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

THOMAS M. SMITH AND RICHARD W. REYNOLDS

*National Climatic Data Center, Asheville, North Carolina*

ABSTRACT

A monthly extended reconstruction of global SST (ERSST) is produced based on Comprehensive Ocean–Atmosphere Data Set (COADS) release 2 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 1880s because data are too sparse to resolve enough modes for that period. Error estimates indicate that except in the North Atlantic 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 decreases through most of the period and is smallest after 1950.

The large-scale variations of ERSST are broadly consistent with those associated with the Hadley Centre Global Sea Ice and Sea Surface Temperature (HadISST) reconstruction produced by the Met Office. There are differences due to both the use of different historical bias corrections as well as 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 that 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 nineteenth century. Improvements include additional data from a new version of the Comprehensive Ocean–Atmosphere Data Set (COADS) release 2, 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 of Smith et al. (1996). The Smith et al. 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 12 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

*Corresponding author address:* Dr. Thomas Smith, NCDC/NESDIS/NOAA, 151 Patton Ave., Asheville, NC 28801.
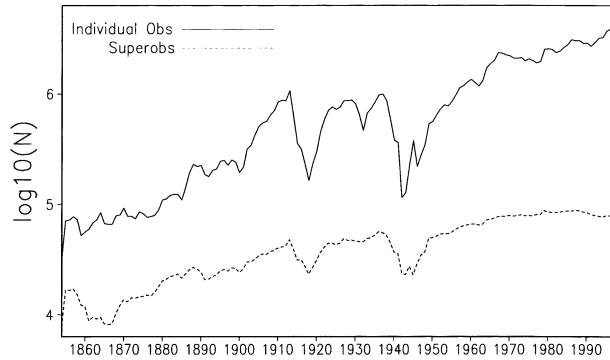E-mail: tom.smith@noaa.gov

FIG. 1. Base 10 log of the number of annual global SST observations, for individual observations (solid line) and 2° superobservations (dashed line).
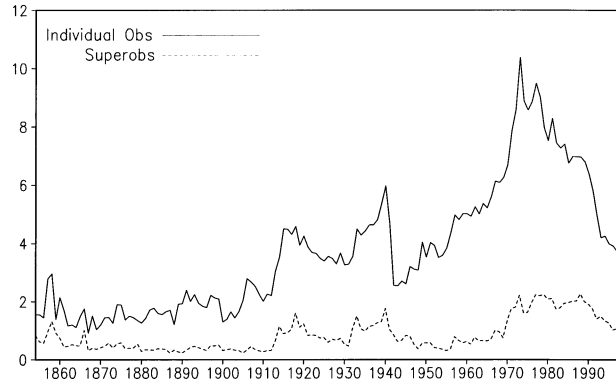


FIG. 2. Percentage of global observations removed by QC, for individual observations (solid line) and 2° superobservations (dashed line).

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. analysis develops statistics from in situ data, which have large gaps in the Southern Hemisphere. In addition, its base period may not span all interdecadal 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 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 COADS release 2 (Slutz et al. 1985; Woodruff et al. 1998), with updates through 1997. We average the SSTs to superobservations, defined as monthly averages on a 2° grid, with grid centers on 88°S, 86°S, . . . , 88°N by 0°, 2°E, . . . , 2°W. This grid is offset by 1° from the standard 2° 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 1915–20 period and the 1940s. The annual number of global SST super observations (dashed line) is less variable, but there are still relative minimums in the same two early twentieth-century periods. These time series include all data, in-

cluding suspect observations that are not used for the reconstruction. Below we discuss the quality control system to remove these suspect observations.

### a. 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 SSTs 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 increase. Some of the increased percentage of flagged observations is due to an increased frequency of flagging at midlatitudes 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, midlatitude 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 near 1980. North of 70°N, typically 50% or more individual observations are flagged by the SST QC.

With superobservations, there is reduced loss due to QC. That is expected, since a superobservation may be formed even if several of its individual observations are discarded. The greatest percentage of lost superobservations occurs after 1970 when about 2% are lost. This increased loss of superobservations 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 superobservations are lost due to the QC.

### b. 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 (see also Rayner et al. 2003, hereafter REA).

We average the monthly 1982–2000 Reynolds et al. (2002) analysis to the same 2° superobservation 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 16 years of this study's analysis period with an additional three years (1998–2000). Because there are no data gaps in the 19-yr climatology, it will resolve SST features globally. Our historical anomaly analysis is computed with respect to this 19-yr climatology. If desired, the anomaly base period can easily be readjusted to any subperiod of the historical analysis (e.g., Smith and Reynolds 1998).

### c. 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 World War II.

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 19 years of SST anomalies, and thus may not adequately span interdecadal variations. Both low- and high-frequency variations are reconstructed using screened COADS 2° superobservations.

### a. 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 10° 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 10° grid covers the globe equatorward of 75° latitude. Poleward of 75° there 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. In the steps that follow multiple filtering steps are used to obtain the low-frequency analysis. These steps are needed to allow global coverage from sparse data. Any signal not retained in this stage will be restored in the high-frequency analysis that follows.

To separate the low-frequency analysis we first form monthly 10° anomalies for squares that contain at least three 2° superobservations, and at least nine individual in situ observations. The 2° superobservations are weighted by their relative area and by their relative sampling. Second, for each calendar year annual 10° anomalies are formed by averaging the monthly 10° anomalies at each location, provided that at least four monthly 10° anomalies could be defined for the year. These 10° annual anomalies are then filtered using zonal and meridional three-point binomial filters, to further reduce small-scale variations. The binomial filtering fills in un-
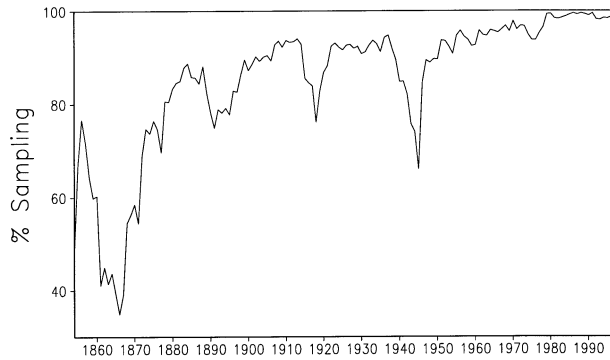
FIG. 3. Percentage of the ocean for which a 10° annual average could be computed.

defined annual 10° regions with filtered values from adjacent defined 10° regions, which slightly expands the spatial extent of the annual field. The percent of the total area for which annual 10° 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-yr 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-yr anomaly for most 10° regions for each calendar year. The final step is to set remaining undefined 15-yr 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 5-yr binomial filter.

