A Global Analysis of Sea Surface Temperature for Numerical Weather Prediction

BRUCE BRASNETT

Canadian Meteorological Centre, Dorval, Quebec, Canada

(Manuscript received 21 February 1996, in final form 2 December 1996)

ABSTRACT

A global analysis of in situ observations of sea surface temperature (SST) developed for use at the Canadian Meteorological Centre is described. The analysis is done on the anomaly, the departure from climatology. The anomaly plus climatology, or resulting SST, is used as the lower boundary condition by the numerical weather prediction model. Since there is no ocean model to provide a background or first-guess field for the analysis, and since anomalies are observed to persist over long periods, the background field is obtained essentially by assuming persistence of the previous anomaly. The analysis algorithm is statistical interpolation. Attention is focused on techniques to control the quality of the observations, including a technique to remove systematic errors from ship observations. The analysis resolution is 0.9° and the correlation e-folding distance is 212 km. Verification of the analysis is presented using independent data from buoys and expendable bathythermographs for a one-year period. Verification is also presented for the National Centers for Environmental Prediction (NCEP, Washington) weekly analysis and for climatology. Results indicate that the analysis has skill over climatology in all regions and skill over the NCEP weekly analysis in the North Atlantic. In the rest of the Northern Hemisphere, analysis error estimates for the two analyses are similar, while in the Southern Hemisphere the NCEP analysis is superior, probably due to its use of satellite data. It is intended that this analysis will be an essential component of a debiasing algorithm for satellite SST observations.

1. Introduction

An accurate sea surface temperature (SST) field has long been known to play an important role in the production of numerical weather forecasts. Danard (1986) and Nuss (1989), for example, demonstrated that the SST distribution can have a significant impact on cyclogenesis in numerical simulations. As model resolutions improve, there is optimism that a variety of sub-synoptic-scale phenomena such as polar lows (Businger and Baik 1991), low-level jets, coastal fronts, and precipitation (Doyle and Warner 1993) will be accurately simulated in operational forecasts. From the latter study, it is clear that the ability of models to accurately forecast these phenomena will depend, in part, on the definition of the SST field.

The use of an SST analysis in numerical weather prediction (NWP) imposes requirements that differ in some ways from other applications such as climate studies or dynamic extended-range forecasting. This is because NWP has traditionally been viewed as an initial value problem, requiring that initial conditions be updated frequently and available without undue delay. In this context, models are integrated one or more times per day, typically with an SST field that is held constant throughout the integration. Dynamic extended-range forecasting models, on the other hand, may be integrated once or twice per month, making it possible to employ a different strategy, such as using data collected over one month to produce an analysis of monthly mean SST (Reynolds 1988).

Another issue that sets apart the problem of analysis for NWP is that of analysis resolution, which is linked to the question of model resolution. The six operational SST analyses for long-range forecasting that were compared by Folland et al. (1993) have resolutions of between 2° and 5°. By contrast, surface fluxes are currently calculated on a 0.9° grid in the Canadian operational global model (Ritchie and Beaudoin 1994). It is desirable to maintain consistency between analysis resolution and model resolution. Moreover, as the resolution of the NWP models improves, an increasing number of relatively small water bodies are resolved by the model. In the context of forecasting for Canadian territory, it is imperative that the SST analysis provide at least as much information as climatology for the Great Lakes, Lake Winnipeg, Great Slave Lake, and a host of smaller lakes. Unfortunately, no observations of water temperature are available for the smaller lakes. Data voids are a problem not just for lakes but over remote regions of the world’s oceans as well. One means of dealing with the paucity of data is to use an analysis of SST in regions where observations are available and to use climatology else-
where. The analysis described in this paper accomplishes this by producing an analysis of the anomaly from climatology. The SST required for model integration, referred to hereafter as the resulting SST, is then computed by adding the climatology to the anomaly analysis. This strategy ensures that where data are never available the anomaly is zero, and therefore the model uses the SST climatology locally.

In this paper, an objective analysis of SST that was developed for use in NWP at the Canadian Meteorological Centre (CMC) is described. The next section provides details of the analysis algorithm, in section 3 the strategy for controlling the quality of the observations is described, section 4 presents results, and there is a discussion in section 5.

2. The analysis algorithm

The analysis uses all data received from ships, moored buoys, and drifting buoys received from the Global Telecommunications System (GTS). The analysis is updated every 24 h using data from a 24-h time window, with a data cutoff time varying from 26 h after observation time for 0600 UTC data to 8 h after observation time for 0000 UTC data.

Climatology is interpolated bilinearly to the observations and observed anomalies are calculated. The climatology used is that of Parker et al. (1995), which is available both as monthly means and daily means on a 1° latitude–longitude grid. Here the monthly means are used, supplemented by values for lakes that were obtained from a variety of sources and added to the basic climatology. The bilinear interpolation scheme uses the values at the four grid points nearest the observation. If the observation is near the coast, some of these grid points may be over land where the climatology is undefined. Hence, the SST climatology, φ, is defined for land points using the method of sequential overrelaxation (see Thompson 1961) to solve ∇2φ = 0, subject to the internal boundary conditions imposed by the climatological values surrounding the landmasses or at lakes. The 12 monthly climatologies, processed in this way, are then interpolated linearly to the analysis time prior to use.

