A Study of Six Operational Sea Surface Temperature Analyses

C. K. FOLLAND
Hadley Centre for Climate Prediction and Research, Meteorological Office, Bracknell, Berks, United Kingdom

R. W. REYNOLDS
National Meteorological Center, NWS/NOAA, Washington, D.C.

M. GORDON* AND D. E. PARKER
Hadley Centre for Climate Prediction and Research, Meteorological Office, Bracknell, Berks, United Kingdom

(Manuscript received 30 December 1990, in final form 25 March 1992)

ABSTRACT

This study results from recommendations made by a 1984 WMO Expert Committee on Ocean–Atmosphere Interaction Relevant to Long-Range Forecasting. The committee suggested that comparisons be carried out between monthly sea surface temperature (SST) analyses routinely made in several different countries in near real time. Emphasis was placed on the improvement of such analyses for use in operational long-range forecasting, especially for initializing dynamical long-range forecasting models. Six different monthly averaged SST analyses have been compared. The extent to which the analyses agree on several space scales and for regions covering the global oceans is shown, together with estimates of the magnitude of various types of errors. Independent estimates of SST obtained from expendable bathythermographs indicate that the monthly mean Meteorological Office (UKMO), Climate Analysis Center (CAC) in situ, and CAC blended analyses showed small differences (biases) from the expendable bathythermograph data. The differences were near to or below the margins of statistical significance over the Northern Hemisphere and the Southern Hemisphere tropics. Apparent negative biases in the analyses were noted, however, in the extratropical Southern Hemisphere.

The authors finish with a discussion of recent improvements to the accuracy and scope of SST analyses for both long-range forecasting and climate studies. These improvements include an integrated analysis of ice limit, in situ and satellite SST data, and the developing use of optimum interpolation as a method of SST analysis.

1. Introduction

A World Meteorological Organization (WMO) Expert Meeting was held in November 1984 in Geneva to review knowledge of the ocean–atmosphere system and to consider how it might be used to improve long-range forecasting (WMO 1986). A major recommendation was that a thorough and objective comparison be made among several near–real-time monthly mean global sea surface temperature (SST) analyses. The meeting requested that methods be devised for improving SST analyses to a level likely to be required in operational long-range forecasting. This study resulted from those recommendations. Because additional data are available from ship’s log books with a delay of up to several years, the results presented here do not test the highest “level” of SST datasets that can be created.

* Present address: Bartletts Cottages, Littlewick Green, Maidenhead, Berks, U.K.

Corresponding author address: David E. Parker, Meteorological Office, Hadley Centre for Climate Prediction and Research, London Road, Bracknell, Berkshire, RG12 2SY United Kingdom.

The comparisons, however, are more realistic for their main purpose, real-time long-range forecasting.

Statistical long-range forecasting requires long, homogeneous historical datasets, as well as near–real-time data. Both are used routinely in forecasting for the United Kingdom (Maryon and Storey 1985; Folland and Woodcock 1986) and for empirical seasonal forecasting for the tropics (Barnett et al. 1988; Folland et al. 1991). Efforts to improve historical SST datasets are being made in the U.K. Meteorological Office and the Massachusetts Institute of Technology (Bottomley et al. 1990; Folland and Parker 1990), and by a consortium of researchers involved in improving and extending the Comprehensive Ocean–Atmosphere Data Set (COADS) marine dataset (Woodruff et al. 1987; IPCC 1990). This work involves problems of data accuracy and coverage additional to those discussed here.

2. Outline of the comparisons

We have compared six different monthly SST analyses for the period January 1982 to December 1984, a period chosen by the WMO Expert Meeting (see Introduction). We could have extended at least some of
the comparisons to more recent years, but major qualitative changes in analysis schemes are only just beginning to be put into effect. Thus, our results remain representative of current performance, as well as illustrative of general principles such as the influence of temporal and spatial resolution and of atmospheric aerosols on satellite data. The period 1982–84 was notable for an exceptionally strong El Niño/Southern Oscillation (ENSO) event (lasting from boreal spring 1982 to autumn 1983) followed by a weak cold phase (much of 1984). The following near-real-time analyses were available on a 5° latitude × 5° longitude (5° box) resolution:

1) U.K. Meteorological Office (UKMO) in situ (ship and buoy) analysis;
2–4) U.S. Climate Analysis Center (CAC) in situ, satellite, and blended (in situ and satellite) analyses;
5) USSR Hydrometeorological Centre in situ analysis;
6) U.S. Navy Fleet Numerical Oceanography Center combined in situ and satellite analysis.

For each dataset, time series of SST anomalies, that is, deviations from climatological averages, were computed for a variety of ocean regions using a common set of 1951–80 averages calculated from the UKMO analysis. Estimates were made of the statistical significance of differences between selected analyses, and to help estimate biases, an independent SST dataset derived from expendable bathythermographs (XBT) was employed.

3. SST data and analyses

a. Sources of data

There were two main sources of SST data: in situ and satellite. In situ data were obtained from ship and buoy reports derived from messages sent over the global telecommunication system (GTS). During 1982–84, over 80% of the in situ observations were taken by ships. Approximately 70% of the ship values are believed to be derived from engine cooling water intakes (WMO 1988). Most remaining observations are thought to have been made using insulated buckets, but a small percentage are known to be derived from temperature sensors mounted on ships' hulls. Positive average differences (several tenths degrees Celsius) between intake and bucket temperatures have been well documented (e.g., Saur 1963; Walden 1966; James and Fox 1972; Tabata 1978a; Barnett 1984). As reported, however, in section 7 and Appendix A, a study of such differences using nearly six million observations for the period 1975–81 shows that average intake temperatures exceeded average bucket temperatures by only about 0.1°C. This may be because modern buckets are mainly “insulated,” and so suffer less cooling due to evaporation and sensible heat loss than did the older uninsulated buckets. The British voluntary observing fleet changed from the predominant use of uninsulated canvas buckets to that of insulated (black) rubber buckets in the 1960s and early 1970s (Bottomley et al. 1990; Folland and Parker 1990; WMO 1988, and earlier editions).

The satellite observations are obtained from the Advanced Very High Resolution Radiometer (AVHRR) on the U.S. National Oceanographic and Atmospheric Administration (NOAA) polar-orbiting satellites. The satellite SST retrieval algorithms are “tuned” by nonlinear regression against quality controlled drifting buoy data using the multichannel processing technique of McClain et al. (1985) and Walton (1988). The tuning is done when a new satellite becomes operational, or occasionally, when verification with the buoy data indicates errors. The multichannel SST retrievals first became operational in November 1981. The algorithms are computed globally and are not a function of position or time.

The coverage and number of in situ and satellite data differ considerably. During 1982–1984, there were typically 10⁵ in situ observations per month. Because the maximum density of in situ observations mostly occurs along major shipping routes (buoy data provide exceptions), their coverage is best in Northern Hemisphere midlatitudes. By contrast, there were several million satellite observations per month with a greater, though not completely uniform, coverage especially in the Southern Hemisphere (see Fig. 1 of Reynolds and Marsico 1992). The AVHRR SST retrievals require cloud-free regions, so satellite SST coverage was reduced in areas that were frequently cloud covered. The number of satellite observations also decreased dramatically after March 1982 in certain, mainly tropical, regions. This was caused by contamination of the data by the influence of stratospheric aerosols following the volcanic eruptions of El Chichón in March and April 1982 (Strong 1983). The monthly total of satellite SST observations fell to only 200 000 in October 1982 but recovered by mid-1983. Moreover, contamination of the remaining observations caused local negative biases in SST values of up to several degrees Celsius (Robock 1989; Reynolds et al. 1989). The following section describes how the various datasets were analyzed.