The 15-yr moving window filter removes interannual and shorter period variations. The signal from these variations will be restored by the high-frequency analysis, which is discussed in the next section. The high-frequency analysis is based on a recent 19-yr period. Because the temporal filter is a 15-yr moving window, the first and last eight years of the low-frequency analysis are estimated from shorter periods truncated at the ends. This is not critical because the last eight years are within the base period of the high-frequency analysis. The first eight years are also not critical because there is very little data in this part of the record and, as will be discussed later, there is limited information in our analysis prior to 1880.

To test our choice of a 15-yr filter, we compare the time series of the leading EOF of the 15-yr analysis with leading EOF time series from similar analyses, with lengths of 5, 11, 19, and 25 yr. 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-yr analysis time series has large variations with periods

8–15 yr and thus is too short. The 11–25-yr analyses all indicate similar variations, and either could be used. However, the 11-yr analysis includes some weak variations with periods 10–15 yr, so a slightly longer period is preferred. The 25-yr analysis begins to damp variations with periods greater than 30 yr, so the shorter 15-yr period is better. The filtering used here greatly reduces variations with spatial scales smaller than 10° and timescales less than 15 yr.

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 use a different bias correction by adjusting our final analysis by the bias-correction difference.

### b. High-frequency analysis

Our analysis of high-frequency anomalies uses the method of Smith et al. (1998). Our 2° 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 superobservations 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 the variance associated with unsupported modes is damped using the mode's autocorrelation. The high-frequency modes are computed from anomalies that are called high frequency because their mean is removed over the base period. Those data are not otherwise filtered to remove low-frequency variations. Note that 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 that 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

TABLE 1. Number of EOT modes with 1-month lag autocorrelation (ac1) in the given range. There are a total of 75 modes.

| Range | $ac1 \leq 0.4$ | $0.4 < ac1 \leq 0.6$ | $0.6 < ac1 \leq 0.8$ | $0.8 < ac1$ |
|---|---|---|---|---|
| Number | 1 | 19 | 47 | 8 |

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 $10°$–$15°$ in latitude by about $30°$ 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 another 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 $2°$ version of the Reynolds et al. (2002) anomaly increments, both spatially (zonally and meridionally) and temporally, so that our patterns will not reflect grid-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 $60°S$, north of $64°N$, the Caspian, Black, and Baltic Seas, the 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. By definition the variance of the EOT mode decreases with the order of the mode. In addition, the spatial size of the mode usually also decreased with the order. Because we want to smooth the data using these modes, we need to truncate the set of EOTs. This decision is subjective. We found that 69 EOTs account for most global incre-

ment variations, with higher modes describing more localized features. These local features (i.e., with spatial scales smaller than about $10°$) are unlikely to be resolved by historical sampling and may increase analysis noise if they were used. Thus, we exclude those higher modes from the analysis.

Because there is some sampling in the Arctic, especially in boreal summer months, we computed a second set of EOT modes with the sampling region restricted to north of $64°N$ 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 1982–97 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) and 0.94 (for the tropical Pacific), with values commonly between 0.6 and 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 timescales will have greater persistence and thus can make greater use of data from months other than the analysis month.

We defined when the sampling was adequate for each mode using the percent of variance supported by the sampling for each mode. High-frequency anomalies associated with adequately sampled modes are updated using the data, while anomalies associated with undersampled modes are damped. We wish to avoid situations in which a mode is only sampled outside of its center of action. In those situations, the variance sampled may be as high as 10%–12%, so we know that at least that sampling for each mode is needed. 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–

TABLE 2. Global rmse (°C) from cross-validation tests for different sampling years and critical values (% sampling).

| Year | 12% | 15% | 18% | 21% |
|------|------|------|------|------|
| 1860 | 0.575 | 0.580 | 0.591 | 0.597 |
| 1918 | 0.510 | 0.513 | 0.526 | 0.528 |
| 1942 | 0.504 | 0.497 | 0.497 | 0.505 |

86 SST anomalies with modes derived from the 1988–2000 Reynolds et al. (2002) analysis. The 5-yr validation period spans interannual variations similar to what may have occurred in the past, and the 13-yr 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.

## 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

$$E^2 = E_{LFS}^2 + E_{HFS}^2 + E_{An}^2. \quad (1)$$

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 dataset. Since the globally complete data must simulate SST trends, we use output from the Geophysical Fluid Dynamics Laboratory 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-yr runs (1866–1990) that are 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 Hadley Centre Global Sea Ice and SST (HadISST; see REA) 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 2° SST anomalies subsampled 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 set of modes was chosen subjectively, to exclude small-scale variations (less than 10° spatially). This requires fewer modes for the model than for the observed SST anomalies. The three sets of full and analyzed model SSTs are averaged. For both the full and analyzed ensemble means, the SSTs are temporally filtered using a 21-yr 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. Some comparisons of the model and ERSST low-frequency anomalies are given by Smith et al. (2002).

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 $F_m \approx aR_m$. 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. Damp-

ing 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

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

where the angle brackets denote averaging over the period, here defined as 21 yr centered on the year of interest. Since the amount of damping depends on sampling, we 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

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

We further smooth this low-frequency mse estimate by averaging over an 11-yr moving period, centered on the year of interest. Since the model may have its own low-frequency errors different from the ERSST estimate, 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 1982–86 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 subsampled to match sampling from each year in analysis period. Validation data are the same data, but without subsampling. 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 subsampling 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

$$E_{HFS}^2 = \langle \varepsilon_{HFS}^2 \rangle, \quad (3)$$

where the averaging in (3) is for the same period used in (2).

