CAC Data 1982-1990 (%)
>89

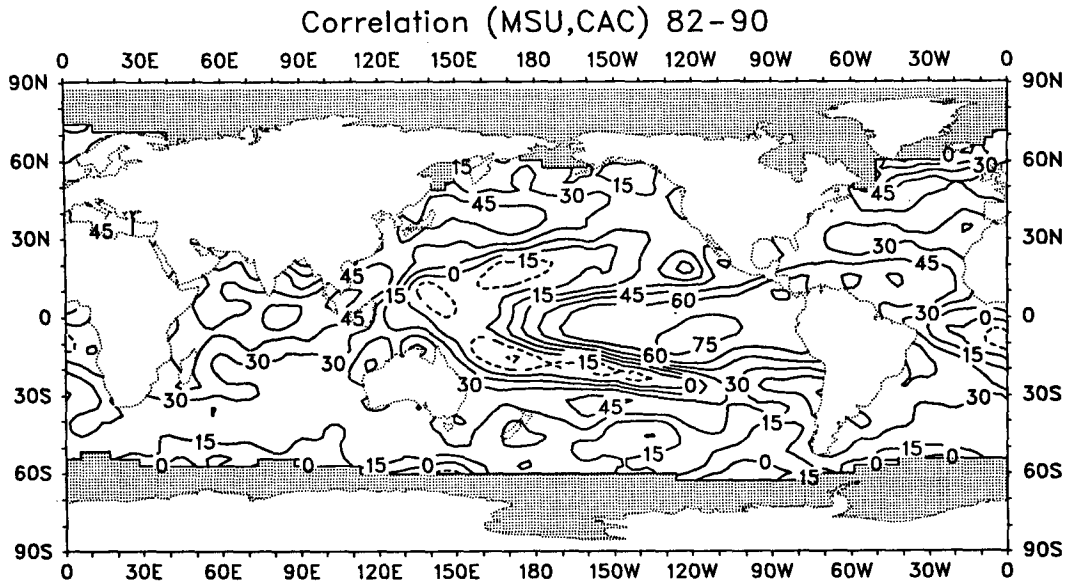


FIG. 10. Correlation coefficient (×100) between SST anomalies from the CAC analyses and the MSU for 1982 through 1990 (108 months). Negative values are dashed and stippled areas indicate less than 70% of months available (i.e., areas of sea ice).

typically almost no observations over the southern oceans (Fig. 8). The panel in Fig. 11 for 1940 is representative of the World War II years, where observations are not available over many areas such as the North Atlantic; 1930 is more typical of the coverage between the two world wars. Since the 1970s, satellite estimates of SSTs have become available globally, but their absolute accuracy is uncertain; substantial corrections are found to be necessary where buoy observations are available to provide calibration. They do, however, provide useful information about patterns and gradients of SSTs (Reynolds 1988).

A certain number of observations are needed to carry out gross quality control of SSTs and the discussion in section 4 assumed that the SSTs were strongly quality controlled. In fact there are many SSTs reported with values outside physically possible limits and location errors are also common. For instance, a substantial number of ship reports are found over landlocked regions (Slutz et al. 1985). In "untrimmed" COADS from 1970 to 1979 in 10° boxes, there were over 1000 reports in some landlocked areas and almost all landlocked boxes contained over 100 "observations." Presumably most, but not all, location errors are removed by tests on the data ("trimming").

Because the number of observations and the coverage decreases (Fig. 11), especially prior to 1950, the tendency has been to use as many observations as possible and require only 3 SST (Bottomley et al. 1990; Jones et al. 1991) observations per box to define an anomaly. Moreover, when computing decadal seasonal-mean 5° anomalies, Bottomley et al. (1990) required only one month to define a season and three seasons (out of 10) to define a decade in computing their decadal seasonal average fields, that is, perhaps as few as nine observations (although average numbers were higher) and an anticipated noise level of at least 0.35°C.