b. Analysis and quality-control procedures

1) U.K. METEOROLOGICAL OFFICE (UKMO) IN SITU ANALYSIS

The UKMO analysis was created in near real time from GTS ship and buoy data. Bottomley et al. (1990) give full details of the analytic process. After elimination of physically unreasonable or climatologically unreasonable values (1.6% of the total), average SST anomalies, from a 1951–80 1° box climatology that varied with each five-day period (pentad) of the year, were calculated for each pentad on a 1° box grid. Anomalies larger in magnitude than 6°C (7.5°C in the
tropical east Pacific) were rejected. Calendar months were approximated by periods six pentads long except for August, which was assigned 7 pentads. One 7-pentad month was needed, as there are 73 pentads per year; August was chosen as the “long” month, as climatological mean SST varies little then in most of the global ocean. Thus, “months” varied in length by up to four days from their true calendar equivalent. For each 5° box and month, 1° pentad SST anomalies were ranked, and the ranked values trimmed using a technique called “winsorization” (Affifi and Azen 1979). This minimizes the influence of extreme 1° pentad anomalies without artificially reducing the magnitude of large coherent SST anomalies, a problem that appears to have occasionally affected the COADS SST dataset (Wolter 1989). The winsorized mean anomaly for a given month and 5° box is the average of four numbers derived from the quartiles of the frequency distribution of 1° pentad anomalies for that month. The four numbers are: (a) the value of the boundary between the first and second quartiles; (b, c) the averages of the second and third quartiles; and (d) the value of the boundary between the third and fourth quartiles. For 5° boxes containing only four 1° pentad anomalies, winsorization is equivalent to simple averaging. In this study, we accepted such boxes, but those with less than four values were rejected.

2) U.S. CLIMATE ANALYSIS CENTER (CAC) IN SITU ANALYSIS

This was also produced in near-real time from in situ GTS reports, but on a 2° grid. After elimination of physically unreasonable data (2% of the total), in situ SST values were averaged on a 2° grid resolution and converted to anomalies from the monthly climatology of Reynolds and Roberts (1987). Very large anomalies were rejected (Reynolds 1988 gives details). The most important step was the use of a nonlinear filter to smooth the data (Reynolds 1988, Appendix B). This was designed to eliminate extreme gridred anomalies without spreading their signal spatially; however, it degraded the original 2° box resolution to about 6°.

3) CAC SATELLITE SST ANALYSIS

All available multichannel AVHRR retrievals were averaged over the same 2° grid for each month and converted to anomalies from the Reynolds and Roberts climatology. Because of the larger number of observations, analysis of the satellite data was simpler than that of in situ data. The most important step was the use of a linear two-dimensional binomial filter with weights 0.25, 0.5, and 0.25 to smooth the anomaly field to eliminate grid-scale noise. Smoothing was required to provide good estimates of the first and second derivatives of the satellite SST anomaly field needed to create the CAC blended analysis.

4) CAC BLENDED SHIP-SATELLITE SST ANALYSIS

This used the CAC in situ and CAC satellite 2° analyses as input fields. The in situ anomalies were used to provide “ground truth” in regions of frequent in situ data. Elsewhere, the CAC satellite anomalies were used to define the shape of the SST anomaly field, and the in situ field was used to minimize relative biases in the satellite values. This was done by adjusting the satellite values to be consistent with boundary values taken from the in situ data. An appropriate form of Poisson’s equation was used to make these adjustments (see section 3.3 and Reynolds 1988). Except where in situ data are present, an absence of analyzed ice limits tends to cause warm biases to occur in the blended SST field near the real ice edge in the Southern Ocean and in the Arctic. This problem is discussed in section 8.

5) USSR HYDROMETEOROLOGICAL CENTRE IN SITU ANALYSIS

The SST field was produced by a hand analysis at the Hydrometeorological Centre in Moscow from in situ data averaged over a 5° grid. The analysis was available for the most of the Northern Hemisphere in January and July 1982–84 in the form of contour charts of absolute monthly SST values. Values in much of the Caribbean, in the North Pacific east of 115°W, and in the sea of Japan were not available.

6) U.S. NAVY FLEET NUMERICAL OCEANOGRAPHY CENTER (FNOC) IN SITU AND SATELLITE ANALYSIS

This was computed globally from daily operational analyses derived from in situ and satellite data and a background field. During the period of study, only limited AVHRR satellite data were available to FNOC. The FNOC analysis was carried out on a polar stereographic grid with a spatial resolution of 2.5° at the equator; daily fields were averaged to produce monthly contoured fields. It was only possible to obtain monthly fields for January and July 1982–1984 in the Northern Hemisphere. At that time, the FNOC analysis was done using a method of blending SST fields from an ocean model and in situ data due to Holl et al. (1979). The analysis was later replaced by optimum interpolation as discussed by Clancy (1987) and Clancy et al. (1990).

To provide a uniform standard of comparison, the three CAC analyses were linearly interpolated to the centre of the UKMO 5° boxes, while the USSR and FNOC analyses were digitized at the centers of the 5° UKMO boxes. The latitudes and longitudes of the boundaries of the boxes are chosen to be integral multiples of 5°.

c. Use of a common reference SST climatology

The monthly mean UKMO climatology for 1951–80 described in Bottomley et al. (1990) was used to
produce all the 5° box SST anomalies. A common climatology eliminates spurious differences between analyses that can result from the use of different reference climatologies. The UKMO climatology was initially produced from in situ data on a 1° grid. The version of the UKMO climatology used here was extended over regions having initially no in situ values through the use of the Laplacian of digitized values, \( A \), of the Alexander and Mobeley (1976, hereafter AM) SST/sea ice climatology in these regions. The AM climatology is globally complete, but was derived from a variety of data during 1854–1972 (Reynolds 1983). To extend the original UKMO climatology, \( C \), its northern and southern limits were used to provide boundary conditions on a 1° × 1° grid to Poisson’s equation
\[
\nabla^2 C = \nabla^2 A.
\] (1)

In this way, the spatial rate of change of the gradient of the AM climatology was reflected in the extended UKMO climatology. Values of the AM climatology were adjusted by an amount related to differences between the two climatologies at nearby boundaries of the datasets. Close to a boundary, the adjustment was almost equal to their differences, but further away, adjustments to the AM climatology decreased. Arctic and Antarctic boundary conditions were defined by the AM climatological ice limit. Here, SST (initially on a 1° × 1° resolution) was set to −1.8°C, the freezing point of seawater with a salinity of 33 psu (practical salinity units).

4. Results

a. Comparisons of analyzed mean values over large ocean regions

Differences in data coverage between the analyses limit the number and location of 5° boxes common to all analyses. Because this paper is concerned with improving methods of analysis, it was felt appropriate to confine the comparisons in a given month to 5° boxes where they all had data. Within each ocean region, this “common area” changed from month to month. When the FNOC and USSR SST fields were not available, common 5° boxes were defined by the availability of data in the UKMO and three CAC analyses alone. Otherwise, common boxes within the Northern Hemisphere during January and July were defined from all six analyses. The majority of comparisons are carried out for ten “standard” ocean regions (Table 1), but other regions are used to make particular points. The acronyms for each region shown in Table 1 are used in the diagrams.

Figures 1a–f show monthly time series for five standard regions (NH, ENP, TNP, TSA, and MSH) and a small tropical northeast Atlantic region (10°N–25°N, 20°W–35°W). The CAC blended, in situ, and UKMO analyses give similar area-mean anomalies in most regions and months, including those standard regions not illustrated, though common box MSH anomalies were less consistent (Fig. 1e). On the other hand, large negative differences can be seen between the CAC satellite and the UKMO and other CAC analyses. Over the Northern Hemisphere as a whole, and in the extratropical North Pacific, the differences reached −1.0°C in early 1983 (Figs. 1a,b). Even larger differences were found in the tropics, for example, −1.6°C in the tropical North Atlantic in summer 1982 (not shown), −1.5°C in the tropical North Indian Ocean (not shown), and −1.2°C in the tropical North Pacific (Fig. 1c). Differences also reached −0.7°C and −0.9°C in the tropical South Atlantic around August 1982 and June 1983, respectively (Fig. 1d). They were generally least in the midlatitude Southern Hemisphere (Fig. 1e), but still evident in late 1982.