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. In these uncertainty estimates, random data error is assumed to be included in the analysis error. Projecting the observed data onto a set of physical modes should greatly reduce that random error component. To estimate the remaining analysis errors, we compare ERSST to two other analyses produced by the U.K. Met Office: Global Sea Ice and SST (GISST; Parker et al. 1994) and HadISST (REA). Those two analyses use different quality control, some differ-
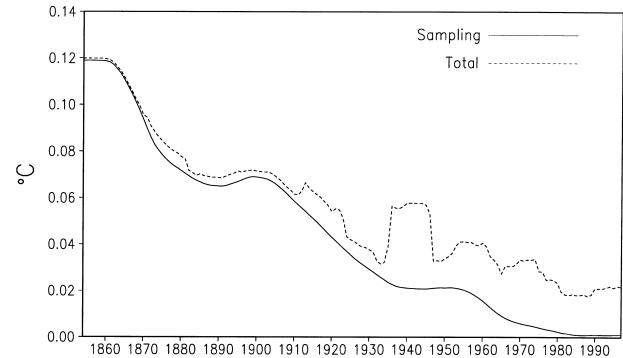


FIG. 4. Global rmse of the ERSST (°C).

ent input data, different historical bias correction methods, and different analysis methods compared to ERSST. The ERSST analysis method mse is defined as

$$E_{An}^2 = \langle (ERSST - Avg)^2 \rangle, \quad (4)$$

where Avg is the average of ERSST, GISST, and HadISST. The averaging period in (4) is as in (2) and (3).

This mse component can be further divided into the bias component of mse, $B_{An}^2 = (\langle R \rangle - \langle Avg \rangle)^2$, and the nonbias 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 cooling water intakes only, for the 1994–97 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°C. 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. Error estimates from the local time-averaging period, the same period used to compute (2), (3), and (4), are used in the regression of error components. 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 makes 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, $E$ in (1), is largest early in the analysis period (Fig. 4, solid line). Before the 1880s sampling severely limits the value of any SST analysis, and in the early twentieth 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°C in all periods. Total error also increases around 1940, near the time when the historical

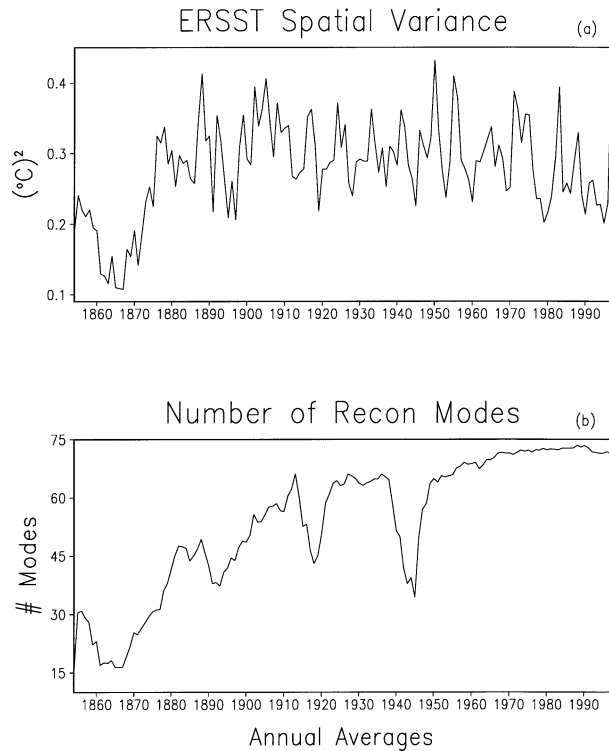FIG. 5. (a) Annual average of ERSST spatial variance (°C)² and (b) the annual average number of modes used.



FIG. 6. Annual and spatial averages of SST anomalies from the original raw COADS, ERSST, and HadISST averaged between 23° and 60°N. The 95% confidence interval is shown in the lower panel; units are °C.

bias corrections are largest and then decrease to zero. These global SST errors for the pre-1880 averages are about 50% larger than the global SST errors estimated by Folland et al. (2001) using different methods, and thus this estimate may be an overestimate of 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 we did not attempt to compute ERSST before 1854. Annual averages of the ERSST global spatial variance (Murphy and Epstein 1989) indicate 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 (°C²), 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. The drop in variance indicates excessive damping before the mid-1870s. Nearly all of this reduction is 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
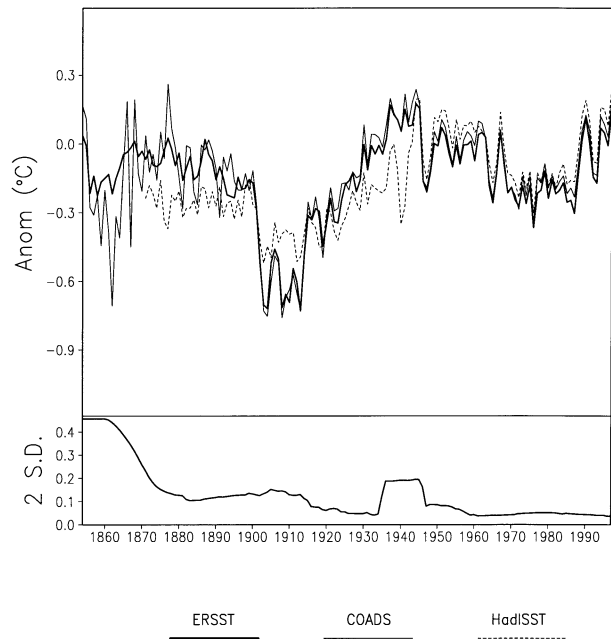
average number of modes used in 1893 (37 modes), 1918 (43 modes), and 1945 (34 modes) due to dips in sampling in those years. However, for each of those years the spatial variance remains relatively strong.

For comparisons, annual and spatial averages of SST anomalies for several regions are computed using the original raw COADS release 2 SST superobservations, ERSST, and the HadISST analysis of REA. The raw COADS superobservations are the same data used for analysis in ERSST, with the same quality control and SR02 historic bias correction. 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°–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 1900–15. 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 twentieth century, but they are striking considering the relatively good Northern Hemisphere sampling in that period.

