Extended Reconstruction of Global Sea Surface Temperatures Based on COADS Data (1854-1997)

Thomas M. Smith and Richard W. Reynolds

National Climate Data Center
151 Patton Ave., Asheville, NC 28801

February 6, 2002
Abstract

A monthly extended reconstruction of global SST (ERSST) is produced based on COADS observations from the 1854-1997 period. Improvements come from the use of updated COADS observations with new quality-control procedures and from improved reconstruction methods. In addition error estimates are computed, which include uncertainty from both sampling and analysis errors. Using this method, little global variance can be reconstructed before the 1870s because of data sparsity. Error estimates indicate that in most regions ERSST is of limited value before 1880, when the uncertainty of the near-global average is almost as large as the signal. In most regions, the uncertainty decrease through most of the period and is smallest after 1950.

The large-scale variations of ERSST are broadly consistent with those associated with the HadISST reconstruction produced by the U.K. Met Office. There are differences due to both the use of different historical bias corrections and different data and analysis procedures, but these differences do not change the overall character of the SST variations. Procedures used here produce a smoother analysis compared to HadISST. The smoother ERSST has the advantage of filtering out more noise at the possible cost of filtering out some real variations when sampling is sparse. A rotated EOF analysis of the ERSST anomalies shows that the dominant modes of variation include ENSO and modes associated with trends. Projection of the HadISST data onto the rotated eigenvectors produces time series similar to those for ERSST, indicating the dominant modes of variation are consistent in both.

1. Introduction

Analyses of historical sea surface temperatures (SSTs) are critically important to global climate-change studies, and several analyses have been performed (e.g., Parker et al. 1994, Smith et al. 1996, Kaplan et al. 1998). These methods indicate generally similar variations in their overlap periods and regions. But there are differences because of some different input data and different analysis methods. Here we produce a global, extended reconstructed SST (hereafter referred to as ERSST), monthly beginning in the 19th century. Improvements include additional data from a new version of COADS, improved quality control of that data, and an improved statistical analysis method. We also produce an error estimate for the reconstruction to show where and when it may be used with confidence.

The analysis method is an outgrowth Smith et al. (1996). The Smith et al. (1996) reconstructed SST (hereafter referred to as RSST) partly overcame the problem of uneven sampling and noisy data by separately analyzing low- and high-frequency variations. Because of their larger scales, the low-frequency variations can be analyzed using simple averaging and smoothing of relatively sparse data. This simple analysis does not require stationary statistics, which may be difficult to define for variations with periods of decades or longer. Interannual and shorter-period variations are spanned by the globally-complete SST analyses based on satellite data (Reynolds and Smith 1994, Reynolds et al. 2002), so stationary statistics based on these data may be used to analyze the high-frequency variations.

In the RSST the high-frequency SST is analyzed by fitting observed high-frequency SST anomalies to a set of empirical orthogonal functions (EOFs), based on the twelve years of spatially complete SST analyses available at that time. For each month the weights for the set of modes are found by fitting the observed data to the modes. This analyzes the high-frequency anomalies for the entire region defined by the modes, while random errors and other variations not represented by the base-period modes are filtered out. The low- and high-frequency components are added for the total RSST anomaly.

A problem with the Smith et al. (1996) method is that it may become unstable if used for
analyses with extremely sparse data. Therefore, the RSST was not computed before 1950 or south of 45°S. The method of Kaplan et al. (1998) is appropriate for producing analyses with sparse data, but since it analyzes all frequencies the same way it requires a much longer base period to develop analysis statistics. Since the satellite data do not cover a long enough base period to be used with their method, the Kaplan et al. (1998) analysis develops statistics from in situ data which have large gaps in the Southern Hemisphere. In addition, its base period may not span all inter-decadal variations. To overcome the instability that can occur using the Smith et al. (1996) method while maintaining its strengths, Smith et al. (1998) modified the method so that it is stable with extremely sparse data. Here we use that modified method to produce an extended reconstructed SST (ERSST) analysis.

In section 2 the data are described, including improved quality-control procedures and historical bias corrections. Section 3 describes the reconstruction methods. The ERSST error estimation is discussed in section 4, and results of the reconstruction are given in section 5. In section 6 large-scale variations are discussed. Conclusions are given in section 7.

2. Data

The SST data used for ERSST are derived from the latest version of the Comprehensive Ocean Atmosphere Data Set (COADS, Slutz et al. 1985, Woodruff et al. 1998), with updates through 1997. We average the SSTs to super observations, defined as 2°E monthly averages, with grid centers on 0°E, 2°E, ..., 2°W by 88°S, 86°S, ..., 88°N. This grid is offset by 1°E from the standard 2°E COADS grid to better resolve equatorial signals (e.g., ENSO). The annual number of individual SST observations (Fig. 1, solid line) is largest after 1960. There are large numbers of observations in several periods after 1900, with lows around the 1915-1920 period and the 1940s. The annual number of global SST super observations (dashed line) is less variable, but there are still relative lows in the same two early 20th century periods. These time series include all data, including suspect observations that are not used for the reconstruction. Below we discuss the quality-control system to remove these suspect observations.

2a. SST Quality Control

Data screening, or quality control, is needed to eliminate outliers. Causes for outliers include misreading of thermometers, errors copying data, or ship position errors. The screening currently provided by COADS excludes most outliers, but it may also exclude some good data in situations when anomalies are strong. Wolter (1997) found that in the eastern-equatorial Pacific, some reasonable SST observations associated with a warm episode in 1878 are discarded by the current COADS data screening. Since historical SST are often sparse, we wish to avoid discarding good observations. Therefore we developed a data screening method that removes outliers while minimizing the rejection of good data.