The background field is the most recent anomaly, modified slightly to provide a gradual return to climatology. The latter is desirable because during the fall and winter months, when ice forms over some ocean regions and lakes, one cannot assume that the anomalies persist. Since no observations are available for these regions until the ice melts, it is prudent to return the anomalies to zero. Therefore, prior to each analysis, the background field is multiplied by 0.99. This is equivalent to an exponential decay of the anomalies with an e-folding time of 100 days, which is consistent with the observed decay times for anomalies (Reynolds 1978). Ideally, wherever there is a high concentration of sea ice, one would prefer that the resulting SST be −1.8°C, the freezing point of seawater, with a salinity of 33 psu (practical salinity units). The method described here does not guarantee that this will be the case. However, the SST values at these grid points are of no consequence to NWP forecasts since the ice field used by the model is specified independently, and where there is ice cover the model makes no use of SST.

The method of statistical interpolation, commonly referred to as optimum interpolation or OI, is applied to correct the background field using the observations. The basic equations of the method have appeared many times in the literature and will not be repeated here [see, e.g., Rutherford (1972) or Lorenc (1981) and, in the context of SST, Clancy et al. (1990) or Reynolds and Smith (1994)]. The first step is to interpolate the background to the observation locations using a bilinear interpolation. The background at each observation location is then subtracted from the corresponding observed anomaly to give the data increments. The OI method is then applied to interpolate the data increments to the grid points of the analysis grid, a global 0.9° latitude–longitude grid. At the end of the procedure, these analysis increments are added to the background to give the updated anomaly.

Computational constraints dictate that not all observations can be used in producing an analysis at each grid point. Accordingly, only observations within a radius of 800 km of each grid point are used, up to a maximum of 24. If there are more than 24 reports within this radius, the 24 reports closest to the grid point are selected.

Variances of the observation error and the background error must be specified in the statistical interpolation formalism. Analysis quality, as measured by independent observations (see section 4), was found to be very sensitive to the ratio of these two error variances. In principle, one should be able to calculate the variances through use of a standard statistical technique carried out on a large ensemble of observed-minus-background differences (see, e.g., Mitchell et al. 1990). Such an approach was attempted with a previous version of the analysis using ship data but the results were found to be unreliable. Reynolds and Smith (1994) reported a similar finding. As a result, the observation and background error variances used here were arrived at through experimentation. Ship observations are assigned an observational error variance of 0.72 K2, and moored and drifting buoys are given an observational error variance of 0.36 K2. Appendix A discusses the specification of these statistics.

Verification of the analysis (see section 4) strongly suggests that the analysis error (and therefore the background error) depends upon the rate of change of the local SST. Anomalies in regions with rapidly changing SST or active circulation systems are not likely to be as persistent as anomalies elsewhere. Such regions are necessarily associated with relatively large standard deviations of the climatological mean SST. Therefore, the
Fig. 1. (a) Background error variance computed from Eq. (1) for 1 February. (b) As in (a) but for 1 August.

background error is parameterized using the SST standard deviations of the *U.S. Navy Marine Climatic Atlas of the World* (Naval Oceanography Command Detachment 1981). The 12 monthly standard deviations on a 1° grid are first interpolated linearly to the analysis date and then transformed to spectral coefficients so that a triangular truncation can be applied at wavenumber 45. This yields a result smooth enough to be used in the optimum interpolation algorithm. The smoothed standard deviations, $E_c$, are then used to compute the background error, $E_b$, at each grid point of the analysis grid according to

$$E_b^2 = \gamma + \delta E_c^2.$$  \hspace{1cm} (1)

Here $\gamma$ and $\delta$ are arbitrary constants taken to be 0.14 K$^2$ and 0.03, respectively. Examples of the background error variance are shown in Fig. 1. Strictly speaking, $E_b$ depends on the day-to-day variability of SST and not on its interannual variability. Nevertheless, use of the climatological standard deviations in this way provides a realistic seasonal trend in the background error, which is clearly seen by comparing Fig. 1a with Fig. 1b. Moreover, Fig. 1 has many of the same features as the guess error standard deviations (Fig. 11) of Reynolds and Smith (1994). It turns out that (1) gives a background error variance bounded by 0.16 K$^2$ and 0.57 K$^2$, but if one examines a typical grid point with an error variance of 0.18 K$^2$, then a single, isolated ship observation located exactly at the grid point receives a weight of 1/5, while the background receives a weight of 4/5. The corresponding weights for buoys would be 1/3 and 2/3, respectively.

Background error correlations between observation locations and grid points and between the observation locations and grid points.
locations themselves are computed with a sum of third-order autoregressive functions used by Mitchell et al. (1990). This function fits air temperature well and it is for this reason that it was tried for SST. The correlation between points $i$ and $j$ is given by

$$
\alpha_{ij} = \frac{5}{6} \left[ 1 + cr + \frac{1}{3} (cr)^2 \right] e^{-cr} \\
+ \frac{1}{6} \left[ 1 + \frac{1}{3} cr + \frac{1}{27} (cr)^2 \right] e^{-cr^2}.
$$

Isotropy is assumed, so the correlation depends only on the distance $r$ separating the two points. The length scale $c$ is 0.016 km$^{-1}$. Figure 2 shows the shape of the correlation function. The more common choice for correlation function is the negative squared exponential function (see, e.g., Clancy et al. 1990 or Reynolds and Smith 1994), which is plotted for comparison as the dashed line in Fig. 2. The $e$-folding distance for both functions is 212 km. The length scale is a global constant, the choice of which is discussed in appendix A. One potential improvement in the current scheme would be to follow the approach of Clancy et al. (1990) and Reynolds and Smith (1994), who allow an anisotropic correlation of the background error as well as correlation length scales that vary with geographical position. It should be noted, however, that while it is reasonable to adopt an anisotropic model for background error if the background is SST, it is not clear that this is the best approach if the background field is anomaly from climatology.