In the 23°S–23°N region (Fig. 7) all three averages are more similar over most of the period, but the raw COADS
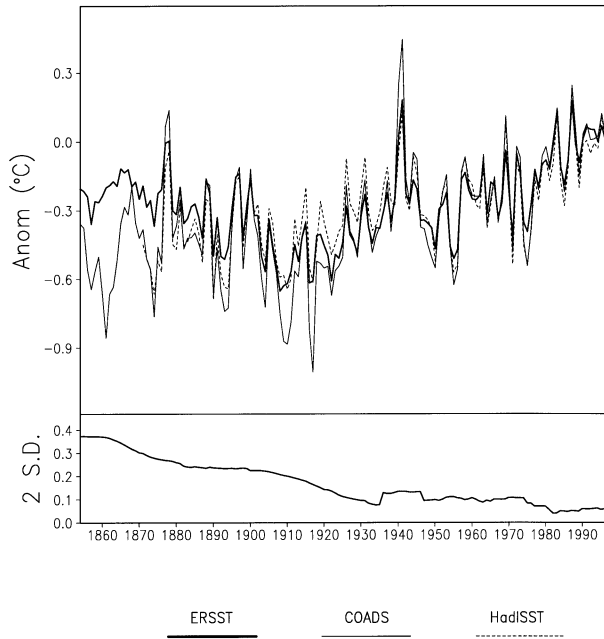
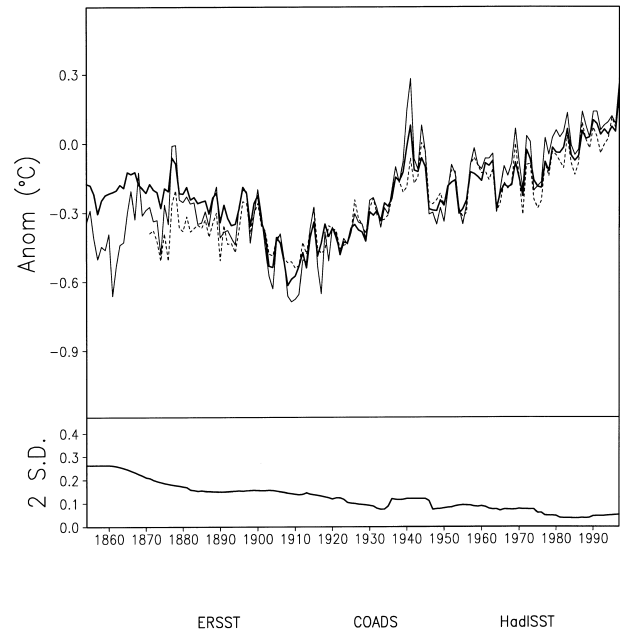FIG. 7. As in Fig. 6 but averaged between 23°S and 23°N.



FIG. 9. As in Fig. 6 but averaged between 60°S and 60°N.

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



FIG. 8. As in Fig. 6 but averaged between 60° and 23°S.

cooler in that period, suggesting that other analysis differences are also important.

For the 60°–23°S region (Fig. 8) all three anomalies show a large warming trend beginning about 1930. Note that the Southern Hemisphere uncertainty is largest in the 1915–60 period. In that period the Southern Hemisphere sampling is reduced, first by the opening of the Panama Canal in 1914 and later by World War II. 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 that in the Northern Hemisphere where overall cooling occurred between 1950 and 1985, preceded and followed by warming. Note that the COADS SST anomalies are slightly warmer than ERSST (and HadISST) in the Southern Hemisphere after about 1960 due to some large anomalies in the region that are too sparse and irregular to be analyzed by our methods.
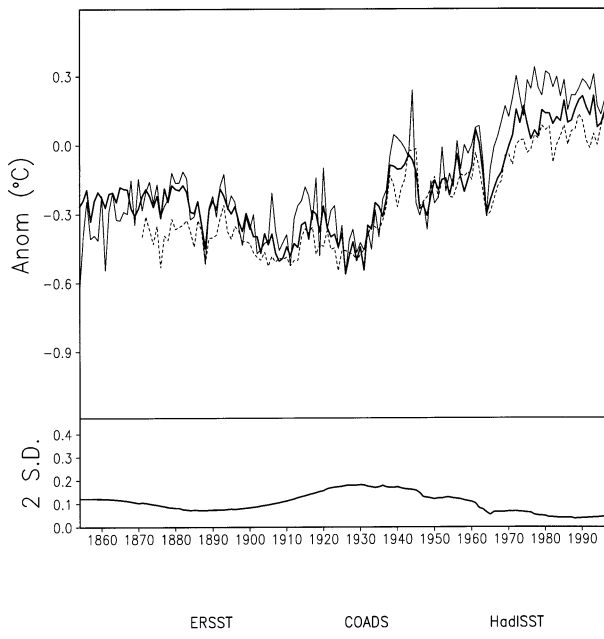
For the near-global area (Fig. 9), there is great similarity among all three after 1900. The uncertainty limits show that the near-global trend over the twentieth century is about 0.6° ± 0.2°C. In the 1982–97 overlap period, the Reynolds et al. (2002) average anomalies are similar to the ERSST and HadISST 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
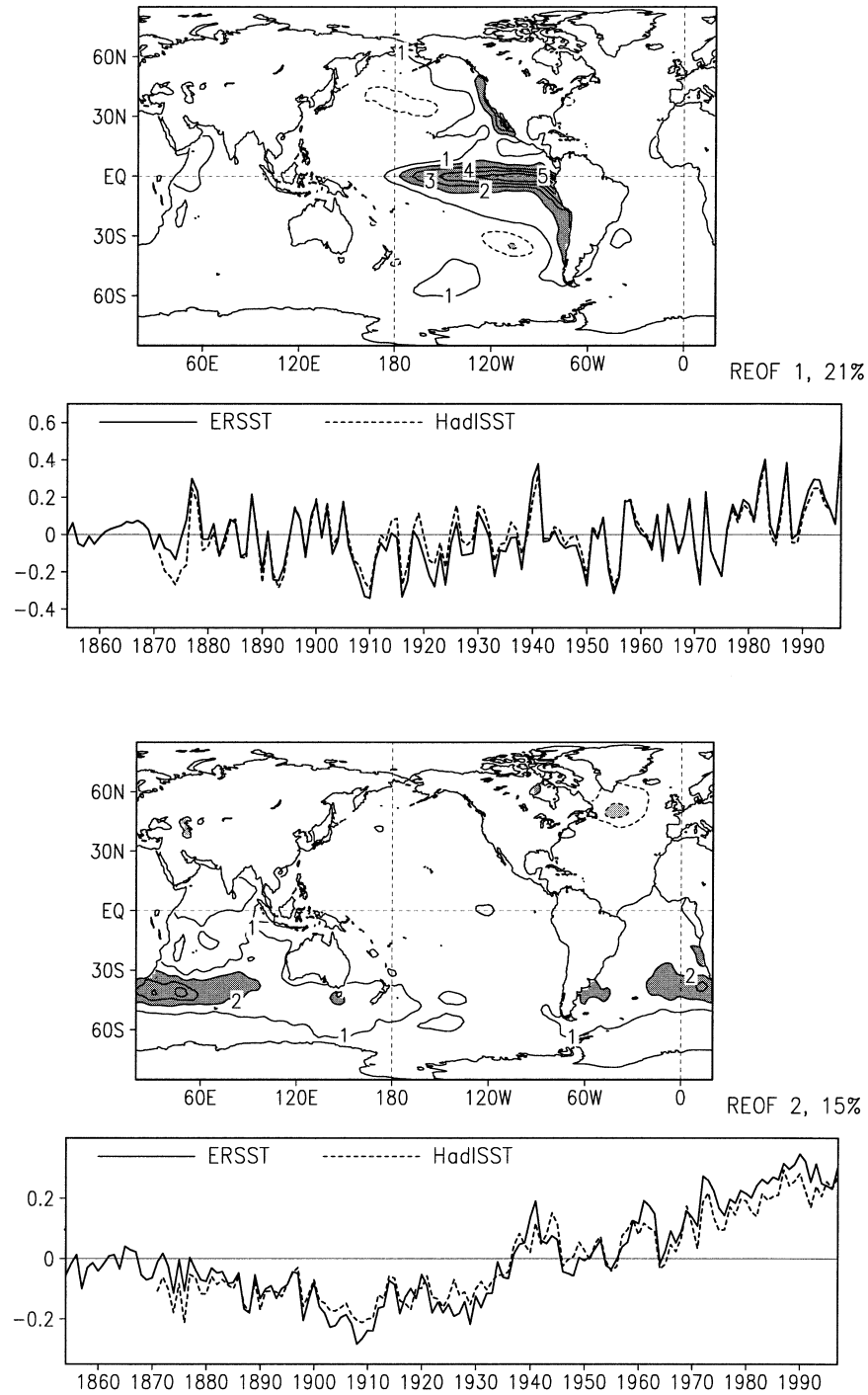
FIG. 10. The first five rotated EOFs of ERSST, explaining 57% of the variance of the annual SSTA. The time series are from projecting the ERSST data onto the eigenvectors (solid line) and from projecting the HadISST data (dashed line). The percent of ERSST variance explained by each mode is indicated.