The quality control (QC) used here is a preliminary version of QC procedures being developed for use with COADS data. The QC method used here for SST is described in more detail in Appendix A. It checks individual normalized anomalies against a normalized local analysis of anomalies. Because individual anomalies are compared to a local analysis, large anomalies that are supported by other data are not flagged as bad, while isolated anomalies greatly different from neighbors are flagged. The annual percentage of individual observations
flagged by our SST QC (Fig. 2, solid line) is lowest, about 2%, before 1900. The percentage increases as data increases. Some of the increased percentage of flagged observations is due to an increased frequency of flagging at mid latitudes as data become more dense. As shown by Appendix A, as data become more dense a higher quality local analysis is available to compare against individual observations. In that case individual anomalies must be closer to the analysis for the observation to pass QC. However, mid-latitude flagging rarely exceeds 10% and is usually less than 5%. Much of the increased percentage of discarded observations in the 1930s and near 1980 is due to heavy flagging of observations north of 70°N in those periods. Before 1910 there are very few data in that region, with many more in the 1930s and in near 1980. North of 70°N, typically 50% or more individual observations are flagged by the SST QC.

With super observations, there is reduced loss due to QC. That is expected, since a super observation may be formed even if several of its individual observations are discarded. The greatest percentage of lost super observations occurs after 1970, when about 2% are lost. This increased loss of super observations is almost entirely poleward of 60° latitude, and much of it is north of 70°N. Over most of the period fewer than 1% of super observations are lost due to the QC.

2b. Satellite and In Situ Analysis

The combined satellite and in situ analysis of Reynolds et al. (2002) is used to develop spatially-complete statistics for our reconstruction. The Reynolds et al. (2002) analysis is an improved version of Reynolds and Smith (1994). Changes are greatest at high latitudes because of an improved sea ice to SST conversion algorithm.

We average the monthly 1982-2000 Reynolds et al. (2002) analysis to the same 2° super observation grid that we use for the COADS data. In addition, we computed a SST climatology for the 1982-2000 period. Thus, the stationary statistics that we compute use the data from the last sixteen years of this study’s analysis period with an additional three years (1998-2000). The nineteen-year climatology will have good resolution in all locations without any data gaps. Our historical anomaly analysis is computed with respect to this nineteen-year climatology. If desired, the anomaly base period can easily be readjusted to any sub-period of the historical analysis (e.g., Smith and Reynolds 1998).

2c. Historical Bias Corrections

Folland et al. (1984), Bottomley et al. (1990), and Folland and Parker (1995, hereafter FP95) show the need for bias corrections for historical SST, and suggest several possible correction methods. These methods apply systematic corrections to SST before 1942, which removes the sharp step in SST that occurs at the end of 1941. In all methods, the global-mean SST cold bias before 1942 is about 0.3°C relative to the SST from 1942 on. The sharp change in SST across the boundary is associated with changes in measurement techniques and data sources associated with the Second World War.

An independent analysis of historic bias corrections (Smith and Reynolds 2002, hereafter SR02) suggested an alternative bias correction method and showed a general consistency with the FP95 bias correction. The largest difference between the SR02 bias correction and that of FP95 is in winter at high latitudes, where the SR02 bias correction is stronger. However, the overall average corrections are similar, and we are unable to determine which correction is more
accurate using the available data. As discussed below, our final analysis uses the SR02 bias correction. Differences between the SR02 and FP95 define uncertainties in the analysis caused by the need for bias corrections.

3. Analysis Method

We analyze monthly anomalies with respect to the 1982-2000 base period using the method of Smith et al. (1998), adapted to a global reconstruction. The anomaly reconstruction is performed separately for the low- and high-frequency components, which are then added together to form the total SST anomaly. The low- and high-frequency components are separated because the stationary statistics used for the high-frequency analysis are based on only nineteen years of SST anomalies, and thus may not adequately span interdecadal variations. Both low- and high-frequency variations are reconstructed using screened COADS 2E super observations.

3a. Low-Frequency Analysis

The low-frequency analysis needs to represent the large-scale, slowly changing SST anomaly variations that may not be represented by the Reynolds et al. (2002) base period. We compute this low-frequency analysis by smoothing and filtering anomalies within 10E spatial regions using 15 years of data to generate one low-frequency analysis per year. This low-frequency anomaly is removed from the observed SST anomalies before analysis of the high-frequency, and will be added back at the end.

The low-frequency 10E grid covers the globe equatorward of 75E latitude. Poleward of 75E is only a small percentage of the global ocean and almost no data. Therefore, for those small regions a low-frequency anomaly of zero is assigned for all times. The 10E anomalies at all other locations are computed in several steps to ensure that there are enough data to filter out high frequencies.

To separate the low-frequency analysis we first form monthly 10E anomalies for squares that contain at least three 2E super observations, and at least nine individual in situ observations. The 2E super observations are weighted by their relative area and by their relative sampling. Second, for each calendar year annual 10E anomalies are formed by averaging the monthly 10E anomalies at each location, provided that at least four monthly 10E anomalies could be defined for the year. These 10E annual anomalies are then filtered using zonal and meridional three-point binomial filters, to further reduce small scale variations. The binomial filtering fills in undefined annual 10E regions with filtered values from adjacent defined 10E regions, which slightly expands the spatial extent of the annual field. The percent of the total area for which annual 10E anomalies could be defined (Fig. 3) shows that the most serious data gaps occur before 1870, with smaller gaps in the 1890s, 1910s and 1940s.

Next, the annual anomalies are temporally median filtered using a moving 15-year window centered on each year. We require that at least three of the 15 years be defined to compute the median, so this time filtering step further fills in the field. Median filtering is preferred to a running mean because it more effectively removes the influence of outliers. This produces a 15-year anomaly for most 10E regions for each calendar year. The final step is to set remaining undefined 15-year anomalies to zero and apply additional spatial and temporal binomial filters to smooth the result. The final spatial filters are three-point zonal and meridional
binomial filters, as described above, and the temporal filter is a five-year binomial filter.