3. Quality control and superobservation formation

Much of the time needed for each analysis is spent not on the gridpoint analysis but on the preliminary steps of eliminating redundancy in the observations and their quality control. The first of these steps is the application of a debiasing algorithm for ship reports. The premise is that instruments used to measure SST aboard ships of the Voluntary Observing Program are not regularly recalibrated, with the result that a given ship may unknowingly transmit SST observations with a significant and nearly constant bias for long periods. Once the ship reports are retrieved from the database in preparation for the analysis, an attempt is made to remove the biases of selected ships using previous data from the same ship and corresponding analyzed SST values. Details of the algorithm may be found in appendix B. Typically, 20% of the observations from ships are modified in this way. Section 4 explores the question of the effect of this procedure on analysis skill.

All reports must first pass several quality control tests prior to the analysis. Reported positions are checked against a high-resolution land–sea mask, and reports that are not over water are discarded. This normally rejects 1%–2% of the ship reports and a small number of drifter reports. The positions reported by drifters are also checked if more than one report is available. If the drifter positions indicate a displacement of more than 100 km in 24 h, the reports from that drifter are discarded.

The elimination of redundant data begins with a scan of the dataset for duplicate ship reports, including those having different call signs but identical reports in all other respects. Once the duplicates have been removed, the superobservation algorithm of Lorenc (1981) is applied to the ship and fixed-buoy data separately. Where two or more observations fall within 10 km of one another, they are combined into a single observation. The observation errors may be reduced by the procedure. If other, they are combined into a single observation. The availability of redundant data makes it possible to check the grouped data for consistency. This check is performed pairwise on all pairs possible and requires that the square of the differences between the two observations be less than the sum of their observational error variances. Observations failing this check are excluded from the group forming the superobservation and may therefore be rejected individually by the final quality control procedure.

In addition to combining the observations themselves, observation errors may be reduced by the procedure. If two platforms are involved, the superobservation has an error variance equal to one-half that of a regular observation, and for three or more platforms the error variance is one-third that of a regular observation. In cases where all observations in a group originate from the same platform, the observations are not truly independent, and the result would be if a different instrument had been used to make each measurement. In these cases, the observation error of the superobservation is equal to that of a regular observation. This procedure eliminates multiple, collocated observations from the dataset used in the analysis without ascribing an unrealistically low error to the superobservations. The superobservation procedure affects 15%–20% of the ship reports and all of the moored buoy reports.
Drifter reports are also combined since large numbers of reports (up to 50 in some cases) may be available daily from a given drifter. Reports are grouped according to drifter identifier, not according to distance. Seeking observations that are distributed as uniformly as possible over the 24-h period, only one report per 3-h interval per platform is retained. This procedure yields from one to eight reports per platform. When there are two or more reports in a group, they are combined using the superobservation algorithm described above. The observations from approximately 85% of the drifting buoys are grouped and averaged in this way.

The final and most comprehensive check is applied only to observations and superobservations that have passed the basic position checks. An observation is rejected if it satisfies

\[(A - O)^2 > T^2[E_o^2 + E_i^2],\]

where \(A\) is the OI interpolated SST anomaly valid at the location of this observation and obtained by using only neighboring observations (if any), \(O\) is the observed anomaly being checked, \(E_o\) is its observational error, \(E_i\) is the error of the interpolated value \(A\) derived from interpolation theory, and \(T\) is a tolerance. See Lorénc (1981, section 3c) for details. Tolerance \(T\) is chosen to be 2.8 for ships and 2.2 for buoys. Observations that initially satisfy (3) are flagged as doubtful and checked again, this time using no doubtful observations in the computation of \(A\). Observations that still satisfy (3) during the second check are rejected. For an isolated observation, this test reduces to a check against the background. In a region where the background error variance is 0.18 K^2, ship observations differing from the background by more than 2.7 K are discarded by the procedure. For buoys, however, isolated observations differing by more than about 1.65 K are rejected. On average, this check rejects 8.9% of ship reports and 9.4% of the reports from drifting and fixed buoys. Compared with other observed parameters, these criteria are exacting. Ship and buoy reports of atmospheric pressure, for example, are rejected less than one-half as often as SST by the operational quality control procedures at CMC.

4. Results

a. Analysis spatial and temporal variation

In this section, some important properties of the analysis are highlighted by examining the spatial and temporal variation of analyzed fields. Figure 3 shows the analysis from 1 March 1995. A large proportion of ocean areas are blank, indicating an anomaly bounded by \(-0.5\) to \(-1.5\) K, dark shading indicates anomalies in the range \(-1.5\) to \(-2.5\) K, and cross-hatching indicates anomalies in the range 1.5 to 2.5 K.
addition, the gradient associated with the northern edge of the Gulf Stream is concentrated in a narrow ribbon in the SST analysis, whereas it is spread over about twice the area in the climatology. Since this is a primary area for cyclogenesis, the consequences of these differences for forecast quality may be quite significant in some cases.