The use of 5° boxes accumulates observations over a fairly broad area and may cause severe problems associated with sampling gradients across a box. In Bottomley et al. (1990), errors arising from this source are reduced by use of a 1° grid for generating the anomalies. Based on section 4, ~25 observations per season/5° box is the minimum necessary to confidently define SSTs locally by reducing the noise to ~0.2°C, provided biases from SST spatial gradients across a box and from the seasonal cycle are accounted for properly. For January 1930 and 1950, Fig. 12 shows the drop in coverage when five observations per month per 2° box are required (compare with Fig. 11). Although three times as many observations might be expected over a season, it is clear from Figs. 11 and 12 that the observations are neither uniformly nor ran-

FIG. 9. Top: Correlation coefficient (×100) between SST anomalies from the UKMO and CAC analyses for 1982 through 1990 (108 months). Values greater than 0.9 are cross hatched, values less than 0.6 are stippled, and areas of no data are fine stippled. Middle: Root-mean-square differences between the monthly anomalies in hundredths of a degree Celsius; values exceeding 0.6°C are cross hatched. Bottom: percentage of months available for the other plots; values greater than 90% are cross hatched.

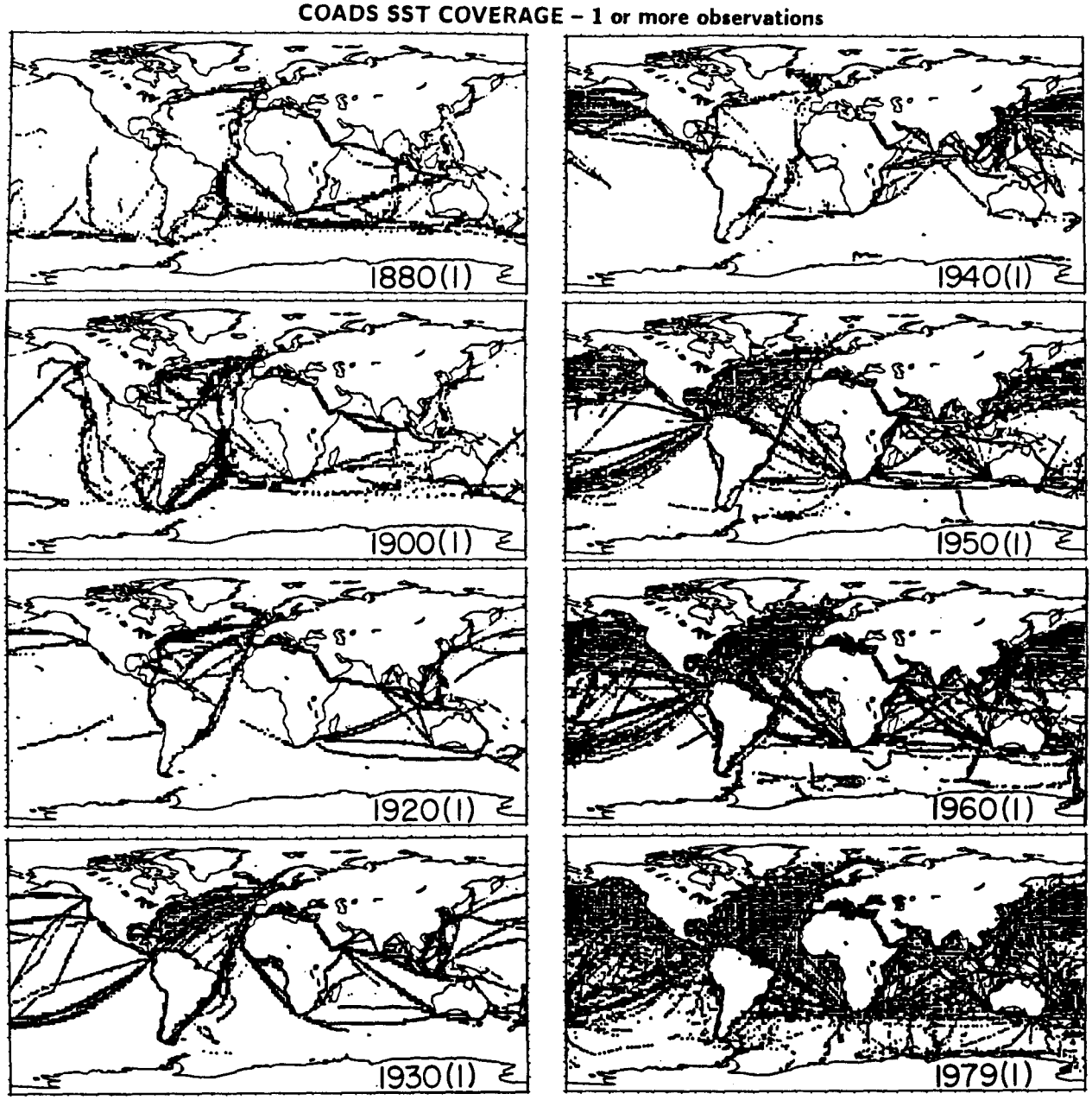


FIG. 11. Ship tracks of SST observations from COADS. In each panel a small x is plotted at the center of gravity of observations of one or more SSTs in each 2° box for the month of January of the year shown.

domly distributed throughout each 5° box. Therefore Fig. 12 gives a more representative estimate of what the reliable historical coverage has been.