The 15-year moving window filter removes interannual and shorter-period variations, which will be spanned by the 19-year base period used for the high-frequency analysis. The choice of a 15-year analysis period is dependent on the length of the high-frequency base period and the need to resolve variations not spanned by the high-frequency base. To test this analysis period we compare the time series of the leading EOF of the 15-year analysis with leading EOF time series from similar analyses, with lengths of 5, 11, 19, and 25 years. In all cases the leading EOF accounts for about half of the analysis variance. While all of the mode 1 time series are similar, the 5-year analysis time series has large variations with 8-15 year periods and thus is too short. The 11- to 25-analyses all indicate similar variations, and either could be used. However, the 11-year analysis includes some weak variations with 10-15 year periods, so a slightly longer period is preferred. The 25-year analysis begins to damp variations with periods greater than 30 years, so the shorter 15-year period is better.

The low-frequency analyses are computed using both the SR02 and FP95 bias corrections (see section 2c). For both, the first three EOFs account for about 80% of the variance, and the two low-frequency analyses produce almost identical modes of variation. Most of the low-frequency variation is associated with warming trends, discussed later. For our analysis we use the SR02 bias corrected SSTs. A user may adjust our analysis to use a different bias correction by adjusting our final analysis by the bias-correction difference.

3b. High-Frequency Analysis

Our analysis of high-frequency anomalies uses the method of Smith et al. (1998). Our 2E version of the 1982-2000 Reynolds et al. (2002) analysis is used to define a set of analysis-increment modes, or spatial patterns. Here, analysis increments are defined as differences between a monthly anomaly and the anomaly of the previous month. In addition, data increments are defined as differences between the data anomaly for a month and the high-frequency analysis anomaly for the previous month. Please recall that all data are super observations from which the climatology and our low-frequency analysis have been removed. The high-frequency analysis increment is computed by fitting the data to the set of spatial modes. Modes that are not supported by sampling are not used, and variance associated with unsupported modes is damped using the mode’s autocorrelation. We compute the high-frequency analysis in both forward and backward temporal directions, and average the results so that temporal information from both directions is included. Details of the analysis method are given by Smith et al. (1998). The high-frequency analysis method is also described in more detail in Appendix B.

The spatial modes need to represent robust patterns of spatial covariance for the high-frequency anomaly increments. Variations which are not common over the entire reconstruction period should be filtered out of the modes as much as possible. For example, patterns representing covariance across several ocean basins or across extremely large regions may not be common over the entire period. In Smith et al. (1996) this problem was overcome by dividing the global ocean into six separate basins with some overlap.

There are several possible methods of computing modes. In Smith et al. (1998) a set of rotated covariance EOFs were used to define the modes. For the tropical Pacific region of that study only a few modes were needed, and the rotation cleanly defined them. Here we wish to
perform a global reconstruction, so a much larger set of modes is necessary. We first examined large sets of rotated EOFs, by rotating sets of the first 55 and the first 100 covariance EOFs. Most of the modes represent covariance patterns with scales of about $10^E - 15^E$ in latitude by about $30^E$ in longitude. However, for both sets a few modes have patterns that stretch across several ocean basins and may not represent covariance common to the entire analysis period.

The method of empirical orthogonal teleconnections (EOTs, van den Dool et al. 2000) can also be used to develop a set of covariance patterns. With EOTs, the point that represents the maximum covariance of the field is chosen as the base point. The regression pattern associated with that point is computed and its variance is removed from the data. Then the process is repeated to compute the next mode. We applied a variation of EOTs to compute a set of covariance patterns. First, we applied a three-point binomial smoother to our 2$^E$ version of the Reynolds et al. (2002) anomaly increments, both spatially and temporally, so that our patterns will not reflect small-scale variations that are unlikely to be robust. In addition, we restricted the selection region for base points to exclude places where there is little historical sampling. Regions excluded are south of $60^E$, north of $64^E N$, the Caspian Sea, Black Sea, Baltic Sea, Sea of Okhotsk, Hudson Bay, and the Great Lakes. For some of those regions isolated modes are computed, as discussed below. In the remainder of those regions variations are only considered if they vary with base points outside of those regions. To eliminate excessively large teleconnections we localize each mode by setting it to zero at distances greater than 8000 km from the base point, and linearly damping it to zero in the range 5000 to 8000 km from the base point. The EOT modes computed this way show many patterns that are almost identical to the rotated EOF modes in both scale and shape. However, we are able to control these EOT modes to avoid the cross-basin linkages that occurred with rotated EOFs. We found that 69 EOTs account for most global increment variations, with higher modes describing more localized features.

Because there is some sampling in the Arctic, especially in the boreal summer months, we computed a second set of EOT modes with the sampling region restricted to north of $64^E N$ and Hudson Bay. In that analysis all EOT variations are computed only in the restricted regions (i.e., the region is isolated). We use only the first five EOTs from the Arctic analysis. A similar isolated EOT analysis was computed for the Caspian Sea, where there is also occasional sampling. One mode was found to account for much of the variance in that region. These three EOT analyses (near-global, Arctic, and Caspian Sea), were merged to give a comprehensive global set of 75 modes. Teleconnections from the global region into the Arctic and Caspian regions were masked out so that all variations in these regions were isolated.

The one-month autocorrelation for each mode, needed for damping unsupported modes, is computed by projecting the high-frequency COADS anomalies onto each mode and then computing the autocorrelation of that time series. The COADS anomalies for 1982-1997 were used because the sampling was sufficient to define an autocorrelation for each mode for this period. Autocorrelation values ranged between 0.17 (for the Caspian Sea) to 0.94 (for the Tropical Pacific), with values commonly between 0.6 to 0.8 (Table 1). Those autocorrelations correspond to e-folding decay times ranging from about one month to over a year, with a typical decay time of about three months. Modes that represent variations with longer time scales will have greater persistence, and thus can make greater use of data from months other than the analysis month.
Table 1. Number of EOT modes with 1 month lag autocorrelation (ac1) in the given range. There are a total of 75 modes.