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
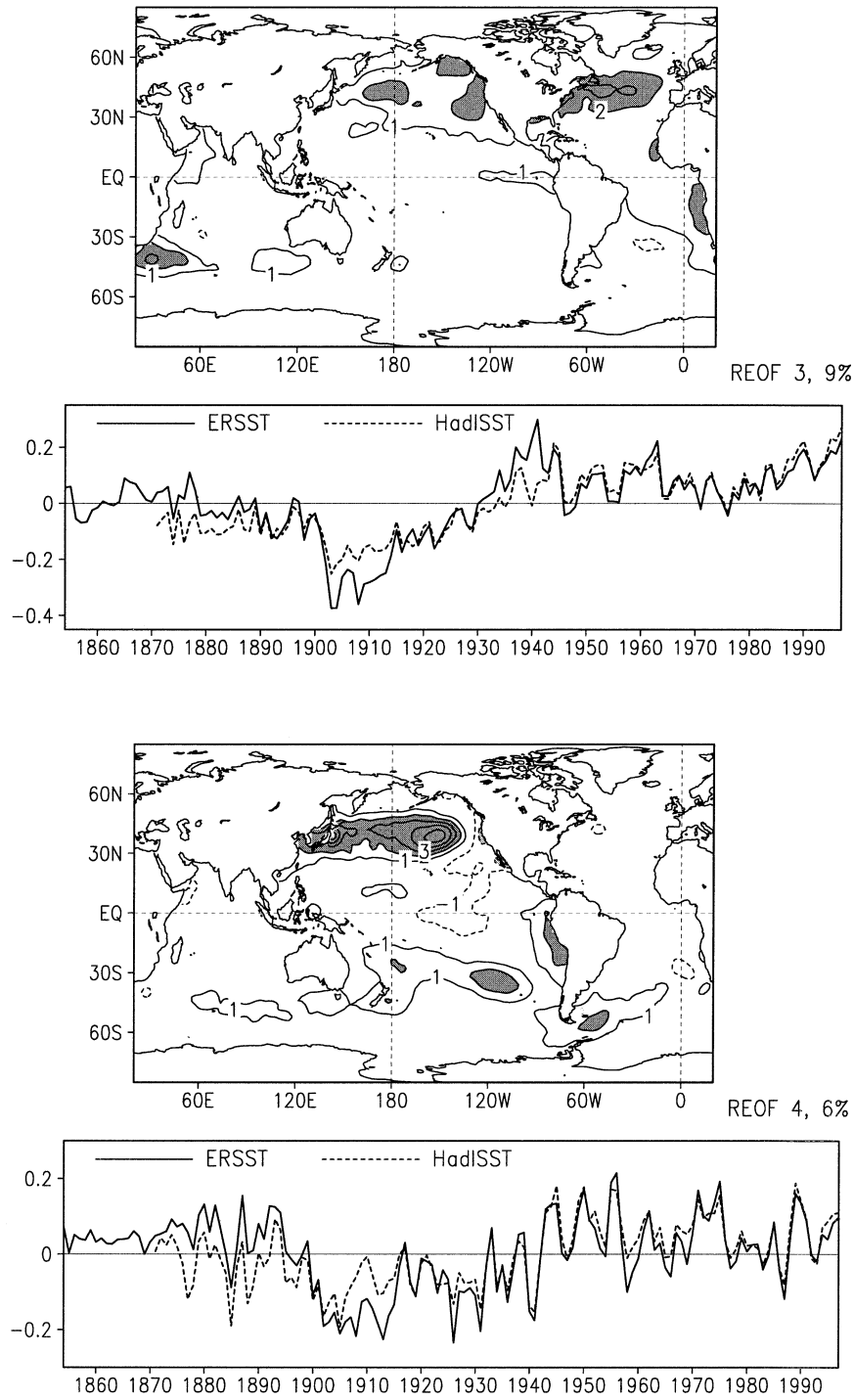
FIG. 10. (*Continued*)

almost 60% of the variance of the annual-average SST anomalies.