Folland et al. (1993) make the point that frequently updated SST analyses such as the one described here may have difficulty exploiting the high degree of temporal coherence of SST anomalies. Figure 5 is an example of the time evolution of the anomaly at 40 grid points in the central Pacific from 29 December 1994 to 15 February 1995. It illustrates that in the equatorial zone, the positive anomaly associated with the El Niño event persisted through the period, with only a slight decrease in intensity noted, mostly occurring at discrete times that are likely related to the availability of observations. At higher latitudes (near 35°N), however, the anomaly went from about 1.5°C to about −1°C, a significant change, but one that occurred in a smooth manner. Figure 5 is evidence that the analysis method described here is capable of representing both persistent and transient anomalies successfully.

It was stated in the introduction that the resulting SST field must provide the NWP model with suitable lake surface temperatures. Given the large number of lakes within Canadian territory, ensuring that the analyzed water temperatures for lakes are reasonable is an abiding concern. For many lakes, both good quality climatological information and real-time in situ data are scarce. During 1995, for example, the only observations of water temperature received for Great Slave Lake were from a buoy moored at 61.1°N, 115.3°W. The buoy was deployed in July and removed from service in October. Figure 6 shows a time series of the data from the buoy at 6-h intervals, along with the resulting SST interpolated to the buoy location (solid line) and climatology (an estimate based on buoy data from previous years, dashed line). The analysis is generally faithful to the trends in the data while smoothing out the erratic part of the signal. An important role of an objective analysis is to filter noise in the observations (Hollingsworth 1987). Here the filtering is accomplished by combining the observations over the 24-h period into a single superobservation as discussed in section 3, discarding outliers, and then weighting observation and background appropriately to produce an analysis at gridpoints. The resulting analysis necessarily exhibits a time lag in its response to fluctuations in the data, which is clearly seen in Fig. 6. When the buoy is removed from service.
on 10 October, the analysis follows the climatological trend with an additional small but perceptible drift toward climatology as described in section 2.

b. Verification

To assess analysis quality, special analyses, wherein all moored and drifting buoy data were withheld, were produced for a 1-yr period. The withheld drifter reports were then used as independent data to estimate analysis error, as was done by Reynolds (1988). Since drifter reports can be infrequent in the vicinity of the Gulf Stream and the Kuroshio, data from nine of the withheld moored buoys in these regions were also used in the verification study. Where possible, the data from each platform was averaged into a single observation for each 24-h period using the superobservation procedure discussed in section 3. Reports that were rejected by the quality control steps of section 3 were not used in the study. To facilitate the verification, the analysis valid the same day as the observation was interpolated to the coordinates of the observation. As a point of comparison, special weekly OI analyses from NCEP (National Centers for Environmental Prediction, Washington) were obtained. These analyses, prepared as described in Reynolds and Smith (1994) but without buoy data, were filtered in time with a binomial filter using weights 1/4, 1/2, 1/4 and then linearly interpolated in time to generate daily analyses, which were then verified against the same buoy data as above. The filtering was applied because it gave slightly better verification results. Finally, the 12 monthly climatologies were also interpolated temporally to the date and spatially to the observation locations to provide a benchmark for the study.

Table 1 gives the definitions for the geographical regions and their acronyms referred to in the verification. Figure 7 shows the geographical distribution of the

---

**Table 1. Definition of ocean regions and their acronyms.**

<table>
<thead>
<tr>
<th>Region</th>
<th>Acronym</th>
</tr>
</thead>
<tbody>
<tr>
<td>Northern Hemisphere</td>
<td>NH</td>
</tr>
<tr>
<td>Southern Hemisphere</td>
<td>SH</td>
</tr>
<tr>
<td>Western North Pacific</td>
<td>WNP</td>
</tr>
<tr>
<td>Eastern North Pacific</td>
<td>ENP</td>
</tr>
<tr>
<td>Western North Atlantic</td>
<td>WNA</td>
</tr>
<tr>
<td>Eastern North Atlantic</td>
<td>ENA</td>
</tr>
<tr>
<td>Tropical North Pacific</td>
<td>TNP</td>
</tr>
<tr>
<td>Tropical North Atlantic</td>
<td>TNA</td>
</tr>
<tr>
<td>Tropical Southern Hemisphere</td>
<td>TSH</td>
</tr>
<tr>
<td>Midlatitude Southern Hemisphere</td>
<td>MSH</td>
</tr>
</tbody>
</table>

---

**Fig. 6.** Time series of resulting SST (solid line), climatology (dashed line), and moored buoy observations (dots) at 61.1°N, 115.3°W (Great Slave Lake) from 1 August 1995 to 14 November 1995. The buoy was the only source of in situ data available during the period.
Southern Hemisphere where ship reports are sparse, the improvement is a more modest 10%. It is also clear from examination of Table 2 that where the climatological variance is small (e.g., Tropics and eastern North Atlantic) the analysis errors tend to be small, and where the climatological variance is large (e.g., near the Gulf Stream or Kuroshio) the analysis errors tend to be large. In the latter regions, the western boundary currents give rise to large temperature gradients and transient eddies. The analysis errors in these regions are partly due to temporal and spatial variability on scales resolved by neither the analysis algorithm nor the observations employed in the analysis.

The magnitude of the errors in Table 2 are sensitive to the quality control performed on the independent data. The quality control measures described in section 3 are not independent of the analysis since the trial field used in the observation check is the previous analysis. One might argue that this fact favors the CMC analysis in the comparison. Moreover, Table 2 leaves unanswered questions about the contribution of the buoy data to analysis quality, especially since buoy data are of high quality and they outnumber ship reports in many regions. Indeed, in the tropical Pacific, the spatial density of the withheld observations is normally much greater than that of the ship data due in part to the array of buoys deployed as part of the TOGA–TAO program (Hayes et al. 1991). Near the Gulf Stream, important differences were occasionally noted between analyses produced using all the buoy data and the special analysis during the study, differences that must be attributed to the withheld observations.