<table>
<thead>
<tr>
<th>Range: ac1</th>
<th>Number:</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.4&lt;ac1</td>
<td>1</td>
</tr>
<tr>
<td>0.6&lt;ac1</td>
<td>19</td>
</tr>
<tr>
<td>0.8&lt;ac1</td>
<td>47</td>
</tr>
<tr>
<td>0.8&lt;ac1</td>
<td>8</td>
</tr>
</tbody>
</table>

We defined when the sampling was adequate for each mode using the percent of variance sampled for each mode. High-frequency anomalies associated with adequately-sampled modes are updated using the data, while anomalies associated with under-sampled modes are damped. We wish to avoid situations in which a mode is only sampled outside of its center of action. In those situations, variance sampled may be as high as 10 to 12%, so we know that we need at least that sampling for each mode. Appendix B defines how the percent of sampled variance is computed for each mode. To better define adequate sampling, we use cross-validation tests (e.g., Smith et al. 1996), validated against the Reynolds et al. (2002) SST analysis. The cross-validation analyses are computed using the 1982-1986 SST anomalies with modes derived from the 1988-2000 Reynolds et al. (2002) analysis. The five-year validation period spans interannual variations similar to what may have occurred in the past, and the thirteen-year period used to derive the modes is independent of the validation period. Data for each cross-validation test are COADS data subsampled to simulate sampling in historical years, which simulates how well those years may be analyzed using our methods. The sampling for years 1860, 1918, and 1942 are used, which represent times when sampling is relatively sparse and thus provides a severe test. The global average error is checked to determine the best overall critical value (Table 2). Overall there is little difference between 12% and 15% sampling, but for higher percentages the error can increase significantly. We use the more conservative value, 15%, to ensure against using modes that could introduce artificial variations. Thus, modes with less than 15% of their variance sampled are not fit to the data and are damped with time.

Table 2. Global RMSE (EC) from cross-validation tests for different sampling years and critical values (% sampling).

<table>
<thead>
<tr>
<th>Year</th>
<th>12%</th>
<th>15%</th>
<th>18%</th>
<th>21%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1860</td>
<td>0.575</td>
<td>0.580</td>
<td>0.591</td>
<td>0.597</td>
</tr>
<tr>
<td>1918</td>
<td>0.510</td>
<td>0.513</td>
<td>0.526</td>
<td>0.528</td>
</tr>
<tr>
<td>1942</td>
<td>0.504</td>
<td>0.497</td>
<td>0.497</td>
<td>0.505</td>
</tr>
</tbody>
</table>

4. Error Estimation

The ERSST mean-squared error (MSE) can be written as the sum of the sampling errors and analysis errors. Sampling errors are due to data gaps, and are most severe early in the analysis period. Analysis errors can have several causes, including uncertainties in bias corrections, quality-control procedures, and statistical methods used to analyze the SSTs. Here we separately compute error components, and estimate the total error from their combined effects. The MSE discussed here is for annual averages of SST anomalies, averaged spatially over regions.
Components of the total MSE include the low-frequency sampling error ($E_{LFS}^2$), the high-frequency sampling error ($E_{HFS}^2$), and the analysis error ($E_{An}^2$). The analysis error includes uncertainties caused by the need for a bias correction. Computation of these three components is discussed below. For each component, any error correlated with the others is removed so that they may be considered to be independent. Thus, the total MSE is

$$E^2 = E_{LFS}^2 + E_{HFS}^2 + E_{An}^2.$$ 

Sampling errors are largest when anomalies are large, since the analysis damps to zero anomaly when data become sparse. The low-frequency analysis adjusts the mean when sampling is adequate, but with extremely sparse sampling such as in the 1860s the low-frequency analysis will be greatly damped and the adjustment will tend to be weak. Damping of the low-frequency analysis leads to low-frequency sampling MSE ($E_{LFS}^2$). To evaluate this component of the MSE we compute the analysis using historical sampling from a globally-complete data set. Since the globally-complete data must simulate SST trends, we use output from the Geophysical Fluid Dynamics Laboratory (GFDL) R30 coupled model, forced with radiative forcing that simulate trends (Delworth et al. 2002). Note that this model output is used only to estimate $E_{LFS}^2$.

We use model SSTs from three 125 year runs (1866-1990), run with identical radiative forcing but initialized differently. The model SSTs are converted to anomalies by subtracting out the model ensemble-mean climatology from the last 19 years of the runs. Averages and EOFs of SST from these runs show low-frequency variations similar to those in ERSST and in HadISST (Parker et al. 1999) in both their spatial extent and timing. The model high-frequency variations do not compare as well as the low-frequency variations. Also, the timing of model high-frequency variations is not linked to radiative forcing, and thus is different from what is observed.

For each ensemble we compute an analysis of the model SSTs following the same procedures used to produce ERSST, with the model monthly 2E SST anomalies sub sampled to match the ERSST sampling for the appropriate month. The statistics for the model analyses are computed from the last 19 years of the model ensemble mean. For the model analyses 67 modes are used, compared to 75 modes for ERSST. The three sets of full and analyzed model SSTs are averaged. For both the full and analyzed ensemble means, the SSTs are time filtered using a 21-year moving window, to remove the high-frequency model variations. This period is slightly longer than was used to define the low frequency for the analysis because the model interannual variations in the tropical Pacific have a slightly longer period than observed.