As already noted, there are many different problems in reconstructing the climate record other than those experienced for the decade from 1979 to 1988. Examination of the observational database reveals a multitude of problems that lead to inhomogenities in the record, such as changes in instrumentation, exposure, and measurement techniques. For SST there have been many changes in measurement techniques over the years, with several different kinds of buckets used to bring up the water and different (and often unknown) instructions for exposure of the thermometer to the water before registering the reading. The greatest inhomogenity arises from the transition to taking measurements from thermometers installed in the intake pipe of seawater used to cool the engine of the ship

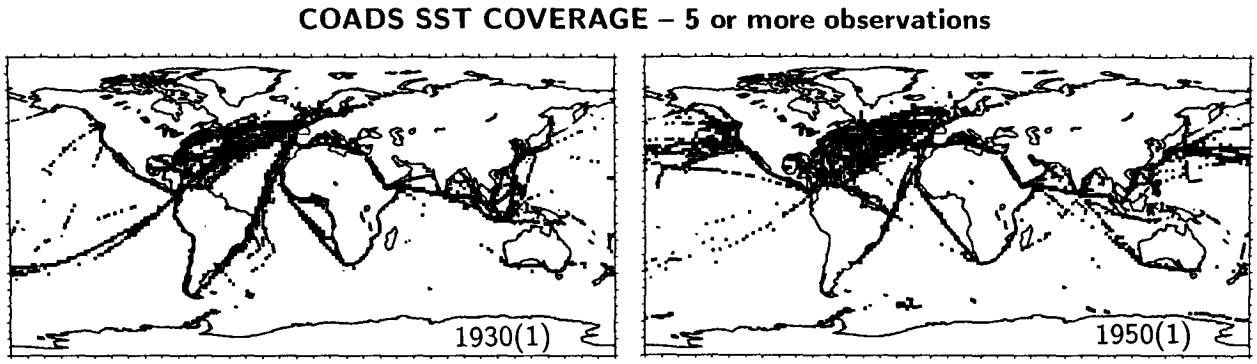


FIG. 12. Ship tracks of SST observations from COADS, as in Fig. 11 for the month of January of 1930 and 1950, except that five or more SSTs in each 2° box are required in each box.

(see Bottomley et al. 1990 and IPCC 1990 for a more complete discussion). Attempts have been made to derive adjustments to the SST record based on an assumed but unknown distribution of bucket versus intake measurements, or by making use of apparent changes in the annual cycle. Other adjustments have used coastal stations as a control, thereby potentially spreading effects such as those due to urbanization over the oceans. Adjustments are unnecessary after World War II, but earlier adjustments are as large as the apparent signal and the results from Jones et al. (1991) and from the UKMO (Folland et al. 1984; Bottomley et al. 1990) differ, especially prior to 1900, where values from the UKMO are ~0.2°C higher.

Combined with the problems in correcting for systematic errors, the random error component in SSTs and limited number of observations show how difficult it is to obtain a reliable estimate of the past surface temperatures over the global oceans.

The preceding comments are all directed toward the level of noise in the surface record. However, the character of the signal may also change. In particular, because of increases in greenhouse gases in the atmosphere, the climate is expected to warm and result in "global warming." If the warming is truly global then only one point or, in the presence of noise, just a few locations should capture that trend. For instance, the IPCC (1990) used a "frozen grid" analysis of the temperature trends using only those grid boxes with data from decades in the late 1800s or early 1900s. The results showed that the long-term trends in the full dataset were well captured. However, in neither the frozen grid nor the full dataset are the southern oceans sampled adequately, yet model experiments in which transient increases in carbon dioxide are included show a very slow and different response over the southern oceans (e.g., Manabe et al. 1991). Therefore, we cannot be sure that the signal on century time scales is truly global, and the frozen grid analysis says nothing about the more regional and local signals, such as those found in the 1980s.