The relatively low satellite SSTs shown by these results are clearest from the second half of 1982 to mid 1983, depending on location. They are believed to result primarily from the influence of stratospheric aerosols ejected by El Chichón on the satellite radiances (Strong 1983) because the calibration algorithms did not properly account for the abnormally large emission of infrared radiation from the stratospheric aerosol at low stratospheric temperatures. Within days of the strongest eruption (4 April 1982), the latitude band centered near 15°N was greatly affected. The volcanic aerosol, and its associated cool satellite SST bias, spread latitudinally for several months and finally dispersed during the latter part of 1983. The widespread and lasting influence of El Chichón aerosols affected an analysis of satellite-derived global SST trends made by

| Table 1. Standard ocean regions. The numbers in parentheses are approximations to the average areas (in 10⁶ km²) of ocean sampled in each month by common boxes when defined by the UKMO and three CAC analyses. |
|---------------------------------|---------------------------------|---------------------------------|
| (NH) Northern Hemisphere (120) | 0°-80°N                        | 0°-60°S                        |
| (SH) Southern Hemisphere (95)  | N of 30°N including all coastal waters and the Mediterranean | N of 30°N, American coast to 120°E |
| (ENA) Extratropical North Atlantic (20) | 0°-30°N including the Caribbean and Gulf of Mexico | 0°-30°N, 30°E–120°E |
| (ENP) Extratropical North Pacific (22) | 0°-30°N, American coast to 120°E | 0°-30°N, American coast to 120°E |
| (TNA) Tropical North Atlantic (22) | 120°E–180°, 20°N–20°S | 0°-30°S |
| (TNI) Tropical North Indian Ocean (15) | (24, austral winter; 30, summer) | (24, austral winter; 30, summer) |
| (TNP) Tropical North Pacific (35) | | |
| (TWP) Tropical West Pacific (16) | | |
| (TSA) Tropical South Atlantic (15) | | |
| (MSH) Midlatitude Southern Hemisphere | | |
Fig. 1. Time series of monthly SST anomalies for six oceans using common 5° box data. Note the varying scales of the ordinates. The acronyms defined in Table 1 are indicated in parentheses. (a) Northern Hemisphere (NH), (b) Extratropical North Pacific (ENP), (c) Tropical North Pacific (TNP), (d) Tropical South Atlantic (TSA), (e) Midlatitude Southern Hemisphere (MSH), (f) Tropical northeast Atlantic.

Other relatively cool or warm satellite temperatures in Fig. 1 are less easily explained. They may be very persistent; a notable example is the tropical eastern North Atlantic near the west coast of North Africa (10°–25°N and 20°–35°W) shown in Fig. 1f. Here the CAC satellite analysis shows large negative differences (henceforth termed biases) from the UKMO and the CAC in situ and blended analyses for the entire record. The cool bias in the satellite data almost certainly results from uncompensated reductions in the satellite radiances by wind-driven tropospheric dust from the Sahara. During the summer of 1982 and in January 1983 the biases reached −2°C; such large values probably result from the joint influence of stratospheric (El Chichón) and tropospheric Saharan dust (McClain 1989) on the satellite radiances. Figure 1 shows that the Poisson adjustment technique almost eliminates these satellite biases from the CAC blended fields in all regions shown. Even in areas with sparse in situ data, biases are much reduced, for example, over the Southern Ocean. Fortunately, Fig. 1e suggests that, in 1982 and 1983, the contamination of satellite data may have been least in the Southern Ocean where the dependence on satellite data is the greatest.

The USSR and FNOC analyses were available only for the Januarys and Julys of each year for Northern Hemisphere areas (asterisks and circles, respectively, in Fig. 1). The USSR analysis, based on ship data, is fairly similar to the blended and the other in situ analyses, though anomalies tend to be slightly warmer. The FNOC analysis, however, shows substantial cool differences, especially in the second half of the study period. The pattern of FNOC differences is different from that of the CAC satellite analysis, suggesting problems unique to the FNOC analysis technique of Holl et al. (1979). Reynolds (1983) found problems when he computed a monthly mean climatology from analyses produced by this technique for the period 1946–78. He compared this with seven other climatologies and found that the climatology based on the Holl et al. analyses was very different from the others, particularly in the Northern Hemisphere summer. The FNOC analysis agrees better with in situ analyses in the extratropical North Pacific (Fig. 1b), and to a lesser extent in the extratropical North Atlantic (not shown) than it does in the tropical Northern Hemisphere oceans.

Table 2 shows mean SST anomalies in the ten standard regions. Three analyses—the UKMO, the CAC in situ, and CAC blend—agree closely in all regions. Agreement is slightly less in smaller regions (not shown), for example, the tropical northeast Atlantic, where sampling problems increase and the intrinsic variability of SST anomalies is greater. Table 2 also shows the root-mean-square differences (RMSD) between the CAC blended and remaining monthly analyses. Values of RMSD for the CAC in situ analysis are understandably small, as the blended analysis is constrained strongly to agree with it. The largest RMSD are in the midlatitude Southern Hemisphere, where the blended analysis is much influenced by satellite data. The UKMO analysis shows only slightly greater values of RMSD than does the CAC in situ analysis, confirming their generally great similarity.

The USSR analysis (for six months only) gives somewhat larger values of RMSD, while the CAC satellite and FNOC analyses (also for six months) give much larger RMSD. The large CAC satellite RMSD occur because the biases caused by stratospheric aero-

<table>
<thead>
<tr>
<th>Region</th>
<th>NH</th>
<th>SH</th>
<th>ENA</th>
<th>ENP</th>
<th>TNA</th>
<th>TNI</th>
<th>TNP</th>
<th>TWP</th>
<th>TSA</th>
<th>MSH</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) Mean anomalies</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CAC blend</td>
<td>.05</td>
<td>.24</td>
<td>−.07</td>
<td>−.20</td>
<td>.07</td>
<td>.23</td>
<td>.19</td>
<td>.10</td>
<td>.26</td>
<td>.10</td>
</tr>
<tr>
<td>UKMO in situ</td>
<td>.05</td>
<td>.29</td>
<td>−.09</td>
<td>−.23</td>
<td>.07</td>
<td>.31</td>
<td>.20</td>
<td>.13</td>
<td>.31</td>
<td>.21</td>
</tr>
<tr>
<td>CAC in situ</td>
<td>.06</td>
<td>.22</td>
<td>−.05</td>
<td>−.20</td>
<td>.08</td>
<td>.25</td>
<td>.20</td>
<td>.10</td>
<td>.26</td>
<td>.05</td>
</tr>
<tr>
<td>CAC satellite</td>
<td>−.43</td>
<td>−.02</td>
<td>−.28</td>
<td>−.47</td>
<td>−.68</td>
<td>−.38</td>
<td>−.36</td>
<td>−.39</td>
<td>−.10</td>
<td>.07</td>
</tr>
<tr>
<td>CAC blend*</td>
<td>.03</td>
<td>#</td>
<td>−.25</td>
<td>.11</td>
<td>.23</td>
<td>.15</td>
<td>#</td>
<td>#</td>
<td>#</td>
<td>#</td>
</tr>
<tr>
<td>USSR*</td>
<td>.09</td>
<td>#</td>
<td>.03</td>
<td>−.23</td>
<td>.26</td>
<td>.36</td>
<td>.19</td>
<td>#</td>
<td>#</td>
<td>#</td>
</tr>
<tr>
<td>FNOC*</td>
<td>−.45</td>
<td>#</td>
<td>−.33</td>
<td>−.51</td>
<td>−.46</td>
<td>−.43</td>
<td>−.46</td>
<td>#</td>
<td>#</td>
<td>#</td>
</tr>
</tbody>
</table>