The difference between the full and the analyzed low-frequency model SSTs are used to compute $E_{LFS}^2$. However, comparing the low-frequency model analysis to the comparable filtered ERSST shows that the model analysis trends are larger than for ERSST, indicating that the model trends are slightly too large, and thus should be adjusted slightly. To make an adjustment we find the constant, $a$, for each time and each region to approximate the full model low-frequency average, $F_m$, using the analyzed model low-frequency average, $R_m$, such that
$F_m = aR_m$. We may do this since the analysis used here produces a low-frequency anomaly similar to the full-data low-frequency anomaly, except the analysis anomaly is damped because of incomplete data. Damping of the analysis anomaly is inversely proportional to the sampling available for the reconstruction. The constant $a$ that minimizes error over a given period is

$$a = \left\langle R_m F_m \right\rangle / \left\langle R_m^2 \right\rangle,$$

where the brackets denote averaging over the period, here defined as 21 years centered on the year of interest. Since the amount of damping depends on sampling, we use assume the same damping for the observed low-frequency anomaly relative to the ERSST low-frequency analysis, $R_o$, and therefore for each year we estimate the adjusted low-frequency error as

$$E_{LFS}^2 = \left\langle (aR_o - R_o)^2 \right\rangle = (a - 1)^2 \left\langle R_o^2 \right\rangle. \quad (1)$$

We further smooth this low-frequency MSE estimate by averaging over an eleven-year moving period, centered on the year of interest. Note that we do not attempt to correct the low-frequency signal using the model output, but only use the model to estimate uncertainty.

Cross validation is used to compute the high-frequency sampling component of MSE. The cross-validation analyses are nearly identical to those used to tune the high-frequency analysis, described in section 3b. The difference is that here the cross-validation analysis is done using the 1982-1986 Reynolds et al. (2002) anomalies. The mean anomaly is removed since differences in the mean are part of the low-frequency error, and the anomalies are sub sampled to match sampling from each year in analysis period. Validation data is the same data, but without sub sampling. Differences between the cross-validation estimate and the full-sampling data indicate high-frequency sampling error. Since we are here interested in errors of annual averages, we compute the annual average for each of the five cross-validation years and for the validation data. Validation data are anomalies from the same five years, but with no sub sampling or analysis. The instantaneous high-frequency error for the year, $\varepsilon_{HFS}$, is defined as the root-mean-squared difference between the five cross-validation and validation years. The high-frequency MSE is defined as

$$E_{HFS}^2 = \left\langle \varepsilon_{HFS}^2 \right\rangle, \quad (2)$$

where the averaging in (2) is for the same period used in (1).

Besides the sampling errors discussed above, there are also errors due to the need to use a bias correction, the need for quality control procedures, and other differences in analysis techniques. To estimate these remaining analysis errors, we compare ERSST to two other analyses produced by the U.K. Met Office: GISST (Parker et al. 1994) and HadISST (Parker et al. 1999). Those two analyses use different quality control, some different input data, different historical bias correction methods, and different analysis methods compared to ERSST. The ERSST MSE is defined as
\[ E_{An}^2 = \left\langle \left( ERSST - \text{Avg} \right)^2 \right\rangle, \]  \hspace{1cm} (3)

where \( \text{Avg} \) is the average of ERSST, GISST and HadISST. The averaging period in (3) is as in (1) and (2).

This MSE component can be further divided into the bias component of MSE, \( B_{An}^2 = \left( \left\langle R \right\rangle - \left\langle \text{Avg} \right\rangle \right)^2 \), and the non-bias MSE component, \( D_{An}^2 = E_{An}^2 - B_{An}^2 \). Uncertainty caused by SST bias is greatest in the pre-1942 period when bias corrections are needed. However, comparison of COADS SSTs from all sources to SSTs from engine intakes only, for the 1994-1997 period, shows that there remain small biases between these two sources of SST. Excluding high latitudes where sampling is sparse, the typical difference is 0.015 E. We use this value as the minimum allowable \( B_{An} \).

To ensure that these individual error components are independent, we regress the error components against each other to remove correlated error. This can be a problem if differences in the analyses affect sampling, and thus the analysis error may include some sampling error. First error explained by the low-frequency error is removed from the high-frequency error. Then error explained by the combined sampling error is removed from the analysis error. In practice this make little difference in the total error, indicating that these components are nearly uncorrelated.

For the global, annual average ERSST anomalies the sampling root-mean-square error (RMSE) is largest early in the analysis period (Fig. 4, solid line). Before 1870 or 1880 sampling severely limits the value of any SST analysis, but in the early 20th century sampling error is greatly reduced. This estimated sampling error is similar to global sampling uncertainties estimated by Duffy et al. (2001). The total global error generally remains above 0.02 E in all periods. Total error also increases around 1940, near the time when the historical bias corrections are largest and then decrease to zero. These global SST errors about 50% or more larger than the global SST errors estimated by Folland et al. (2001) using different methods, and thus this estimate may be an overestimate the actual analysis uncertainty.

5. Results

We used the methods of section 3 to compute the monthly average SST anomalies from 1854 to 1997. Here we discuss some aspects of the annual-average ERSST anomalies, in order to illustrate their overall character. We also show why be did not attempt to compute ERSST before 1854. Annual averages of the ERSST global spatial variance (Murphy and Epstein 1989) indicates periods when the global signal is excessively damped due to insufficient sampling. The global spatial variance from 1876 on is usually within 0.2 and 0.4 E², but before 1876 the variance is systematically less (Fig. 5a). Oscillations in the spatial variations after the 1870s may be caused by interannual variations such as ENSO. This indicates excessive damping before the mid 1870s. Nearly all of this reduction in variance is due from the high-frequency variance. The annual average number of modes used (Fig. 5b) is over 30 from 1876 on, and generally less than 30 before then, suggesting that at least 30 modes are needed to reconstruct the global SST anomaly. There are also lows in the average number of modes used in 1893 (37
modes), 1918 (43 modes), and 1945 (34 modes), due to dips in sampling in those years. However, for each of those years the spatial variance remains relatively strong.

For comparisons, annual and spatial averages of SST anomalies for several regions are computed using unanalyzed COADS SST super observations, ERSST, and the HadISST analysis of Parker et al. (1999). Note that the HadISST analysis is independent of ERSST since it is based on U.K. Met Office SSTs, the data are analyzed differently, and it employs the FP95 bias corrections. The HadISST analysis is available monthly for the period beginning 1871. The same climatology is used for all three. For the region 23°N-60°N (Fig. 6) the analyses are similar after 1950. Before 1950 the HadISST average is systematically cooler for most of the period, and systematically warmer in 1900-1915. The differences, about 0.3°C for much of the period, are larger than the annual differences in the bias corrections used, about 0.1°C or less. For much of the pre-1950 period, those Northern Hemisphere differences are near the 95% confidence interval (shown in the lower panel). These differences do not change the overall character of the SST anomaly variation through the 20th century, but they are striking considering the relatively good Northern Hemisphere sampling in that period.