6. Concluding remarks

At present, it is widely acknowledged that we do not have a system for monitoring global climate or even global surface climate. Moreover, there is no system in place, nor plans to put one in place, that will avoid the multitude of problems experienced with past data. Data from MSU channel 2 have emerged as a potentially valuable tool for monitoring global temperatures (Spencer and Christy 1990). However, it monitors a physically different quantity than surface temperature. As shown here, there is a fairly strong correlation between the temperature anomalies from the MSU and SFC datasets. The strength of that correlation depends a lot on the size of the signal. Larger signals are found over the extratropical continental regions, while the signal over the oceans is much smaller in general except where the El Niño phenomenon is prominent. On a monthly time scale, we have noted that there is also a real physical difference between surface air temperature and SST anomalies, but that SST anomalies are used to make up the SFC record. The MSU data provide an excellent complement to our climate monitoring capabilities.

The results of section 4 have implications for how SST data should be processed into means. The COADS SSTs are assembled into 2° boxes and monthly means and are therefore subject to the significant sampling errors discussed in section 4a, including errors of type (iv) (not allowing for annual cycle variations during the month), and type (v) (not allowing for gradients across a box). The UKMO SSTs reduce but do not eliminate errors from these sources by using a working grid of 1° and 5-day time periods. Nevertheless, these two kinds of errors could be avoided altogether by defining a good background climatological mean field for

each month and calculating the anomaly value for each observation relative to the exact interpolated mean value for that location and that day of the year. Errors of type (ii) from the diurnal cycle might be reduced by ensuring, wherever possible, that uniform sampling of the time of day occurs. Other sampling errors [types (i), measurement errors; and type (iii), real within-month variability] can be reduced only by sufficient numbers of observations to beat down the random component.

We discussed the sources of noise in SSTs, assessed the numbers of observations available, and then estimated the total noise in SST fields that use only in situ data. The lower correlations between the MSU and the SFC data over the oceans in the SH and in the western and central tropical Pacific are brought about partly by insufficient SST observations, which results in noise that dominates the modest climate signal. The differing ENSO signal between the MSU and the SSTs in the western tropical Pacific is another factor. The large discrepancies between the MSU and SFC data occurring in winter in the NH likely arise from a real physical difference associated with shallow wintertime inversions over land. We conclude that the general pattern in Fig. 1 reflects not only noise in the SFC record but also physical differences between the SFC and MSU, some of which are simply related to the difference between surface air temperature and SST over the oceans. Our assessments of noise were reinforced by the root-mean-square differences and correlations between SST anomalies in the CAC and UKMO monthly mean analyses.

It must be recalled that the different analyses of past data use the same observations and similar methods, so that just because they exhibit agreement does not ensure that the results are correct. On the contrary, the analysis here suggests that previous claims about our knowledge of past temperatures have been overstated.

Acknowledgments. We thank Amy Solomon and Dennis Shea for processing the COADS and other datasets. We also thank Phil Jones and Chris Folland for comments, and David Parker for advice on the SFC dataset. This research was partially sponsored by NASA Grants W-17661 and W-17214.

APPENDIX

Signal-to-Noise Ratio

We consider the common signal in two time series. The approach is related to a method called "common elements" (Snedecor and Cochran 1973, p. 181) and can be contrasted with regression analysis in which the two time series are not treated equally (one is the predictor and one is the predictand) and the variance of one series explained by the other, if both are normalized, is r^2, where r is the correlation coefficient.

Given two time series $x(t)$ and $y(t)$ with zero mean, that ostensibly measure the same quantity $z(t)$ in the presence of noise. The signal z will be measured by its variance S. We put

$$x(t) = z(t) + \epsilon(t)$$
$$y(t) = z(t) + e(t)$$

where $\epsilon(t)$ and $e(t)$ are the noise, assumed random, and give rise to noise variances N_x and N_y. Then the correlation coefficient between x and y is

$$r = \frac{S}{\sigma_x \sigma_y} \tag{A1}$$

where σ is the standard deviation. Because $\sigma_x^2 = S + N_x$ and $\sigma_y^2 = S + N_y$, a measure of the average noise is

$$N = 0.5(N_x + N_y) = 0.5(\sigma_x^2 + \sigma_y^2 - 2r\sigma_x\sigma_y)$$

using (A1), and the signal-to-noise ratio is

$$S/N = 2r \bigg/ \left(\frac{\sigma_x}{\sigma_y} + \frac{\sigma_y}{\sigma_x} - 2r\right) \tag{A2}$$

or, in the case where $\sigma_x \approx \sigma_y$, then

$$S/N = \frac{r}{1 - r}. \tag{A3}$$

Note also that the variance of the difference between x and y is simply $2N$, and this provides an additional reason for defining the noise in this way. Thus, the root-mean-square difference between two series is a direct measure of the square root of the noise.