In order to validate the results in Table 2 and address the issues raised above, the analysis that includes all the buoy data was verified against data from expendable bathythermographs (XBT) obtained from the GTS. XBT observations are not currently used in the analysis because only a handful are available at the data cutoff time and because they provide a convenient source of independent data for real-time monitoring of analysis quality. To choose a temperature at an appropriate depth...
while minimizing the impact of spurious data that can result from the transient when the probe hits the water, the algorithm of Folland et al. (1993) was employed. The second most shallow temperature was selected if there were two or more temperatures reported in the first 10 m (over 80% of the cases), otherwise the temperature nearest the surface was chosen. Some rudimentary quality control is necessary before these data can be used in a verification study. The quality control consisted of two checks: the position of each report was verified to ensure it is over water, and the selected temperature was compared with the Parker et al. (1995) climatology to ensure it was within two standard deviations of climatology. Again, the SST standard deviations of the Naval Oceanography Command Detachment (1981) were used. If an XBT report contained no temperature within 10 m of the surface, it was discarded.

Figure 8 shows the geographical distribution of the XBT data, and Table 3 shows the results of the verification. The XBTs give a 20% larger error estimate for the Northern Hemisphere than the buoy data, evidence that checking them against climatology is less than ideal (see the discussion in appendix A). In general, however, the results of Table 3 share some of the characteristics of the buoy verification. For example, the NCEP analysis has the smallest error estimate for the Southern Hemisphere, the TNA, and the ENP regions, while the CMC analysis has the smallest error in the extratropical North Atlantic. The discrepancies between the two results are in the WNP and TNP regions, where the XBT verification indicates that the CMC analysis has a smaller error than the NCEP analysis, while the opposite is true in the buoy verification. Closer examination of the XBT verification also reveals that the CMC and NCEP errors are much closer in the ENA and MSH regions. It turns out that these discrepancies are an artifact of the geographical distribution of the data. XBT reports tend to be available mostly along shipping routes (Fig. 8). In the data-sparse Southern Ocean, this fact skews the results since the regions devoid of ship data are not

![Figure 8](https://example.com/figure8.png)

**Fig. 8.** Distribution of SST reports from expendable bathythermograph (XBT) reports used to verify analysis quality, with verification regions outlined. The period covered is 1 January–31 December 1995.

**Table 3.** The rms errors (°C) from the verification against XBT data by ocean region and approximate 95% confidence limits.

<table>
<thead>
<tr>
<th>Region</th>
<th>Sample</th>
<th>CMC daily analysis</th>
<th>NCEP weekly analysis</th>
<th>Climatology</th>
</tr>
</thead>
<tbody>
<tr>
<td>NH</td>
<td>13 719</td>
<td>0.92 ± 0.01</td>
<td>0.99 ± 0.01</td>
<td>1.18 ± 0.01</td>
</tr>
<tr>
<td>WNP</td>
<td>2656</td>
<td>1.28 ± 0.03</td>
<td>1.43 ± 0.04</td>
<td>1.56 ± 0.04</td>
</tr>
<tr>
<td>ENP</td>
<td>1294</td>
<td>0.68 ± 0.03</td>
<td>0.61 ± 0.02</td>
<td>0.89 ± 0.03</td>
</tr>
<tr>
<td>WNA</td>
<td>683</td>
<td>1.52 ± 0.08</td>
<td>1.79 ± 0.10</td>
<td>2.23 ± 0.12</td>
</tr>
<tr>
<td>ENA</td>
<td>2768</td>
<td>0.84 ± 0.02</td>
<td>0.88 ± 0.02</td>
<td>1.13 ± 0.03</td>
</tr>
<tr>
<td>TNP</td>
<td>3436</td>
<td>0.77 ± 0.02</td>
<td>0.78 ± 0.02</td>
<td>0.95 ± 0.02</td>
</tr>
<tr>
<td>TNA</td>
<td>1239</td>
<td>0.52 ± 0.02</td>
<td>0.50 ± 0.02</td>
<td>0.71 ± 0.03</td>
</tr>
<tr>
<td>SH</td>
<td>5389</td>
<td>0.62 ± 0.01</td>
<td>0.58 ± 0.01</td>
<td>0.77 ± 0.02</td>
</tr>
<tr>
<td>TSH</td>
<td>3637</td>
<td>0.56 ± 0.01</td>
<td>0.50 ± 0.01</td>
<td>0.69 ± 0.02</td>
</tr>
<tr>
<td>MSH</td>
<td>1752</td>
<td>0.74 ± 0.02</td>
<td>0.72 ± 0.02</td>
<td>0.90 ± 0.03</td>
</tr>
</tbody>
</table>
sampled. In the ENA region, two-thirds of the data comes from the southern part of the region where the NCEP error is close to the CMC error (cf. TNA region). It is possible to confirm that the cause of the discrepancies between Tables 2 and 3 is the spatial distribution of the XBT data and not other factors (such as the inclusion of buoy data). This is done simply by verifying the analyses of Table 2 with the data of Table 3. The results (not shown) are very similar to those of Table 3, with the rankings the same in all regions. This implies that the spatial distribution of the XBTs, which is generally less uniform than that of the drifter data, is a factor that must be considered when interpreting these analysis error estimates. The true analysis error cannot be known. Nevertheless, the statistics presented here are helpful in gauging the relative quality of the three SST fields.