In the 23°E-23°N region (Fig. 7) all three averages are more similar over most of the period, but the unanalyzed COADS variations are larger. The annual and regional average bias corrections in the tropics differ by about 0.1°C, in an opposite direction to the Northern Hemisphere difference. The Northern Hemisphere bias correction differences will tend to make HadISST cooler than the others, while in the tropics the difference tends to make HadISST slightly warmer. This partly explains the slightly warmer HadISST anomalies before 1942. However, HadISST is sometimes cooler in that period, suggesting that other analysis differences are also important.

For the 60°S-23°E region (Fig. 8) all three anomalies show a large warming trend beginning about 1930. Note that the Southern Hemisphere uncertainty is largest in the 1915-1960 period. In that period the Southern Hemisphere sampling is greatly reduced, first by the opening of the Panama Canal in 1914 and later by the Second World War. Even considering the uncertainty due to sampling and analysis error, a warming trend is clear. The Southern Hemisphere trend is similar to the weaker tropical trend, but is different from in the Northern Hemisphere where overall cooling occurred between 1950 and 1985, preceded and followed by warming.

For the near-global area (Fig. 9), there is great similarity between all three after 1900. The uncertainty limits show that the near-global trend over the 20th century is about 0.6°C ± 0.2°C. In the 1982-1997 overlap period, the Reynolds et al. (2002) average anomalies are similar to the ERSST and HadISST average anomalies in all regions except the Southern Hemisphere, where the Reynolds et al. (2002) analysis is biased about 0.08°C cooler, due to some residual bias in the Southern Hemisphere satellite data.

6. Large-Scale SST Variations

Much of the large-scale SST variation of ERSST is from the low-frequency analysis. However, there are other climatic changes relating to changes in the frequency and intensity of interannual variations. This is illustrated by EOFs of annual averages of the ERSST anomaly. Rotation of the first five EOFs of ERSST (Fig. 10) gives three modes similar to the first three
rotated modes of the low-frequency analysis (modes 2, 3, and 5). Rotated EOFs 1 and 4 contain both interannual and decadal variations, indicating important changes in interannual variations over the period. Together this set of five modes accounts for almost 60% of the variance of the annual-average SST anomalies.

For all modes there is little variance before the mid 1870s followed by greater and more uniform variance afterwards, consistent with the change in spatial variance about that time (Fig. 5). The rotated EOF time series are computed by projecting both the ERSST and HadISST anomalies onto the eigenvectors. Consistencies between the two time series are an indication of the degree to which their variations the same.

Mode 1 is an ENSO mode that shows changes in the frequency and intensity of tropical warm and cool episodes over the period. With the exception of the strong 1940-1941 warm episode, there is a tendency for more cool episodes in the 1905-1960 period. The earlier and later periods are both slightly warmer, and there is a sharp increase in the late 1970s. This is consistent with the tropical variations shown in Fig. 7. The time series variance is also larger in this most recent period. Since the late 1970s there have been several large warm episodes, including 1982-1983, 1986-1987, 1992, and 1997. Such strong warm episodes are less frequent in earlier periods, although there are strong warm episodes prior to 1980.

The most important trend mode is mode 2, which indicates Southern Hemisphere warming throughout the 20th century. The trend is consistent for both the ERSST and HadISST time series, although there are higher-frequency differences between the two and the ERSST trend is slightly stronger. Southern Hemisphere warming indicated by mode 2 may be due to a slight shift in the Antarctic Circumpolar Front. A coupled ocean-atmosphere model with changing radiative forcing, to simulate increasing carbon dioxide, showed an expansion of the Southern Hemisphere upper-ocean warm layer in this region (Manabe and Stouffer 1994). In addition, increasing ocean heat content over the second half of the 20th century has been shown by Levitus et al. (2000), who also showed cooling in the Atlantic Ocean north of about 45°N.

The simultaneous North Atlantic cooling indicated by mode 2 could result from a slight freshening of Atlantic near Greenland, slowing down the Atlantic thermohaline circulation. That slowdown may cause the flow of warm water across the North Atlantic to become more zonal, causing a cooling at high latitudes. The Manabe and Stouffer (1994) model also shows a slow down in the North Atlantic thermohaline circulation. Recent observations of decreased Faroe Bank channel overflow, east of Iceland, are also consistent with reduced thermohaline circulation in the region (Hansen et al. 2001). However, the North Atlantic cooling in mode 2 is countered and slightly reversed before 1940 by warming in that region indicated by modes 3 and 5.

Mode 3 most strongly affects the Northern Hemisphere and its time series has greatest variance in the 1900 to 1940 period. Mode 4 indicates strong interannual variations after about 1930, with weaker interannual variations before then. The spatial pattern and time series of mode 4 suggest that it may be associated with interannual and longer-period teleconnections in the Pacific (Zhang et al. 1997). The North Pacific pattern is also reminiscent of a Pacific-decadal oscillation pattern, but the mode indicates wider teleconnections into the Southern Hemisphere.

The mode 5 time series shows a warming trend similar to mode 2, but mostly affecting the tropics and Northern Hemisphere. The variance accounted for by mode 5 is much lower than
for mode 2, despite the similar time series. Both mode 5 and mode 3 have spatial loadings that most strongly affect the Northern Hemisphere and both indicate warm trends with their time series. However, most of the mode 3 warming occurs between about 1900 and 1940 while the mode 5 warming occurs over two periods: 1900-1940 and after about 1970. In addition, mode 5 indicates some local cooling in the North Pacific and the tropical Pacific. However, the variance associated with the local cooling in mode 5 is less than the interannual Pacific variations in modes 1 and 4.