All modes have small amplitudes before the mid-1870s followed by greater and more uniform variance afterward, consistent with the change in spatial variance about that time (Fig. 5). That is because the sparse sampling before 1870 does not support the analysis of strong anomalies, greatly damping ERSST. The rotated EOF time series are computed by projecting both the ERSST and HadISST anomalies onto the same ERSST-based eigenvectors. Consistencies between the two time series are an indication of the degree to which their variations are the same.
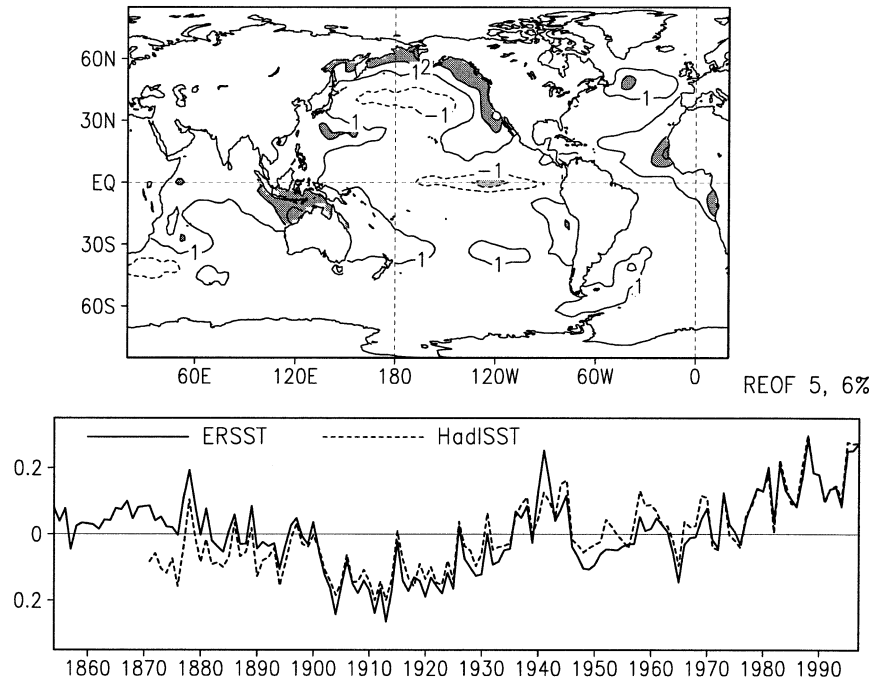
REOF 5, 6%

FIG. 10. (*Continued*)

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 1940–41 warm episode, there is a tendency for more cool episodes in the 1905–60 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. Since the late 1970s there have been several large warm episodes, including 1982–83, 1986–87, 1992, and 1997. Such strong warm episodes are less frequent in earlier periods, although there are strong warm episodes prior to 1980. This mode appears to contain both high- and low-frequency variations, which were separated in the analysis of Zhang et al. (1997).

The most important trend mode is mode 2, which indicates Southern Hemisphere warming throughout the twentieth 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. A similar mode was also identified by Cai and Whetton (2001), who used an earlier version on the Met Office SSTs for their analysis. 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 twentieth 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 the 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 slowdown 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, especially in the 1900–40 period. It is similar to the SST EOF mode 2 of Yasunaka and Hanawa (2002), although there are differences do to their use of winter only SSTs and a different COADS-based analysis. They identify that mode as associated with the Arctic Oscillation. 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. This mode is similar to the Yasunaka and Hanawa (2002) mode 1. They associate their

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

| Mode | 1854–1997 | 1854–1925 | 1926–97 |
|------|-----------|-----------|---------|
| 1 | 0.94 | 0.87 | 0.98 |
| 2 | 0.96 | 0.82 | 0.95 |
| 3 | 0.88 | 0.86 | 0.78 |
| 4 | 0.84 | 0.74 | 0.96 |
| 5 | 0.88 | 0.82 | 0.94 |

modes 1 and 2 (similar to our REOF modes 3 and 4) with regime shifts in SST. This analysis suggests that such regime shifts may be detectable back to the late nineteenth century.

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: 1900–40 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 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 nineteenth century. For most modes the correlations are slightly higher in the second half of the period (with values of 0.78–0.98), 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).

## 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 greater spatial coverage. The RSST is limited to the period 1950 on, and RSST anomalies are only computed for the region 45°S–70°N. Annual averages of anomaly
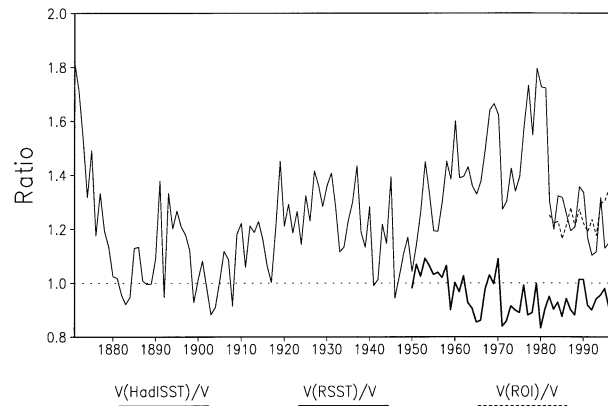


FIG. 11. Annual averages of spatial variance ratios for the 60°S–60°N region. Ratios are of the analyses to ERSST for HadISST, Smith et al. (1996, labeled RSST), and Reynolds et al. (2002, labeled ROI).

spatial variance (Murphy and Epstein 1989) indicates how well each analysis represents the SST variations in a given year. The 60°S–60°N 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 a large-scale signal to be represented. As indicated by Fig. 5a, ERSST variance after the late 1870s reflects interannual variations, but it is otherwise fairly stationary. 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 (1982–97). 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. In 1949 the HadISST analysis spatial resolution was changed from 4° to 2° (REA). This may explain the generally higher HadISST/ERSST ratios after 1949.