The buoy and XBT data were also used to estimate analysis biases. These were generally small for all regions (not shown). For biases calculated with respect to the buoy data, both the CMC and NCEP analyses exhibit a slight warm bias of less than 0.1°C for most regions. For XBT data, however, the analyses show larger warm biases of near 0.2°C in most regions. This discrepancy can be attributed to the use of the temperature at the second most shallow depth of the XBT, as biases were found to be in agreement with those calculated with buoy reports if one uses the temperature nearest the surface. Folland et al. (1993) also computed analysis biases relative to XBT reports. One curious finding from their study, which spanned the period January 1982–December 1984, was the presence of an apparent large cold bias (~0.5°C) in the MSH region, which suggested a problem with the XBT data there. There is no evidence of a similar problem with the 1995 XBT data. The biases in the MSH region are 0.15°C for the CMC analysis and 0.14°C for the NCEP analysis. Observing practices appear to have changed significantly since 1984, when only 25% of reports contained two or more temperatures in the first 10 m, as opposed to 1995 when the corresponding number was 80%.

Independent data can help to estimate the impact of some of the strategies adopted for use in this analysis. Table 4 quantifies the effect of the 24-h time window for observations, the contribution made by the ship debiasing procedure, and the importance of analysis resolution. As was the case for Table 2, the independent data come from drifters and a few moored buoys.

In the first test analysis of Table 4, the role of a 7-day time window for observations is explored. This is the strategy of the NCEP weekly analysis, which uses observations from Sunday to Saturday and is valid on Wednesday (R. W. Reynolds 1996, personal communication). It is worthwhile to assess the benefit of using data from up to three days after the analysis valid time. To approximate this strategy, the CMC daily analysis was produced as before, but the anomaly fields from seven consecutive days were averaged and the result was verified against the data from the fourth day of the period. This is an analog of the NCEP analysis. Both approaches lag real time by 3.5 days and both behave as temporal filters, yielding daily analyses that evolve smoothly. The results in the table show a modest gain from this approach, at least in the Northern Hemisphere. Such an analysis could be of value in climate studies, where timeliness is less of a concern. In the NWP context, however, there has traditionally been an emphasis on the timeliness of initial conditions, mainly because forecast quality decreases markedly with increasing lead time. Clearly, the 3.5-day lag cannot be reconciled with this concern. In addition, a temporal filter will reduce the amplitudes of short-lived anomalies. The latter may occur over lakes, along ocean fronts such as the Gulf Stream, and in the wake of hurricanes. It is the goal of this analysis to analyze these transients as faithfully as the data permit. For these reasons, this strategy was not adopted.

In order to examine the effect of debiasing ship reports, a special analysis was produced for the period of the study wherein the SST observations from ships were used in their original, unmodified form. Table 4 contains the results by region. It is clear that the removal of the apparent bias from ship reports is beneficial, accounting for a decrease of 0.04 K or about 5% in the analysis error for the Northern Hemisphere. This decrease is significant at the 95% level of confidence. The ENP region is most affected by the ship debiasing procedure, reinforcing the finding of Trenberth et al. (1992) that SST reports from ships in the North Pacific are of poorer quality than others. The bias removal technique has a positive impact in each of the regions tabulated, although the differences are not significant at the 95% level of confidence in every region.

Finally, since there is a slight difference between the resolution of the NCEP analysis and that of the CMC analysis, it is worthwhile to test whether this could account for the skill of the CMC analysis in the extratropical North Atlantic. Accordingly, the CMC daily analysis was produced on a global 1° latitude–longitude grid for the period of the study. The results appear in Table 4. The change of analysis grid is a small one and, with the exception of the WNA region, the impact is not significant. For the WNA region, however, the analysis error is approximately 4% higher because of the lower-resolution analysis grid. Thus, resolution alone cannot account for the 30% difference between the CMC analysis and the NCEP analysis in the WNA region in Table 2.

5. Discussion and conclusions

A statistical interpolation analysis of SST has been presented that satisfies some of the specific needs of a numerical weather prediction system. These include the requirement for a global analysis at a relatively high resolution and the need for an analysis that is current,
that includes a seasonal trend even where observations are not available, and that integrates real-time lake temperature data. Several simplifying assumptions are implicit in the analysis procedure, namely that the observation errors and correlation length scales are global constants and that the correlation of the background error is isotropic. Despite its simplicity, the performance of the analysis is encouraging. This performance is partly due to the careful quality control of the observations. The two main components of the quality control are the removal of the bias for ships whose SST reports have a nearly constant bias and the application of a strict check against the background and against neighboring reports. The choice of rejection criteria for this check was guided by the following considerations: in the absence of observations, the analysis eventually returns to a climatological state and the globally averaged error of climatology is near 1°C, an error that is acceptable for NWP purposes. Therefore, for data-sparse regions, the error incurred by rejecting a correct observation is limited to the error of climatology, locally, whereas the error incurred by accepting an incorrect observation may easily exceed the error of climatology. Moreover, it was found that even in data-rich areas, the quality of the analysis benefitted from the use of strict rejection criteria.

Several analysis characteristics important to NWP are illustrated in section 4. It is shown that the analysis can represent both persistent and transient anomalies and does so without compromising its role as a filter of observation error. Unlike the algorithms of Reynolds and Smith (1994) and Clancy et al. (1990), the present scheme uses only 24 h worth of data, making it responsive to sharp trends in the data. Also, the use of a 212-km length scale for the correlation function can yield a very intense and detailed temperature gradient associated with the Gulf Stream.