Although the variance explained by some of these modes is similar, the test of North et al. (1982) shows that they are all different enough to be regarded as distinct modes. For all of these modes the HadISST time series show a similar but slightly weaker low-frequency variations than the ERSST time series. The correlation between ERSST and HadISST time series (Table 3) is lowest for mode 4 (correlation = 0.84) and highest for mode 2 (correlation = 0.96). For modes 1, 3, and 5 the correlations are 0.94, 0.88, and 0.88, respectively. These high correlations are encouraging considering differences between these two analyses, including different historical SST bias corrections, quality control and analysis procedures. They suggest that the dominant SST variations in the analysis are robust beginning in the late 19th century. For most modes the correlations are slightly higher in the second half of the period, when data are best and trends tend to be strongest, but even for the first half of the period correlations are strong (between 0.74 and 0.87).

Table 3. Correlation of ERSST and HadISST time series associated with the five rotated EOF eigenvectors, for the given periods.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.94</td>
<td>0.87</td>
<td>0.98</td>
</tr>
<tr>
<td>2</td>
<td>0.96</td>
<td>0.82</td>
<td>0.95</td>
</tr>
<tr>
<td>3</td>
<td>0.88</td>
<td>0.86</td>
<td>0.78</td>
</tr>
<tr>
<td>4</td>
<td>0.84</td>
<td>0.74</td>
<td>0.96</td>
</tr>
<tr>
<td>5</td>
<td>0.88</td>
<td>0.82</td>
<td>0.94</td>
</tr>
</tbody>
</table>

7. Conclusions

The extended reconstructed SST (ERSST) shown here is an improvement over the Smith et al. (1996) reconstruction (RSST) because of its longer period and its slightly greater coverage. The RSST is limited to the period 1950 on, and RSST anomalies are only computed for the region 45E-70E. Annual averages of anomaly spatial variance (Murphy and Epstein 1989) indicates how well each analysis represents the SST variations in a given year. The 60E-60EN spatial variance is computed for several SST analyses, and the ratio to ERSST is computed to show the relative variance of the other analyses (Fig. 11). For RSST the ratio is near 1.0, showing that a similar amount of variance is represented in each.

The ERSST filters data using a set of modes, and it also uses only the incomplete in situ sampling. Filtering with modes is designed to greatly reduce small-scale noise while allowing
large-scale signal to be represented. The Reynolds et al. (2002) analysis uses both satellite and in situ sampling and requires less filtering because of the more complete spatial coverage, and therefore its spatial variance ratio is greater than 1.0 for the brief overlap period (1982-1997). For the period when satellite data are available the HadISST analysis uses that data, and therefore its ratio is similar to the Reynolds et al. (2002) ratio for the common period. However, before the satellite period the HadISST spatial variance ratio is often larger. This generally larger ratio indicates that, relative to ERSST, HadISST retains more signal because of less filtering but may also retain more noise.

The signal/noise variance ratio is evaluated using the method of Thièbaux and Pedder (1987). With that method, the spatial correlations as a function of distance, for distances greater than zero, are computed and fit to a function. Here we fit the correlations to a Gaussian function, as in Reynolds and Smith (1994). The value of the function at zero distance is less than one due to noise in the analysis. If that zero-distance value is $A$ then the correlated/uncorrelated variance ratio is equal to $A/(1-A)$. If we assume that the correlated variance is signal and the uncorrelated variance is noise then this is also the signal/noise variance ratio. Here we compute that ratio for several periods and average it between 60E and 60W for ERSST, HadISST, and RSST.

Table 4 shows that the signal/noise ratios for ERSST is nearly constant for all periods beginning in the late 19th century. Compared to ERSST, the RSST signal/noise ratio is similar but slightly less. Since their variance is similar, this indicates that RSST is slightly more noisy than ERSST. The HadISST signal/noise ratio is largest in the early period, when data are most sparse. This may be because when data are sparse more filtering is needed to fill in the analysis, which also filters out more noise. The HadISST signal/noise variance ratio is similar to the ratio estimated for an earlier Met Office SST analysis by Folland et al. (1993). In all periods, the HadISST signal/noise ratios are less than for ERSST, indicating that it is a slightly more noisy analysis. Table 4 and Fig. 11 indicate that ERSST has the advantage over HadISST of being less noisy at the cost of a reduced signal.

```
Table 4. Average (60E-60W) signal/noise variance ratios for ERSST, HadISST, and RSST, for the given periods.

<table>
<thead>
<tr>
<th>Period</th>
<th>ERSST</th>
<th>HadISST</th>
<th>RSST</th>
</tr>
</thead>
<tbody>
<tr>
<td>1871-1900</td>
<td>34</td>
<td>26</td>
<td>--</td>
</tr>
<tr>
<td>1901-1930</td>
<td>32</td>
<td>25</td>
<td>--</td>
</tr>
<tr>
<td>1931-1960</td>
<td>33</td>
<td>21</td>
<td>--</td>
</tr>
<tr>
<td>1961-1990</td>
<td>34</td>
<td>18</td>
<td>30</td>
</tr>
<tr>
<td>1982-1997</td>
<td>34</td>
<td>21</td>
<td>29</td>
</tr>
</tbody>
</table>
```

Appendix A: Quality Control

The SSTs used in this study are screened by comparing individual observed SST anomalies to a local analysis of SST anomalies. Before screening a monthly SST climatology is formed on the 2E spatial grid, using all COADS SSTs for a period when sampling is dense (1961-1990). The climatology is median filtered to minimize the influence of outliers and interpolated spatially to fill regions not sampled in this period. Using this climatology, the SST
anomaly monthly standard deviation on the 2E grid is defined for the same period \( (\sigma_a) \). A monthly and 2E optimal interpolation (OI) analysis of the SST anomalies for this period is also computed (see, e.g., Reynolds and Smith 1994 for a description of OI). This OI analysis is performed using local data, from within a 10E square surrounding the 2E square. To keep extreme outliers from contaminating the local analysis, anomalies with a magnitude exceeding six \( \sigma_a \) are excluded from the OI. Since this period is densely sampled it approximates the best possible analysis. In periods with sparse sampling the local OI analysis will damp to zero anomaly as the available data are reduced. The difference between observed SST anomalies and this analysis in the 1961-1990 period is used to define an anomaly difference standard deviation \( (\sigma_d) \) over that well-sampled period. Note that because the OI analysis incorporates data from over a spatial region for the entire month, \( \sigma_d \) will not be zero although is should be less than \( \sigma_a \). These statistics are used to screen the SST observations.