The signal/noise variance ratio is evaluated using the method of Thiébaux 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

TABLE 4. Average (60°S–60°N) signal/noise variance ratios for ERSST, HadISST, and RSST, for the given periods.

| Period | ERSST | HadISST | RSST |
|---|---|---|---|
| 1871–1900 | 34 | 26 | — |
| 1901–30 | 32 | 25 | — |
| 1931–60 | 33 | 21 | — |
| 1961–90 | 34 | 18 | 30 |
| 1982–97 | 34 | 21 | 29 |

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 60°S and 60°N 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 nineteenth 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 U.K. 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. In the HadISST analysis (REA) the initial analysis is done using reduced space interpolation (Kaplan et al. 1997), which should give a similar signal/noise ratio as that using our EOT fitting procedure. However, after this stage is complete, the original data are reintroduced and then smoothed. The reintroduction tends to decrease the HadISST signal to noise ratio. Table 4 and Fig. 11 indicate that ERSST has the advantage over HadISST of being less noisy at the cost of a reduced signal. The higher filtering of ERSST noise is due to projection of data onto physical modes, which filters out most random noise. As discussed above, the HadISST analysis more closely follows the available data and therefore can include variations not associated with a fixed set of modes. This additional variance can be associated with both signal and noise.

REA have developed fields of sea ice concentrations for HadISST. These concentrations have been adjusted to make the in situ data and satellite data as homogeneous as possible over the historic period of record. The sea ice concentrations are converted into SSTs using regional climatological relationships, as described in REA, and then merged with the completed SST analysis. The procedure was also used in the improved optimal interpolation version 2 (Reynolds et al. 2002). We plan to also use a similar method to add sea ice to the ERSST analysis in the near future. Both HadISST and ERSST with sea ice will be tested as boundary conditions to atmospheric general circulation models under Climate of the Twentieth Century Project, sponsored by the International Re-

search Program on Climate Variability and Predictability. This may help to resolve the importance of the differences in the analyses, including the signal/noise differences. The ERSST data are available online at http://www.ncdc. noaa.gov/oa/climate/research/sst/sst.html.

## 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. Here the quality control (QC) method is outlined.

### a. Define a climatology

Before screening a monthly SST climatology is formed on the 2° spatial grid, using all COADS release 2 SST superobservations for a period when sampling is dense (1961–90). The climatology is median filtered to minimize the influence of outliers and interpolated spatially to fill regions not sampled in this period. The climatology could have been defined using other data in addition to COADS. However, during the base period the sampling is sufficient to define a global climatology.

### b. Define QC statistics

Using this climatology to define anomalies, the SST anomaly monthly standard deviation on the 2° grid is computed for the same 1961–90 base period ($\sigma_a$). A monthly and 2° 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 10° square surrounding the 2° 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 1961–90 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.

### c. Data screening

Screening is done using the statistic

$$Q = \frac{|T_a - A_a|}{\sigma}, \quad (\text{A1})$$

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 ($E_{OI}^2$) that has a minimum value of 0 when data are dense, as approximated by the 1961–90 period. In this case the $\sigma$ to use in Eq. (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 Eq. (A1) is $\sigma_a$. The standard deviation to be used in Eq. (A1) for all other cases is computed from the normalized analysis error by

$$\sigma^2 = E_{OI}^2 \sigma_a^2 + (1 - E_{OI}^2)\sigma_d^2. \quad (\text{A2})$$

If anomalies are locally supported by other consistent anomalies then the numerator of A1 is reduced, which makes the observation more likely to be accepted (holding the denominator equal). But, if the local supporting observations are dense, then the denominator will be reduced and $Q$ increased, so in that case the observation must be closer to the local analysis to be accepted.

### APPENDIX B

### High-Frequency Analysis

For this analysis we adapt the methods of Smith et al. (1998) for analysis of global SST anomalies. These methods analyze the anomaly by computing time increments of the anomaly at each data point, analyzing the increments to fill a regular grid, and adding the analyzed increment back onto the previous month's anomaly. As discussed by Thiébaux (1997), an increment analysis can provide more accuracy if statistics of the increments can be accurately defined. This analysis of increments incorporates both spatial and temporal correlation information, both of which are helpful for analyzing SST anomalies. Below an outline of the method is given.

### a. Define anomaly increments

In the analysis data increments are projected onto a set of spatial covariance modes. We define data increments at each spatial point, $x$, as

$$I(x) = D(x) - G(x), \quad (\text{B1})$$

where $D$ is the data and $G$ the first guess. The first guess is the previous month's anomaly analysis projected onto the set of spatial modes.

### b. Define the guess using modes

The first guess is defined as a linear combination of the spatial modes. Since the previous month's anomaly is defined everywhere, that anomaly may be projected onto the full set of modes. Although these are increment modes, the anomalies are built from linear combinations of these modes and thus they span both the increment and anomaly variance. By representing the guess as modes we are able to damp the variance associated with modes that are not adequately sampled. The first guess weights are the set of weights that minimize the error of the first guess when it is represented as the weighted sum of the modes.

If the first guess weights for each mode are $wg_m$, for $m = 1, 2, \ldots, M$ (where $M$ is the total number of modes), then we may represent the guess as

$$G(x) = \sum_{m=1}^{M} [\Delta_m + (1 - \Delta_m)c_m]wg_m\psi_m(x), \quad (\text{B2})$$

where $\Delta_m = 1$ if mode $m$ is supported by sampling and 0 otherwise, $c_m$ is the one-month autocorrelation for mode $m$, and $\psi_m$ is the spatial covariance-based mode $m$. Thus, if sampling is erratic in time, with adequate sampling one month, and poor sampling the next, damped persistence of that mode's anomaly is used to spread information temporally. However, if sampling for a mode remains inadequate for $n$ months, then the variance for that mode will be damped by a factor of $c_m^n$.

### c. Define adequate sampling

Whether or not to use a mode to analyze increments is determined from the fraction of that mode's variance supported by the sampling

$$f(m) = \frac{\sum_{x \in A} \delta(x)\psi_m^2(x)a(x)}{\sum_{x \in A} \psi_m^2(x)a(x)}, \quad (\text{B3})$$

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 to analyze increments and $\Delta_m = 0$. 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.