The analysis has been verified against two sets of independent data for a period of one year. The results were compared to identical verifications of the NCEP analysis and climatology. The CMC analysis does not compare well to the NCEP analysis in the Southern Hemisphere, where the inclusion of satellite data is clearly very important. In the Northern Hemisphere, however, the CMC analysis fares better, having the lowest error estimate for the extratropical North Atlantic for both sets of independent data. In two other regions, the Pacific from the equator to 30°N and the extratropical Pacific from the date line to the Asian coast, the comparison was inconclusive, with the NCEP analysis verifying better against buoys and the CMC analysis verifying better against XBT data. The findings of section 4 also indicate that observations from ships contain sufficient information to improve upon climatology by about 27% in the Northern Hemisphere, proving the value of ship data in this application.

The effect of three unique characteristics of the analysis was examined in section 4. It was found that the analysis could be improved slightly by averaging the anomaly fields over a 7-day period, a strategy analogous to the 7-day time window for observations used in the NCEP analysis. The cost of such an approach is a lag of 3.5 days from real time and a filtering of transient anomalies, both of which are undesirable for NWP. The ship debiasing scheme was also examined and found to contribute significantly to analysis quality (a 5% lower error estimate in the Northern Hemisphere). Finally, using a 1° latitude–longitude grid rather than the 0.9° grid used here was found to have a significant effect only near the Gulf Stream, where errors increased by 4%. Of the three test analysis configurations examined, the procedure to debias ship observations had the most important impact.

There is potential for improving this analysis by allowing some spatial variation of the observation error variances and by relaxing the assumption of isotropy for the background error correlations. These ideas need to be investigated. Moreover, satellite data is a rich source of additional SST information, particularly in data-void or data-sparse regions. However, satellite data like the MCSSTs of McClain et al. (1985), currently carried on the GTS, contain biases (Reynolds et al. 1989). Reynolds and Smith (1994) remove these biases using a blended analysis (Reynolds 1988), which relies on the in situ data where sufficient in situ data are present, and is faithful to the temperature gradient computed from the satellite data else-

<table>
<thead>
<tr>
<th>Region</th>
<th>CMC daily analysis</th>
<th>7-day average of CMC anomaly fields</th>
<th>CMC daily analysis with no ship debiasing</th>
<th>CMC daily analysis using a 1° × 1° grid</th>
</tr>
</thead>
<tbody>
<tr>
<td>NH</td>
<td>0.77 ± 0.004</td>
<td>0.75 ± 0.004</td>
<td>0.81 ± 0.004</td>
<td>0.78 ± 0.004</td>
</tr>
<tr>
<td>WNP</td>
<td>1.09 ± 0.025</td>
<td>1.06 ± 0.024</td>
<td>1.11 ± 0.026</td>
<td>1.10 ± 0.025</td>
</tr>
<tr>
<td>ENP</td>
<td>0.72 ± 0.009</td>
<td>0.70 ± 0.009</td>
<td>0.79 ± 0.010</td>
<td>0.72 ± 0.009</td>
</tr>
<tr>
<td>WNA</td>
<td>1.36 ± 0.026</td>
<td>1.33 ± 0.025</td>
<td>1.44 ± 0.028</td>
<td>1.42 ± 0.027</td>
</tr>
<tr>
<td>ENA</td>
<td>0.80 ± 0.007</td>
<td>0.78 ± 0.007</td>
<td>0.83 ± 0.007</td>
<td>0.80 ± 0.007</td>
</tr>
<tr>
<td>TNP</td>
<td>0.58 ± 0.005</td>
<td>0.57 ± 0.005</td>
<td>0.63 ± 0.005</td>
<td>0.58 ± 0.005</td>
</tr>
<tr>
<td>TNA</td>
<td>0.55 ± 0.009</td>
<td>0.54 ± 0.009</td>
<td>0.60 ± 0.010</td>
<td>0.55 ± 0.009</td>
</tr>
<tr>
<td>SH</td>
<td>0.78 ± 0.004</td>
<td>0.78 ± 0.004</td>
<td>0.81 ± 0.004</td>
<td>0.79 ± 0.004</td>
</tr>
<tr>
<td>TSH</td>
<td>0.63 ± 0.005</td>
<td>0.63 ± 0.005</td>
<td>0.67 ± 0.005</td>
<td>0.63 ± 0.005</td>
</tr>
<tr>
<td>MSH</td>
<td>0.94 ± 0.008</td>
<td>0.94 ± 0.008</td>
<td>0.96 ± 0.008</td>
<td>0.94 ± 0.011</td>
</tr>
</tbody>
</table>
where. A further problem with the remotely sensed SST data is that coverage drops to zero at high latitudes (see Fig. 3 from Reynolds 1988), where the need for additional information is acute. Nevertheless, inclusion of the MCSSTs would undoubtedly benefit the analysis, and work is currently under way toward this end.

Acknowledgments. The analysis was improved by consultation with several people. In particular, I thank R. W. Reynolds for his advice, for proposing to include the NCEP analyses in the study, and for suggesting the algorithm used to return the anomalies to zero. I am indebted to an anonymous reviewer who contributed useful advice and to D. E. Parker for providing the climatology. Special thanks to H. L. Mitchell for his encouragement and for constructive suggestions on an earlier draft of this paper.