(b) RMSD of CAC blended monthly average anomalies minus those of other analyses

| UKMO in situ    | .04 | .06 | .06 | .08 | .04 | .11 | .06 | .10 | .09 | .13 |
| CAC in situ     | .02 | .03 | .02 | .01 | .02 | .04 | .04 | .05 | .05 | .09 |
| CAC satellite   | .57 | .31 | .35 | .50 | .86 | .73 | .68 | .64 | .46 | .21 |
| USSR*           | .13 | #   | .24 | .14 | .16 | .19 | .11 | #   | #   | #   |
| FNOC*           | .57 | #   | .34 | .38 | .71 | .92 | .73 | #   | #   | #   |

* For Jan and Jul 1982–84 only.
# Not applicable.
sols changed strongly with time (Fig. 1). The large RMSD for the FNOC analysis also reflect large temporal changes in the anomalies of FNOC values relative to CAC blended values, which are evident in Fig. 1.

b. Ability of analyses to distinguish differences between successive monthly anomalies

For long-range forecasting, it is pertinent to ask whether the more consistent analyses can meaningfully distinguish the differences between successive monthly or seasonal mean SST anomalies over large areas. Harrison et al. (1990) concluded that changes in SST between successive months in $4^\circ$ latitude $\times 10^5$ longitude areas in the eastern equatorial Pacific needed to be at least 1$^\circ$C per month before the signs of changes shown by CAC blended and in situ analyses agreed; farther west, agreement of sign above a chance level was difficult to find.

For our larger standard areas, we have defined a “signal” of SST anomaly change as the standard deviation of changes in successive monthly ($\sigma_m$) or seasonal ($\sigma_s$) area-mean SST anomalies provided by the CAC blended analysis. Because these changes have been calculated using common $5^\circ$ boxes only, they are not truly representative owing to incomplete coverage especially in the extratropical Southern Ocean. Given this proviso, a measure of the extent to which an analysis can skillfully reproduce changes between successive monthly or seasonal mean SST is the signal-to-noise ratio

$$ R = \frac{\sigma_s}{\sigma_n}. $$

Here, $\sigma_n$ is defined to be the noise. It is the standard deviation of the differences between the area-mean SST changes given by the blended and nominally independent analysis. A reasonable criterion for a minimum acceptable “true” signal in the independent analysis is that the noise should be half the signal, that is, $R = 2$. However, $\sigma_n$ may itself contain some noise, so a rather higher value of $R$ may be needed for this demonstration. The standard deviation of the interseasonal changes tends to be similar to, or rather larger in size than, intermonthly ones (not shown).

Figure 2 shows $R$ for the UKMO and CAC satellite analyses on the monthly and seasonal time scales for the ten regions. The CAC satellite analysis is not quite independent of the blended analysis, but because only those $5^\circ$ boxes are considered where ship data exist, the CAC ship analysis tends to contribute most to the blended analysis. Despite its lack of complete independence, the CAC satellite analysis is so different from the blended analysis that $R$ is usually much less than unity on both monthly and seasonal time scales. In fact, $R$ only slightly exceeds unity on the seasonal time scale in the tropical South Atlantic ($R = 1.3$). Therefore, the satellite analysis is not sufficiently consistent with the blended analysis anywhere to identify the same changes of monthly and seasonal mean SST over the above regions (varying in area from $10^7$ to $10^8$ km$^2$).

Figure 1 indicates that $\sigma_n$ tends to be large because the satellite bias varies strongly in time in most regions. The apparent average biases in the satellite data are a separate cause of incompatibility, as they do not contribute to $\sigma_n$. Over the Northern Hemisphere $R$ is as small as 0.13 and 0.29 on monthly and seasonal time scales, respectively, though most of this region is well observed by in situ data. These results confirm the remarkable ability of the CAC blending procedure to remove time-varying biases in satellite SST data relative to values measured by ships and buoys.

The independent UKMO in situ analysis gives values of $R$ in the range 2–8 on seasonal time scales and 1–3 on monthly ones, a much larger signal-to-noise ratio than for the satellite data. This is not surprising, as the satellite data contributing to the blended analysis are corrected to largely agree with in situ data. Nevertheless, on the monthly time scale $R$ reaches a maximum of only 3 in the tropical North Atlantic for the UKMO analysis. Overall, agreement between the UKMO and the blended analysis in their assessment of intermonthly changes of SST, relative to the size of those changes, is at best moderate. For the Southern Hemisphere extratropics, a value of $R = 0.93$ indicates little agreement between the analyses. For the Southern Hemisphere as a whole, $R$ is still only 1.2. For seasonal changes, the situation is much better; $R$ fails to reach 3 only in the extratropical Southern Hemisphere ($R = 2.56$) and tropical North Pacific ($R = 2.42$). The best in situ data give the highest values of $R$; thus, the tropical North Atlantic gives considerably higher monthly and seasonal values ($R = 3.13$ and 8.15, respectively) than does the tropical North Indian Ocean ($R = 1.57$ and 3.60), though the latter region shows more variability.
The smaller ocean areas analyzed in Fig. 2 tend to show the largest interseasonal and intermonthly values of $R$, indicating that the greater real temporal variability of SST in smaller regions provides a better chance of picking out the signal against errors in the data and a less than perfect analysis procedure. In yet smaller regions where in situ data are not plentiful, however, $R$ may decrease again due to the joint influences of mesoscale SST variability and the greater influence of undetected errors in individual in situ values (section 4c).

An alternative definition of the signal could be the true mean modulus of the intermonthly or interseasonal changes, estimated from the blended analysis, and the noise could be defined as the mean modulus of the differences between changes in the blended and the independent analysis. As long as the mean change is near zero and the changes are approximately normally distributed, however, there should be no appreciable difference in resulting values of $R$.

For studies of climate, the above analysis needs to be repeated on monthly to at least annual time scales for a considerably longer period and over a wider range of areal sizes. Nevertheless the UKMO analysis seems acceptably similar to the blended analysis for detecting interseasonal SST changes, though only marginally so in the tropical North Pacific and midlatitude Southern Hemisphere. The monthly time scale, however, is often too short for the UKMO analysis to provide reliable estimates of SST changes. The CAC in situ analysis is likely to be similar, but Eq. (2) cannot be used to deduce this directly because the CAC in situ analysis is not created independently of the blended analysis.

c. Effects of analysis spatial resolution

We investigate the general character of differences between the analyses as a function of spatial resolution, and enquire as to whether a 5° box resolution is justified given the limited amount of information about sea surface temperature that is currently available.

1) VARIATION IN THE RMSD BETWEEN THE BLENDED AND OTHER ANALYSES WITH RESOLUTION

Figure 3a shows the RMSD between the UKMO and the CAC blended analysis for the ten standard regions for (i) each region as a whole, (ii) constituent common 5° boxes, (iii) 10° boxes (four times less resolution than 5° boxes), and (d) 10° latitude × 20° longitude boxes (eight times less resolution). Figure 3b gives a corresponding comparison between the blended and the CAC satellite analysis. The CAC in situ data were not analyzed, as they are insufficiently independent of the blended analysis. Values over the areas larger than 5° boxes were created from available 5° boxes; thus, the larger areas may not always contain a full complement of 5° box anomalies. Boxes have been weighted for their varying areas according to latitude and the fraction of ocean contained in a box when part of its area is land. The 10° latitude × 20° longitude boxes are included because SST anomalies tend to have a greater correlation length in an east–west compared with a north–south direction in the tropics (Clancy et al. 1990). Because the 10° and 10° × 20° boxes were computed by averaging constituent 5° boxes, the coarser grids smooth the 5° values and root-mean-square differences should decrease.