### d. Anomaly analysis

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, $H$, is the sum of the increment analysis and the first guess,

$$H(x) = \sum_{m=1}^{M} \psi_m(x)[\Delta_m(w_m + wg_m)$$
$$+ (1 - \Delta_m)c_m wg_m]. \quad \text{(B4)}$$

This analysis is computed for each time beginning at the first month and going to the last, which includes temporal information in one direction. To use temporal information in the other direction, we then compute $H$ going backward in time, from the last month to the first and average the forward and backward analyses.

## REFERENCES

Bottomley, M., C. K. Folland, J. Hsiung, R. E. Newell, and D. E. Parker, 1990: *Global Ocean Surface Temperature Atlas "GOS-STA."* Joint project of the U.K. Meteorological Office and the Massachusetts Institute of Technology, Her Majesty's Stationery Office, 20 pp. and 313 plates.

Cai, W., and P. H. Whetton, 2001: Modes of SST variability and the fluctuation of global mean temperature. *Climate Dyn.,* **17,** 889–901.

Delworth, T. L., R. J. Stouffer, K. W. Dixon, M. J. Spelman, T. R. Knutson, A. J. Broccoli, P. J. Kushner, and R. T. Wetherald, 2002: Review of simulations of climate variablility and change with the GFDL R30 coupled climate model. *Climate Dyn.,* **19,** 555–574.

Duffy, P. B., C. Doutriaux, I. K. Fodor, and B. D. Santer, 2001: Effect of missing data on estimates of near-surface temperature change since 1900. *J. Climate,* **14,** 2809–2814.

Folland, C. K., and D. E. Parker, 1995: Correction of instrumental biases in historical sea surface temperature data. *Quart. J. Roy. Meteor. Soc.,* **121,** 319–367.

——, ——, and F. E. Kates, 1984: Worldwide marine surface temperature fluctuations 1856–1981. *Nature,* **310,** 670–673.

——, R. W. Reynolds, M. Gordon, and D. E. Parker, 1993: A study of six operational sea surface temperature analyses. *J. Climate,* **6,** 96–113.

——, and Coauthors, 2001: Global temperature change and its uncertainties since 1861. *Geophys. Res. Lett.,* **28,** 2621–2624.

Hansen, B., W. R. Turrell, and S. Østerhus, 2001: Decreasing overflow from the Nordic seas into the Atlantic Ocean through the Faroe Bank channel since 1950. *Nature,* **411,** 927–930.

Kaplan, A., Y. Kushnir, M. A. Cane, and M. B. Blumenthal, 1997: Reduced space optimal analysis for historical data sets: 136 years of Atlantic sea surface temperatures. *J. Geophys. Res.,* **102,** 27 835–27 860.

——, M. A. Cane, Y. Kushnir, A. C. Clement, M. B. Blumenthal, and B. Rajagopalan, 1998: Analyses of global sea surface temperature 1856–1991. *J. Geophys. Res.,* **103,** 18 567–18 589.

Levitus, S., J. I. Antonov, T. J. Boyer, and C. Stephens, 2000: Warming of the world ocean. *Science,* **287,** 2225–2229.

Manabe, S., and R. J. Stouffer, 1994: Multiple-century response of a coupled ocean–atmosphere model to an increase of atmospheric carbon dioxide. *J. Climate,* **7,** 5–23.

Murphy, A. H., and E. S. Epstein, 1989: Skill scores and correlation coefficients in model verification. *Mon. Wea. Rev.,* **117,** 572–581.

North, G. R., T. L. Bell, R. F. Cahalan, and F. J. Moeng, 1982: Sampling errors in the estimation of empirical orthogonal functions. *Mon. Wea. Rev.,* **110,** 699–706.

Parker, D. E., P. D. Jones, C. K. Folland, and A. Bevan, 1994: Interdecadal changes of surface temperature since the late nineteenth century. *J. Geophys. Res.,* **99,** 14 373–14 399.

Rayner, N. A., D. E. Parker, E. B. Horton, C. K. Folland, L. V. Alexander, D. P. Rowell, E. C. Kent, and A. Kaplan, 2003: Global analyses of SST sea ice and night marine air temperature since the late nineteenth century. *J. Geophys. Res.,* in press.

Reynolds, R. W., and T. M. Smith, 1994: Improved global sea surface temperature analyses using optimum interpolation. *J. Climate,* **7,** 929–948.

——, N. A. Rayner, T. M. Smith, D. C. Stokes, and W. Wang, 2002: An improved in situ and satellite SST analysis. *J. Climate,* **15,** 1609–1625.

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

Smith, T. M., and R. W. Reynolds, 1998: A high-resolution global sea surface temperature climatology for the 1961–90 base period. *J. Climate,* **11,** 3320–3323.

——, and ——, 2002: Bias corrections for historic sea surface temperatures based on marine air temperatures. *J. Climate,* **15,** 73–87.

——, ——, R. E. Livezey, and D. C. Stokes, 1996: Reconstruction of historical sea surface temperatures using empirical orthogonal functions. *J. Climate,* **9,** 1403–1420.

——, R. E. Livezey, and S. S. Shen, 1998: An improved method for analyzing sparse and irregularly distributed SST data on a regular grid: The tropical Pacific Ocean. *J. Climate,* **11,** 1717–1729.

——, T. R. Karl, and R. W. Reynolds, 2002: How accurate are climate simulations? *Science,* **296,** 483–484.

Thiébaux, H. J., 1997: The power of the duality in spatial-temporal estimation. *J. Climate,* **10,** 567–573.

——, and M. A. Pedder, 1987: *Spatial Objective Analysis with Applications in Atmospheric Science.* Academic Press, 299 pp.

Van den Dool, H. M., S. Saha, and Å. Johansson, 2000: Empirical orthogonal teleconnections. *J. Climate,* **13,** 1421–1435.

Wolter, K., 1997: Trimming problems and remedies in COADS. *J. Climate,* **10,** 1980–1997.

Woodruff, S. D., H. F. Diaz, J. D. Elms, and S. J. Worley, 1998: COADS Release 2 data and metadata enhancements for improvements of marine surface flux fields. *Phys. Chem. Earth,* **23,** 517–527.

Yasunaka, S., and K. Hanawa, 2002: Regime shifts found in the Northern Hemisphere SST field. *J. Meteor. Soc. Japan,* **80,** 119–135.

Zhang, Y., J. M. Wallace, and D. S. Battisti, 1997: ENSO-like interdecadal variability: 1900–93. *J. Climate,* **10,** 1004–1020.