APPENDIX A

Some Considerations in the Specification of User-Defined Constants

A particularly important analysis parameter is the ratio of the observational error variance to the prediction error variance, which can be considered as a noise to signal ratio. For ships, a globally averaged value for this ratio was determined by experiment and can only be defended empirically. Its value of 4.0 strikes a balance between the need to maximize the use of observations from ships, the principal data source in this analysis, and the need to filter the random errors in the observations. Assuming that the background and ship errors are uncorrelated, the ship error variance and background error variance were constrained to sum to the actual total variance. The latter is calculated using

$$E^2_{\text{total}} = \frac{1}{N} \sum_{i=1}^{N} (O_i - B_i)^2,$$

(A1)

where $O_i$ is the ship reported anomaly, $B_i$ is the background at the location of the ship, and the variance is computed for a large ensemble of $N$ ship reports that have passed the quality control checks. For a one-month global sample of 55 700 ship reports, the total variance is 0.90 K$^2$. Using the noise to signal ratio of 4.0, this yields a global average background error variance of 0.18 K$^2$ and a ship error variance of 0.72 K$^2$.

The estimate of ship error variance obtained by this method is much lower than the value of 1.39 K$^2$ reported by Trenberth et al. (1992). The most important cause of this discrepancy is the comprehensive quality control used here. To illustrate the effect of the quality control scheme on the total variance, a different quality control algorithm with the same rejection rate as the comprehensive check is applied to the same one-month sample of ship reports used above. This alternative scheme is the test against climatology that was applied to the XBT data in section 4. It yields a total variance of 1.4 K$^2$, which is 55% higher than the value obtained with the comprehensive check. Subtracting the background error variance, this corresponds to a ship error variance of 1.22 K$^2$, which is quite close to the 1.39 K$^2$ estimated by Trenberth et al. (1992). It should also be noted that Trenberth et al. (1992) are using ship data to estimate monthly mean SSTs and not daily mean SSTs as is the case here. This implies that the contribution to their estimate of the variance due to “incomplete sampling of the within-month variance” is not present in the ship error variance here.

The observation error variance for buoys requires a different approach to that used for ships. This is because the positions of the buoys change little from day to day (not at all in the case of moored buoys), which means that the background reflects recent data from the buoys, and therefore observation error is likely to be correlated with background error, making any estimate of the observation error unreliable. However, if the background were independent of the buoy data, as is the case for the analysis produced with ship data only, one could use the buoy data to compute the total variance given by (A1). By subtracting the globally averaged background error variance from this total variance, one would have the required estimate for the buoy error variance. The globally averaged total error variance, $E^2_{\text{total}}$, computed using drifters is 0.54 K$^2$. This gives a drifter variance of 0.36 K$^2$. The fixed buoy variance is taken to be the same as the drifter variance.

The constants appearing in (1) were chosen to satisfy the requirement that the globally averaged background error variance be 0.18 K$^2$, as discussed above. Since $E^2_{\text{g}}$ averages to approximately 1 K$^2$, this means that $\gamma$ and $\delta$ must sum to approximately 0.18. Over most of the globe, the choice of these constants has little effect. However, over the Gulf Stream, the weighting of $E^2_{\text{g}}$ is more important. Therefore, using the climatological variance $E^2_{\text{c}}$ for February, the values for $\gamma$ and $\delta$ were chosen to give a ratio of the variance in the Gulf Stream to the variance in the Tropics of 4.0, a ratio typical of those of Clancy et al. (1992, Fig. A3) and Reynolds and Smith (1994, Fig. 11).

Another user specified constant is the correlation function $\epsilon$-folding distance of 212 km. This value is based on the north–south correlation scales of Fig. A2 of Clancy et al. (1992). The east–west correlation scales of Clancy et al. (1992) are considerably larger than the north–south scales, but isotropy is assumed, making it necessary to choose just one length scale for east–west and north–south correlations. The smaller length scale is chosen in order to represent smaller scales where data permit.

APPENDIX B

An Algorithm for Debiasing SST Reports from Ships

A file of recent observations is consulted immediately following the retrieval of the observations from the da-
tabase and prior to any analysis or quality control procedures. The file is organized by ship call sign with nonunique call signs (such as SHIP) excluded. Observations are stored along with the value of the background (the previous analysis) interpolated to the observation location, so that \( O - B \), the difference between the observed SST and the background, may be computed. The historical record for each ship goes back 30 days or 360 observations, whichever is less. The observations stored in the file have undergone neither the data redundancy reduction steps nor the quality control steps described in section 3. Therefore, observations are excluded when

\[
|O - B| > 4^\circ C,
\]

so that outliers do not contaminate the estimate of the apparent bias

\[
O - B = \frac{1}{N} \sum_{i=1}^{N} (O_i - B_i),
\]

where \( N \) is the number of historical observation-background pairs satisfying (B1). Although the variances of neither the ship observations nor the background are known, it is nevertheless useful to estimate the scatter in the sample using

\[
s = \left[ \frac{1}{N-1} \sum_{i=1}^{N} (O_i - B_i - (O - B))^2 \right]^{1/2}.
\]

The current observation is then debiased by subtracting \( O - B \) if the following conditions are met: \( N \geq 11, s < 1.0^\circ C, O - B > 0.5^\circ C, \) and \( O - B - 2s < a - b < O - B + 2s \), where \( o \) is the current observation and \( b \) is the current background interpolated to the observation location. Note that if \( a - b \) is inconsistent with the biases of the recent past, or if the apparent bias is small or was estimated from a sample with a large scatter, no attempt is made to apply a correction.

REFERENCES