The expression (A3) can be contrasted with that expected from regression analysis in which the r would be replaced with r^2. It is closely related to the expression for the signal-to-noise ratio in Trenberth (1984), which has a value of 1 when there is no signal. In (A3) S/N is zero for $r = 0$.

REFERENCES

Bengtsson, L., and J. Shukla, 1988: Integration of space and in situ observations to study global climate change. *Bull. Amer. Meteor. Soc.*, **69**, 1130–1143.

Bernstein, R. L., and D. B. Chelton, 1985: Large-scale sea surface temperature variability from satellite and shipboard measurements. *J. Geophys. Res.*, **90**, 11 619–11 630.

Blackmon, M. L., R. A. Madden, J. M. Wallace, and D. S. Gutzler, 1979: Geographic variations in the vertical structure of geopotential height fluctuations. *J. Atmos. Sci.*, **36**, 2450–2466.

Bottomley, M., C. K. Folland, J. Hsiung, R. E. Newell, and D. E. Parker, 1990: Global ocean surface temperature atlas. U.K. Meteor. Office. 20 pp., 313 plates.

Cornillon, P., and L. Stramma, 1985: The distribution of diurnal sea surface warming events in the western Sargasso Sea. *J. Geophys. Res.*, **90**, 11 811–11 815.

Fissel, D., S. Pond, and M. Miyake, 1976: Spectra of surface atmospheric quantities at Ocean Weathership P. *Atmosphere*, **14**, 77–97.

Folland, C. K., D. E. Parker, and F. E. Kates, 1984: Worldwide

marine temperature fluctuations, 1856–1981. *Nature,* **310,** 670–673.

Fu, L-L, D. B. Chelton, and V. Zlotnicki, 1988: Satellite altimetry: Observing ocean variability from space. *Oceanography,* **1,** 4–11.

Gutzler, D. S., R. D. Rosen, and D. A. Salstein, 1988: Patterns of variability in the Northern Hemisphere wintertime 850-mb temperature field. *J. Climate,* **1,** 949–964.

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

——, and ——, 1988: Global surface air temperatures: Update through 1987. *Geophys. Res. Lett.,* **15,** 323–326.

Horel, J. D., and J. M. Wallace, 1981: Planetary scale atmospheric phenomena associated with the Southern Oscillation. *Mon. Wea. Rev.,* **109,** 813–829.

Hurrell, J. W., and K. E. Trenberth, 1992: An evaluation of monthly mean MSU and ECMWF global atmospheric temperatures for monitoring climate. *J. Climate,* **5,** 1424–1440.

IPCC, 1990; Scientific Assessment of Climate Change. IPCC WG I, WMO, UNEP. J. T. Houghton, G. J. Jenkins, and J. J. Ephraums, Eds. U. Cambridge Press, 365 pp.

James, R. W., and P. T. Fox, 1972: Comparative sea surface temperature measurements. Marine Science Affairs Rep. No. 5, WMO 336, 27 pp.

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

——, S. C. B. Raper, R. S. Bradley, H. F. Diaz, P. M. Kelly, and T. M. L. Wigley, 1986a: Northern Hemisphere surface air temperature variations: 1851–1984. *J. Climate Appl. Meteor.,* **25,** 161–179.

——, ——, and T. M. L. Wigley, 1986b: Southern Hemisphere surface air temperature variations: 1851–1984. *J. Climate Appl. Meteor.,* **25,** 1213–1230.

——, T. M. L. Wigley, and G. Farmer, 1991: Marine and land temperature data sets: A comparison and a look at recent trends. *Greenhouse-Gas-Induced Climatic Change: A Critical Appraisal of Simulations and Observations.* M. E. Schlesinger, Ed. Elsevier, 153–172.