Screening is done using the statistic

\[
Q = \frac{|T_a - A_a|}{\sigma} \tag{A1}
\]

where \( T_a \) is the observed individual SST anomaly, \( A_a \) is the local monthly OI analysis of SST anomalies, computed as described above, and \( \sigma \) is a standard deviation, described below. If \( Q \) exceeds a threshold, here set to 3, the individual observation is not used. The local OI analysis used to compute \( A_a \) produces a normalized error estimate \( (E_{OI}^2) \) which has a minimum value of 0 when data are dense, as approximated by the 1961-1990 period. In this case the appropriate \( \sigma \) to use in equation (A1) is \( \sigma_d \). The maximum \( E_{OI}^2 \) value is 1, which occurs when there are no data for analysis. In this case \( A_a \) damps to 0 anomaly and the appropriate \( \sigma \) to use in equation (A1) is \( \sigma_a \). The standard deviation to be used in equation (A1) for all other cases is computed from the normalized analysis error by

\[
\sigma^2 = E_{OI}^2 \sigma_a^2 + (1 - E_{OI}^2) \sigma_d^2. \tag{A2}
\]

Appendix B: High-Frequency Analysis

In the analysis methods of Smith et al. (1998), data increments are defined at each spatial point, \( x \), as

\[
I(x) = D(x) - G(x), \tag{B1}
\]
where $D$ is the data and $G$ the first guess. The first guess is the previous month’s analysis. Modes of that analysis are damped when they are not supported by the data,

$$G(x) = \sum_{m=1}^{M} \left[ \Delta_m + (1 - \Delta_m)c_m \right]w_m \psi_m(x),$$

(B2)

where $\Delta_m$ is 1 if mode $m$ is supported by sampling and 0 otherwise, $c_m$ is the one-month autocorrelation for mode $m$, $w_m$ is the weight obtained for mode $m$ by projecting the previous month’s analysis onto the set of modes, and $\psi_m$ is the spatial covariance-based mode $m$. The total number of modes is $M$. Whether or not to use a mode is determined from the fraction of that mode’s variance supported by the sampling,

$$f(m) = \frac{\sum_{x \in A} \delta(x) \psi_m^2(x)a(x)}{\sum_{x \in A} \psi_m^2(x)a(x)},$$

(B3)

where $a(x)$ is the relative area represented by point $x$, and $\delta(x) = 1$ if point $x$ is sampled and 0 otherwise. If this variance sampling is less than a critical value then the mode is not used. The critical value of variance sampled for each mode is determined using cross-validation tests. In our analysis we use a critical value of 0.15, as discussed in section 3b.

Those modes that are adequately sampled by the data are used to compute the increment by fitting the data increments to the set of modes, as in Smith et al. (1996). The analysis increment is defined by a weighted sum of the sampled modes, where the increment weights, $w_m$, are chosen to minimize the error of the increment analysis. The total high-frequency anomaly analysis is

$$H(x) = \sum_{m=1}^{M} \psi_m(x)[\Delta_m (w_m + w_m c_m) + (1 - \Delta_m)c_m w_m].$$

(B4)

Acknowledgments

We thank Scott Woodruff for assistance with the COADS data, and Xiao-Wei Quan for useful discussions about the data quality control. Suggestions by Tom Karl helped with the development of the error-estimation method. We thank Vern Kousky, Matt Menne, and Ned Guttman for comments on the analysis. We also thank the NOAA office of Global Programs, which provided support for some of this work.

References


FigureCaptions

Fig.1. Base 10 log of the number of annual global SST observations, for individual observations (solid line) and 2 E super observations (dashed line).

Fig.2. Percentage of global observations removed by QC, for individual observations (solid line) and 2 E super observations (dashed line).

Fig.3. Percentage of the ocean for which a 10 E annual average could be computed.

Fig.4. Global RMSE of the ERSST (EC).

Fig.5. Annual average of ERSST spatial variance, EC^2 (a), and the annual average number of modes used (b).

Fig.6. Annual and spatial averages of SST anomalies from unanalyzed COADS, ERSST, and HadISST, averaged 23 EN-60 EN. The 95% confidence interval is shown in the lower panel. Units are EC.

Fig.7. Annual and spatial averages of SST anomalies from unanalyzed COADS, ERSST, and HadISST, averaged 23 ES-23 EN. The 95% confidence interval is shown in the lower panel. Units are EC.

Fig.8. Annual and spatial averages of SST anomalies from unanalyzed COADS, ERSST, and HadISST, averaged 60 ES-23 ES. The 95% confidence interval is shown in the lower panel. Units are EC.

Fig.9. Annual and spatial averages of SST anomalies from unanalyzed COADS, ERSST, and HadISST, averaged 60 ES-60 EN. The 95% confidence interval is shown in the lower panel. Units are EC.

Fig.10. The first five rotated EOFs of ERSST, explaining 57% of the variance. The time series are from projecting the ERSST data onto the eigenvectors (solid line) and from projecting the HadISST data (dashed line). The percent of ERSST variance explained by each mode is indicated.

Fig.11. Annual averages of spatial variance ratios for the 60 ES-60 EN region. Ratios are of the analyses to ERSST for: HadISST, Smith et al. (1996, RSST), and Reynolds et al. (2002, ROI).