The RMSD of the monthly CAC satellite (Fig. 3b) and FNOC analyses (not shown) from the blended
analysis are often five or more times larger for each region as a whole than for the UKMO in situ analysis. The USSR in situ analysis behaves more like the UKMO analysis, though RMSD (not shown) tend to be about twice as great. For the U.K. and USSR analyses, RMSDs decrease rapidly, as expected, as the spatial scale of the analysis region increases from 5° to that of the region. Because the CAC satellite and FNOC analyses are dominated by large time-varying average differences from the blended analysis, however, their RMSDs behave quite differently and decrease little as the resolution coarsens. We now investigate whether the larger RMSD in the UKMO analysis on the 5° resolution merely reflects an inevitably larger noise that accompanies a larger signal as resolution decreases, or whether the noise is so large on the 5° resolution that the signal has become obscured.

2) Reproducibility of 5° Box Resolution SST in the UKMO and Blended Analysis

These two analyses were chosen because they are created independently and data was most plentiful. To assess the reproducibility of 5° resolution SST between the two analyses, we introduce a second type of signal-to-noise ratio, $S_2$. Here, $S_2$ is defined as the ratio of the standard deviation of all monthly mean anomalies for a 5° area given by the blended analysis, $\sigma_{bls}$, to that of the differences between the UKMO and the blended analysis, $\sigma_{ds}$. The average value of $S_2$ over the whole region is then

$$S_{2av} = \frac{1}{N} \left\{ \sum \left( \frac{\sigma_{bls}}{\sigma_{ds}} \right)^2 \right\}^{0.5}.$$  (3a)

A similarly based ratio, $S$, has also been calculated for a resolution degraded to that of the area as a whole from

$$S = \frac{\sigma_{bl}}{\sigma_d},$$  (3b)

where $S$ is the analogue of $R$. For many purposes, measures of error like $S$ may be more appropriate than measures like $R$. Thus, although it may be difficult to get the timing or detailed character of monthly SST changes correct, the fact that a monthly SST value in January of one year is correctly assessed to be substantially different from that in the previous January is likely to have considerable value in many studies. An example is the difference between an El Niño state in one January and an anti-El Niño in the previous January. The SST accuracy needed for studies of detailed physical mechanisms is, however, more likely to be reflected by $R$.

Figure 3c shows that with a resolution reduced to that of the regions as a whole, $S$ [Eq. (3b)] varies in much the same way as does $R$ (Fig. 2), but with larger values. Thus, monthly values of $S$ tend to be similar to seasonal values of $R$, confirming that it is easier to skillfully capture the overall character of the variation of monthly SST anomalies than to skillfully estimate intermonthly SST changes. Looked at this way, Fig. 3c shows that monthly UKMO analyses are useful on the regional scale except in the midlatitude Southern Hemisphere. Now $S_{2av}$ (Fig. 3c) is similar to $S$ in the Northern and Southern hemispheres and the extratropical North Atlantic and Pacific, but smaller elsewhere. Indeed, $S_{2av}$ is generally less than 2. Accepting the blended analysis as the better of the two because of its extra information, this indicates that monthly in situ UKMO data are often unreliable on a 5° space scale. Figure 3c, however, indicates that the UKMO analysis may be considerably more useful on the regional resolution, and that large-scale patterns of SST anomalies may often be quite well captured, though not in the midlatitude Southern Hemisphere.

5. Correlations of anomalies on a 5° resolution

Time series of the spatial correlations between fields of anomalies from different analyses on a 5° resolution show appreciable temporal changes and different patterns from region to region. Figure 4a shows time series of spatial correlations between the CAC blended and (a) UKMO, (b) CAC in situ, and (c) CAC satellite analysis for the Northern Hemisphere. Correlations between the CAC blended and in situ analyses consistently exceed 0.9. This further demonstrates that the blended analysis preserves the major spatial features of the CAC in situ analysis. The correlations between the blended and UKMO analyses are close to 0.8, except for some lower values in early 1982. Correlations between the blended and satellite analyses are generally smallest; they reached their lowest values in August 1982 when the El Chichón aerosol strongly affected the satellite retrievals, causing cooler anomalies that were frequently of opposite sign to those of other analyses. With the dispersal of the aerosols, the correlations between the blended and satellite analyses improved to about 0.7 in 1983 and 1984. Figure 4a also helps to place the RMSD shown in Figs. 3a and 3b in perspective. The mean rms error for the whole period between the UKMO and CAC blended analysis of 0.43°C for 5° boxes in the Northern Hemisphere (Fig. 3a) translates to a mean correlation between the two analyses on this resolution of 0.80.

To illustrate the effect of a strong SST anomaly signal on the 5° box correlations, time series of correlations between the UKMO and the three CAC analyses for the tropical South Pacific (120°E to South American coast and 0°–30°S) are shown in Fig. 4b. Use of this region avoided the degrading effect of the El Chichón aerosols on the 1982 portion of the satellite analyses. Between about August 1982 and August 1983, strong warm anomalies occurred over much of the tropical South Pacific due to the 1982–83 El Niño, with fairly
Fig. 4. (a) Monthly time series of spatial-anomaly correlation for the Northern Hemisphere ocean. The CAC blended analysis was correlated with the UKMO, CAC in situ, and CAC satellite analyses. (b) As in (a) but for the tropical South Pacific (30°S–30°N). The UKMO analysis was correlated with the CAC in situ, satellite, and blended analyses. (c) Monthly SST anomalies from the UKMO analysis for January 1983. Contour interval is 1°C; areas with anomalies cooler than −1°C are stippled, and areas with anomalies warmer than 1°C are hatched. Regions with limited data are dotted. (d) As in (c) but for the CAC blended analysis.
strong negative anomalies in its southwest. This gave a very pronounced SST anomaly pattern (Bottomley et al. 1990, plates 220–243) and correlations between the 5° box UKMO analyses and any of the 5° box CAC analyses in this region were close to 0.9. Before and after the event, correlations were much lower; this is to be expected as weak anomalies, although not dissimilar in magnitude, may be of opposite sign in given 5° boxes. Thus, May 1982 shows a pronounced minimum correlation at a time of many weak anomalies.

Figures 4c and 4d show the UKMO and CAC blended analyses for January 1983, a month close to the peak strength of the 1982–83 El Niño, highlighting some differences between their character. The blended analysis is smoother because of the use of a strong spatial filter (Reynolds 1988), that effectively reduces its resolution to about 6°. There is also a greater local influence of questionable in situ values in the UKMO analysis. The disagreement in sign west of Chile illustrates the unreliability of sparse in situ data there. Dotted regions show 5° boxes where the individual analyses were poorly defined. In the UKMO analysis, dots denote 5° boxes with less than four observations; these box values were rejected, so isopleths were interpolated if possible, or else omitted. In the blended analysis, dotted regions include ice-covered ocean, but in the tropical Pacific near 10°N, there was a lack of satellite retrievals because of interference from the El Chichón stratospheric aerosol. Comparison of the extent of the dotted regions shows the extensive improvement in data coverage that can be obtained from using (adjusted) satellite data.

6. Analysis biases: Comparisons with independent expendable bathythermograph data

To help determine the accuracy of the analyses, expendable bathythermographs (XBTs) were used to provide near-surface “reference” SST data. These data were not used in any of the analyses. In addition to ship identifier, time, and location, XBT messages consist of paired measurements of temperature versus ocean depth usually starting very close to the surface. Although the XBT observations were taken along normal shipping routes and were limited in number compared with conventional surface marine observations (about 5%), XBT measurements are considered to be of generally higher accuracy (Heinmiller et al. 1983) if appropriate quality control is carried out. An exception is thermal-recorder engine intake data that can, after quality control, have the same accuracy as XBT data or better (Tabata 1978b). The XBT data used here were taken from GTS radio messages and had a widespread geographical distribution. Figure 5 shows the locations of all XBT data available, each represented by one dot.