Kidson, J. W., and K. E. Trenberth, 1988, Effects of missing data on estimates of monthly mean general circulation statistics. *J. Climate,* **1,** 1261–1275.

Kraus, E. B., and R. E. Morrison, 1966: Local interactions between the sea and the air at monthly and annual time scales. *Quart. J. Roy. Meteor. Soc.,* **92,** 114–127.

Large, W. G., J. C. McWilliams, and P. P. Niiler, 1986: Upper ocean thermal response to strong autumnal forcing of the northeast Pacific. *J. Phys. Oceanogr.,* **16,** 1524–1550.

Manabe, S., R. J. Stouffer, M. J. Spelman, and K. Bryan, 1991: Transient responses of a coupled ocean–atmosphere model to gradual changes of atmospheric CO_2. Part I: Annual mean response. *J. Climate,* **4,** 785–818.

Morrissey, M. L., 1990: An evaluation of ship data in the equatorial Pacific. *J. Climate,* **3,** 99–112.

Newell, R. E., and B. C. Weare, 1976: Factors governing tropospheric mean temperatures. *Science* **194,** 1413–1414.

Ostapoff, F., and S. Worthem, 1974: The intradiurnal temperature variation in the upper ocean layer. *J. Phys. Oceanogr.,* **94,** 2005–2021.

Parker, D. E., 1984: The statistical effects of incomplete sampling of coherent data series. *J. Climatol.,* **4,** 445–449.

Price, J. F., R. A. Weller, and R. Pinkel, 1986: Diurnal cycling: Observations and models of the upper ocean response to diurnal heating, cooling, and wind mixing. *J. Geophys. Res.,* **91,** 8411–8427.

Rasmusson, E. M., and T. H. Carpenter, 1982: Variations in tropical sea surface temperature and surface wind fields associated with the Southern Oscillation/El Niño. *Mon. Wea. Rev.,* **110,** 354–384.

Reynolds, R. W., 1988: A real-time global sea surface temperature analysis. *J. Climate,* **1,** 75–86.

Saur, J. F. T., 1963: A study of the quality of sea water temperatures reported in logs of ships' weather observations. *J. Appl. Meteor.,* **2,** 417–427.

Snedecor, G. W., and W. G. Cochran, 1973: *Statistical Methods,* sixth ed., The Iowa State University Press. 593 pp.

Shea, D. J., K. E. Trenberth, and R. W. Reynolds, 1990: A global monthly sea surface temperature climatology. NCAR Tech. Note NCAR/TN-345+STR, 167 pp.

Slutz, R. J., S. J. Lubker, J. D. Hiscox, S. D. Woodruff, R. L. Jenne, D. H. Joseph, P. M. Steurer, and J. D. Elms, 1985: *COADS: Comprehensive Ocean–Atmosphere Data Set.* Release 1, 262 pp. [Available from Climate Research Program, Environmental Research Laboratories, 325 Broadway, Boulder, CO 80303.]

Spencer, R. W., and J. R. Christy, 1990: Precise monitoring of global temperature trends from satellites. *Science,* **247,** 1558–1562.

——, and ——, 1992a: Precision and radiosonde validation of satellite gridpoint temperature anomalies, Pt. I: MSU channel 2. *J. Climate,* **5,** 847–857.

——, and ——, 1992b: Precision and radiosonde validation of satellite gridpoint temperature anomalies, Pt. II: A tropospheric retrieval and trends during 1979–90. *J. Climate,* **5,** 858–866.

——, ——, and N. C. Grody, 1990: Global atmospheric temperature monitoring with satellite microwave measurements: Method and results 1979–1984. *J. Climate,* **3,** 1111–1128.

Tabata, S., 1978: Comparison of observations of sea surface temperatures at Ocean Station P and NOAA buoy stations and those made by merchant ships traveling in their vicinities, in the northeast Pacific Ocean. *J. Appl. Meteor.,* **17,** 374–385.

Trenberth, K. E. 1984: Signal versus noise in the Southern Oscillation. *Mon. Wea. Rev.,* **112,** 326–332.

——, and J. G. Olson, 1988: An evaluation and intercomparison of global analyses from NMC and ECMWF. *Bull. Amer. Meteor. Soc.,* **69,** 1047–1057.

Vinnikov, K. Ya, P. Ya. Groisman, and K. M. Lugina, 1990: Empirical data on contemporary global climate changes (temperature and precipitation). *J. Climate,* **3,** 662–677.