Often XBTs are dropped into the water without any attempt to preadjust the temperature of the probe to that of the ocean surface temperature. Consequently, the uppermost XBT data value can show unrealistically high or low temperatures; this value is often measured at a depth of around 2 m or 2 seconds after the probe hits the water (Roemmich and Cornuelle 1987). The problem is shown by a surface temperature “transient,” occasionally of up to several degrees in size. The size

![Fig. 5. Locations of all available XBT data (Jan. 1982–Dec. 1984). Each value is shown by a point on the map.](image-url)
of the effect depends partly on the thermal capacity of the XBT nose, and so can vary between XBT types for given conditions (Roemmich and Cornuelle 1987). To minimize bad data that can result from surface temperature transients, all temperature and depth pairs within the first 10 m were quality controlled. If there was only one pair of data in the first 10 m (75% of the total number of XBT observations), the corresponding temperature value was selected. If there were two or more pairs, the temperature at the second most shallow depth was selected (usually about 10 m). The value was converted to an anomaly by interpolating the climatological mean temperature for the day and geographical point appropriate to the XBT observation using the monthly 5° UKMO climatology. Because radio messages are subject to transmission errors, the XBT SST value was also subjected to a gross error check. This was done by eliminating any value whose anomaly differed from the collocated blended anomaly by more than 5°C. Because Reynolds (1988) found that the typical monthly blended rms error calculated relative to monthly mean buoy temperatures was 0.7°C, this quality control procedure is conservative since differences have to be larger than seven times the typical blended “error” before elimination.

Comparisons of the six analyses with an XBT SST anomaly at a given location were carried out only if a linearly interpolated value could be computed from all analyses for a given month. This condition was met when the four nearest 5° box values all had an analyzed value. The USSR and FNOC analyses contributed to this condition only in January and July of each year in regions located in the Northern Hemisphere. The collocated analysis anomaly minus the XBT anomaly or “bias” was used to generate areally averaged monthly and whole period mean differences and standard-error statistics. The standard error statistic was defined as the standard deviation of the monthly areally averaged bias B divided by the square root of the equivalent number n' of independent values of B through time, allowing for persistence in B following the method of Quenouille (1952). For the CAC satellite analysis, persistence reduced the number of independent values of B to as little as six from a maximum of 36 for several regions. Thus, values six months apart were usually independent. For the other CAC analyses and the UKMO analysis, n' was estimated from the UKMO analysis. In many regions no intermonthly persistence of B was found but some Northern Hemisphere regions did show noticeable persistence. The six values of B in each region derived from the FNOC and USSR analyses were regarded as effectively independent, being observed six months apart. Statistical significance was computed for the mean bias of a given analysis, \( B_m \), by assuming that monthly mean biases were normally distributed. Here, \( B_m \) was regarded as significantly different from zero at the 95% confidence level if it exceeded twice the standard error statistic.

To help cast further light on recent controversies about the accuracy of satellite data and other SST analyses (Strong 1989; Reynolds et al. 1989; Robock 1989), Figs. 6a–6d show \( B_m \) and its standard error for up to twelve regions and four of the analyses. The three extra regions are as follows:

(ECP) Equatorial Central Pacific, 10°N–10°S, 170°–130°W
(TSI) Tropical South Indian Ocean, African coast–120°E, 0°–30°S
(TSP) Tropical South Pacific, 120°E–South American coast, 0°–30°S

Table 3 shows, additionally, the average number of XBT data used in each month and region, \( n_x \). Note that the actual numbers varied considerably from month to month and that fewer XBT data were actually used than were available because of the need to have neighboring 5° boxes with data and because some XBTs were rejected for being mislocated over land (Fig. 5). A total of about 47 100 XBTs were used in the Northern Hemisphere and 5600 in the Southern Hemisphere.

Figure 6d shows that over the three-year period as a whole, the CAC satellite data are biased cold relative to the XBT data by more than two standard errors in most regions north and south of the equator. Negative biases occur in all regions, including the midlatitude Southern Hemisphere, though they are very small (−0.08°C) in the extratropical North Atlantic. This may be because the satellite retrievals were least affected here by El Chichón. In the tropics, the typical satellite bias is −0.5°C, the value also seen in the midlatitude Southern Hemisphere (but see below).

The in situ and blended data are in much better agreement in all regions (Figs. 6a–6c) except the midlatitude Southern Hemisphere where they also average a surprising 0.5°C too cold (see below). When all data for the Northern Hemisphere are put together, there is a hint that the overall differences of 0.07°C, 0.10°C, and 0.10°C for the UKMO, CAC in situ, and CAC blended analyses really are warm (in most regions), but these small warm biases are only just significant for the two CAC analyses over the hemisphere as a whole. The average biases for the Southern Hemisphere tropics (giving equal weight for this calculation to the three oceans) are 0.01°C, −0.03°C, and −0.03°C, respectively, giving remarkable agreement, though a three-year total of only 4100 widely spread (in time and space) accepted XBT data contribute to this result. These results should therefore be viewed cautiously. All analyses appear to be much too cold relative to the XBT data in the midlatitude Southern Hemisphere (Fig. 6). This conclusion needs verification against XBT or buoy data for other periods. Recently, Reverdin et al. (1988) have compared two ship datasets with XBT data in the tropical Atlantic. They found that one set was biased cold relative to XBT data by −0.2°C on average. The other set, for almost the same
region, had a bias of +0.11°C. Some influence of different XBT types or samples might be suspected. Systematic differences of 0.2°C have been reported by Heimiller et al. (1983) and Tabata (1978b). The USSR and FNOC analyses are only available for three Januaries and three Julys in the Northern Hemisphere, so XBT samples are too small to draw reliable conclusions.

Figure 7 shows the time-varying character of the biases for the Northern Hemisphere. Biases for the UKMO, CAC in situ, and blended analyses appear to show some persistence between adjacent months. The character of the persistence is different for the satellite data; these data are biased cold in 1982 and into 1983, but are in good agreement with the XBT data from mid-1983 to the end of 1984, with near-random differences in the latter period. Similar results are seen for the extratropical North Pacific (not shown). There is a distinct suggestion of a seasonal variation of the in situ biases. This is investigated hereafter in more detail.

7. Differences between in situ bucket and other in situ data

An investigation has been made of the causes and character of in situ SST biases, including possible seasonal biases. During the period of interest (1982–1984), most in situ surface marine observations were

| Table 3. Average monthly number of quality controlled XBT comparisons. |
|--------------------------|----------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| Region                  | Globe    | NH     | SH     | ENA    | ENP    | TNA    | TNI    | TNP    | ECP    | TWP    | TSI    | TSP    | MSH    |
| n                      | 1464     | 1309   | 155    | 414    | 535    | 99     | 43     | 195    | 20     | 28     | 22     | 67     | 41     |
made from ships. Almost all the observations can be divided into two classes: those derived from water sampled in buckets, and those derived from engine cooling water intake measurements (WMO 1988). To help determine differences between “bucket” and “nonbucket” observations, use was made of an unpublished earlier study of these data for January 1975–December 1981, which we carried out using logbook values for which instrument type indicators were available. This gave a total of about $2 \times 10^6$ bucket and $4 \times 10^6$ nonbucket data. Although the mix of instrumentation for 1975–1981 is likely to have differed a little from that for 1982–1984, the changes are unlikely to be large (earlier editions of WMO 1988), while the characteristic differences between bucket and nonbucket data should be similar. A “bucket” and a “nonbucket” analysis were produced using the UKMO 5° box analysis scheme. Further details are given in appendix A.

Differences between the two analyses are summarized in Fig. 8. This shows zonally averaged differences (buckets minus nonbuckets) averaged over the six-year period for the three-month seasons December–February, March–May, etc. Despite some uncertainties (appendix A), the pattern of results shown in Fig. 8 is physically reasonable. Bucket SSTs were colder on average than the nonbucket SSTs by 0.08°C, a smaller difference than that indicated by Wright (1986), James and Fox (1972), Walden (1966), or Saur (1963). Most of the buckets currently used, however, are believed to be insulated types, in view of the reduced seasonal cycles in SST compared with pre-1940s uninsulated bucket data, and most U.K.-selected ships now use an insulated black rubber bucket (Folland and Parker 1990). Previous authors are not explicit, but it is likely that in older studies, many, but probably not all, buckets tested were uninsulated. These were often single-walled canvas types, which cool more quickly in many locations when exposed on deck (Bottomley et al. 1990; Folland and Parker 1990). In Northern Hemisphere midlatitudes, however, Fig. 8 still shows a distinct seasonal cycle in differences between “bucket” and “non-

![Figure 7. Monthly time series of biases in SST analyses relative to XBT SSTs (analysis minus XBT) for six analyses for the Northern Hemisphere.](image)

![Figure 8. Seasonal, zonal averages of the difference between “bucket” and “nonbucket” SST values for 1975–81. Values are plotted every 5° latitude, and are based on 5° boxes with >=8 monthly values available in a box.](image)
bucket” values. As a result, bucket SSTs are on average warmer than nonbucket SSTs in Northern Hemisphere summer, and cooler at other times of the year. This may be because buckets tend to sample water above the shallow summer thermocline, while engine cooling water intake thermometers tend to measure within it. North of 50°N, bucket SSTs appeared to be warmer than nonbucket SSTs for all seasons, though least so in winter. Unfortunately, there were insufficient data to examine this effect in southern midlatitudes.

Although Fig. 8 shows only relative differences in SST, the results may help to explain the seasonal variation of bias relative to XBTs suggested by Fig. 7. Unfortunately, as discussed in appendix A, uncertainty exists as to the accuracy of the classification of ships’ SSTs into bucket and nonbucket categories during 1975–1981; this uncertainty is greatest in the North Pacific. Improvements in the procedure for indicating the type of SST measurement were introduced in 1982 for logbook data exchanged under WMO Resolution 35. Sufficient numbers of these more recent data will soon have accumulated for a similar study to be very desirable.

8. Conclusions and recent developments

a. Accuracy of existing analyses

This study shows that three SST analyses—the UKMO, the CAC in situ, and the CAC blend—consistently agreed well in the study period. Averaged over relatively small regions, differences between them occasionally exceeded 1°C in individual months, but over regions as large as a hemisphere, differences rarely exceeded 0.1°C. These differences cannot be considered significant based on the studies using XBT data in section 6. The other analyses investigated—the CAC satellite, FNOC and USSR in situ—did differ when compared with the UKMO, and the CAC in situ and blended analyses, but firm conclusions could only be drawn about the satellite analysis. Over the Northern Hemisphere and tropical Southern Hemisphere, at least, comparisons with XBT data showed that the UKMO, CAC in situ, and CAC blended analyses seem to have only small biases of near 0.1°C, or less. If the XBT data were to be entirely trusted, the small differences show these data to be a little too warm in the Northern Hemisphere. Relative to the XBT and in situ data, the satellite data were clearly much too cold in the first half of the study period in many regions of the world, especially between mid-1982 and mid-1983 in the Northern and Southern Hemisphere tropics, but the biases disappeared late in the period. Although data from the extratropical Southern Hemisphere was least affected, there is a hint that El Chichón aerosols may have briefly reduced satellite SST values in late 1982 (Fig. 1e). Otherwise collocated in situ and satellite data agreed reasonably well here.

A disturbing and unexpected result concerned the apparently similar, statistically significant, average negative biases of all three CAC analyses and the UKMO analysis of about −0.5°C in the midlatitude Southern Hemisphere when compared with XBT data. Although great care was taken to compare XBT and analyzed anomalies colocated in space and time, the result may not be trustworthy, as the total number of comparisons with XBT data in this region was only 1470 for the whole period. An explanation, however, should be sought as soon as practicable from a longer period of data, and a more detailed study made of the character of XBT temperature traces. The types of XBT used and the observational practices associated with their use may also need scrutiny (Tabata 1978a).

The comparisons have only been made in regions where all analyses had sufficient observations to produce monthly mean SST anomalies. Thus the advantages of satellite SST data in data-sparse regions could not be fully demonstrated. The distribution of in situ observations, however, is not sufficient to adequately define the SST field everywhere, especially in the Southern Ocean, the eastern South Pacific, and parts of the equatorial Pacific. In particular, this study provides evidence that in situ data poorly define changes in SST between consecutive months on many space scales, though they more adequately sample interseasonal changes. Therefore, for use in studies of long-range forecasting and climate mechanisms, we believe that a global SST field based on in situ data must, when possible, be augmented with satellite data that have been corrected as far as possible for their biases.

b. Toward an improved analysis

This study suggests that the blending technique used in the CAC blended analysis consistently combines satellite and in situ data. Furthermore, within the limitations set by data availability, the UKMO and CAC in situ analyses provide a sound foundation for the analysis of in situ SST. There are, however, several ways in which these analyses are being improved:

(i) Sea ice limits have been added routinely to the monthly CAC blended analysis beginning in January 1989, though the definition of an “ice limit” can never be exact, given even perfect ice data, because there is usually an extended region where open water and ice floes coexist. The limit currently used corresponds to 50% ice cover. Inclusion of an ice limit affects the blended values of SST outside the ice edge, because SST at the ice edge is fixed at the freezing point of sea water and used as the external boundary condition when blending. Lack of inclusion of an ice limit had caused errors in blended SSTs at grid points between the ice edge and the nearest in situ data. This arose because the external boundary condition previously used was the high-latitude limit of preferably in situ, but sometimes satellite, data so that biases in satellite data sometimes affected the analysis. Incorporation of an ice edge markedly reduces analyzed SST values in
parts of the Southern Ocean north of the edge where in situ SST data are absent. There is a similar, smaller effect south of the Arctic ice edge.

The CAC blended fields from 1982 to 1988 have also been recomputed to include ice data [Reynolds and Marsico (1992) give details]. These fields are further being merged with the operational and historic UKMO SST analyses to create a globally complete, historically consistent database for numerical model simulations. It should be noted that the CAC fields that include ice data were not used in this paper. Because our comparisons were limited to regions with adequate in situ data, however, the use of the improved CAC fields would have very little impact on our results.

We expect that a new blended analysis appropriately incorporating ice limits will be adequate for long-range forecasting in most situations and also for the simulation of monthly to decadal atmospheric variability using climate models. It should also be adequate for studies of observed climate variability and change except perhaps on small space scales and in some equatorial regions of the Pacific.

(ii) The most recent version of the CAC blended analysis uses a lighter spatial smoothing and so is more highly resolved. Ideally a "data adaptive" smoothing and quality control technique is needed to smooth little where data are plentiful and more where they are not.

(iii) The CAC in situ analysis has a higher spatial resolution (2°) than the UKMO analysis (5°), but as a result uses fewer SST data per box. In most regions, the spatial coherence of SST anomalies is sufficiently high that the UKMO technique allows more effective quality control of the within-box SST data without excessive smoothing. In some regions, however, a 5° box is too coarse, for example, near the equator in the tropical eastern Pacific, especially during an anti–El Niño (van Woert 1990). Thus, a technique is needed that combines the advantages of the larger spatial data scan of the UKMO analysis with the finer intrinsic resolution of the CAC in situ analysis.

(iv) Insufficient use is made of the temporal coherence of SST anomalies in all the analyses. Existing techniques can cause sudden localized changes in apparent SST anomalies in regions with few in situ data. This is especially evident when the analysis is made daily using a rolling 15-day or shorter period. Sudden local SST changes tend to be observed when an in situ data value, relatively isolated in space and time, is either lost or gained as the analysis time window moves forward. A jump will occur (i) if the satellite data in this region are appreciably biased and are suddenly adjusted, or no longer adjusted, by the in situ data value, or (ii) if the satellite data are not biased, but the in situ data value is erroneous or unrepresentative. A method of temporal smoothing is required.

To improve the spatial and temporal resolution of the CAC analysis, a 1° resolution optimum interpolation (OI) analysis (Gandin 1963; Leetmaa and Ji 1988) has recently been developed at the U.S. National Meteorological Center (NMC). This analysis is produced both daily and weekly using the last week of in situ and satellite SST observations and includes sea ice extent. The daily analysis is produced operationally. To reduce the number of observations (especially satellite observations) that are used in the OI, averages over 1° squares are computed. These "super observations" are computed independently for each ship and buoy identification code and for day and night satellite data. Both analyses use the preceding analysis to provide temporal continuity.

The OI method assumes that the data are unbiased. Because, as we have shown, there are biases in satellite SST data, a preliminary step was added before applying the OI to correct any large-scale satellite SST biases. A smoothed version of the CAC blended analysis analyzed over 10° boxes is used to provide separate smooth correction fields for daytime and nighttime satellite data. This bias adjustment was originally only used with the weekly analysis. In July 1991, however, the bias adjustment was added to the daily analysis because large-scale tropical satellite SST biases (often around −1° to −2°C) were caused by the presence of stratospheric aerosols from the June 1991 eruptions of Mt. Pinatubo in the Philippines. Independent buoy data confirmed the presence of the satellite biases and showed a significant improvement in the OI analysis when the satellite bias adjustment was used.

APPENDIX

Comparison of Recent Measurements of Sea Surface Temperature Derived from Buckets and Other Methods

a. Introduction

Separate global datasets of monthly 5° box SST values for 1975 to 1981 were created based on (a) ships' observations that were flagged as being derived from "buckets" or from unknown instruments, and (b) unflagged ships' observations that were assumed not to be derived from buckets. The flags were available in the U.K. Meteorological Office Main Marine Data Bank (MOMMDB) for data exchanged under WMO Resolution 35 (Shearman 1983). Unflagged data included hull sensor and other specialized data sources as well as engine-intake data. Scrutiny of the periodical publication WMO No. 47 (WMO 1988) indicates that the great majority of unflagged data originated from engine intakes. WMO No. 47 gives the reported method of measurement used on each selected supplementary and auxiliary ship in the world voluntary observing fleet. Flags set before 1975 appear to be unreliable, as there is a sudden, very large and persistent change in the ratio of bucket to nonbucket observations in the MOMMDB between December 1974 and January 1975. The datasets used here were created in 1983. In these datasets, the ratio of the numbers of nonbucket to all observations is about 0.6 in 1975 and 1976, de-
creasing slowly thereafter. The decrease may be because the most complete data after 1976 came from U.K. ships, most of which use insulated double-walled rubber buckets. Bucket and nonbucket datasets were created separately and quality controlled as described by Bottomley et al. (1990).

b. Procedure

The datasets were analyzed for each season (December–February, March–May, June–August, September–November) as follows:

(i) For a given 5° box and season, only months with data available for both instrumental classes were analyzed, to avoid sampling error. If the box contained fewer than 8 such months out of the 21 months possible in a given season (7 years × 3 months), it was not analyzed. Individual months’ values could be based on a single data value in view of the temporary persistence of SST anomalies (Bottomley et al. 1990).

(ii) Means, \(u\), and standard deviations, \(s\), of monthly differences between bucket and nonbucket SST were computed for each eligible 5° box.

(iii) The \(t\) variate, \(t = u/(s/n^{0.5})\), was computed where \(n\) was the number of months available (8 < \(n\) < 21); this \(t\) test compares \((u, s)\) with a value of zero and the significance of \(u\) is estimated in each available 5° box. A test comparing \((u_{\text{bucket}}, s_{\text{bucket}})\) with \((u_{\text{nonbucket}}, s_{\text{nonbucket}})\) was not considered valid because the bucket and nonbucket data contained common variance due to meteorological effects, rendering \(t\) artificially small.

c. Results

Zonal means of \(u\) are shown in Fig. 8. Salient features are as follows:

(i) Bucket data were on average nearly 0.1°C colder than nonbucket data. Average differences were -0.11°, -0.06°, -0.08°, and -0.08°C for 60°N to 40°N for December–February, March–May, June–August, and September–November, respectively.

(ii) The relative coldness of bucket data was greatest in winter in low midlatitudes with statistically significant differences mainly in the North Atlantic and the Kuroshio.

(iii) Buckets were warmer than nonbuckets in higher midlatitudes of the Northern Hemisphere, particularly in summer, but to some extent in winter. Statistically significant differences were largely confined to the Gulf of Alaska.

(iv) In equatorial regions, buckets were on average about 0.05°C colder than nonbuckets; differences were only statistically significant in the Atlantic.

Buckets also tended to be warmer than nonbuckets in the eastern half of the Pacific (maps not shown). Statistically significant differences were mainly confined to the southern tropics.

d. Discussion

The small average differences (relative to those, for instance, found in Barnett 1984) suggest that the buckets used to measure SST in 1975–81 were largely insulated. Suggested reasons why buckets still indicate lower SST overall than the combination of engine-intake and other methods include:

(i) Some heating due to the ship’s engine, etc., will affect the dominant nonbucket sensor, the engine cooling water intake thermometer.

(ii) A small amount of evaporation and sensible-heat transfer, with associated heat losses, may occur from the exposed top of the water sample in the insulated bucket when on deck during a measurement.

On the other hand, Fig. 8 indicates that bucket temperatures tended to be warmer in midlatitude summer. Insulated buckets could read higher if the surface ocean layers sampled were warmer than the deeper layers sampled by engine intakes or hull sensors. The sampling by buckets of shallow, warm, water surface layers would be especially frequent in late spring and early summer at a time when engine intakes may be placed within a thin seasonal thermocline, or even below it, in modern ships with a large draught. Bucket temperatures may also be higher by day due to a thin diurnal thermocline, and buckets (some of which are black) may warm in strong insolation while on deck before a reading is taken.

The relative warmth of bucket data north of 50°N in winter is more difficult to explain. It is particularly marked in the Gulf of Alaska, where data were plentiful and were largely derived from Japanese, American, and a few Canadian ships. According to flags set in the MOMMDB, the American data largely came from engine cooling water intake temperature measurements, while the Canadian data mostly came from buckets or unknown instrumentation. These indications agree with those in WMO No. 47. MOMMDB flags classify the Japanese data as deriving from buckets or unknown instrumentation, in contrast to WMO No. 47, which indicates that 90% of Japanese ships used engine cooling water intakes. The Japanese (and the relatively few Canadian) data appear to be warmer than the U.S. data. If most of the plentiful Japanese data are really derived from cooling water intakes, the relative warmth of “bucket” data in winter in the Gulf of Alaska has been exaggerated in our analysis.

Finally, inclusion of “unknown” instrumentation in the “bucket” class when they may really be engine intakes implies that Fig. 8 may give differences between buckets and nonbuckets that are generally too small. The ratio of bucket-to-nonbucket observations in 1975 and 1976, when data were most complete, should be at least 7:3 according to WMO No. 47, but is indicated to be about 6:4 in the MOMMDB. Thus, perhaps 25% of “bucket” data are really derived from “nonbuckets.”
If so, a more realistic global annual mean difference between non-bucket- and bucket-derived SST data in 1975–1981 is 0.11°C.

REFERENCES


