A Global Merged Land–Air–Sea Surface Temperature Reconstruction Based on Historical Observations (1880–1997)

THOMAS M. SMITH

NOAA/NESDIS/National Climatic Data Center, Asheville, North Carolina, and CICS/ESSIC, University of Maryland, College Park, College Park, Maryland

RICHARD W. REYNOLDS

NOAA/NESDIS/National Climatic Data Center, Asheville, North Carolina

(Manuscript received 19 May 2004, in final form 13 October 2004)

ABSTRACT

A merged land–air–sea surface temperature reconstruction analysis is developed for monthly anomalies. The reconstruction is global and spatially complete. Reconstructed anomalies damp toward zero in regions with insufficient sampling. Error estimates account for the damping associated with sparse sampling, and also for bias uncertainty in both the land and sea observations. Averages of the reconstruction are similar to simple averages of the unanalyzed data for most of the analysis period. For the nineteenth century, when sampling is most sparse and the error estimates are largest, the differences between the averaged reconstruction and the simple averages are largest. Sampling is always sparse poleward of 60° latitude, and historic reconstructions for the polar regions should be used with caution.

1. Introduction

Analysis of past climate variations is an important part of understanding climate change, and an important indicator of climate is the surface temperature. Long analyses of the surface temperature are used for monitoring the present climate and for comparison between the present and past climatic variations. Here an analysis of the merged surface temperature is produced. The analysis uses a sea surface temperature (SST) analysis over the ocean and a separate land surface air temperature (LST) analysis over the land. Surface marine air temperatures are not used because of biases in the daytime temperatures (Rayner et al. 2003). The two analyses are merged to form a monthly merged analysis, from 1880 to 1997.

In recent years a number of groups have developed surface temperature analyses. For example, SSTs have been analyzed by Bottomley et al. (1990), Smith et al. (1996), Kaplan et al. (1998), and Rayner et al. (2003). More recently, Smith and Reynolds (2003, 2004) applied improved reconstruction methods to the historical SST data, and produced an extended global reconstruction of SST anomalies. These methods remove almost all noise by fitting the data to a set of covariance modes, based on a densely sampled period. However, they damp the anomalies when and where data are too sparse for a reliable reconstruction. Uncertainty estimates reflect that damping, as well as uncertainty, is caused by possible biases in the observed data.

LSTs have been analyzed by Jones et al. (1990), Peterson and Vose (1997), Hansen et al. (1999), Jones et al. (2001), and Jones and Moberg (2003). Some details of updated LSTs and what they show about climate change are given by Parker et al. (2004). Averages of these LST analyses are often used to indicate climate change. In addition, merged LST and SST anomaly datasets have been produced (e.g., Parker et al. 1994; Jones et al. 2001; Jones and Moberg 2003). All of these studies help to better explain historical temperature variations.

In this study we use the improved analysis and error-estimation methods of Smith and Reynolds (2004) to analyze both SST and LST anomalies, and to then

Corresponding author address: Dr. Thomas M. Smith, National Climatic Data Center, ESSIC, 4115 Computer and Space Sciences Bldg., University of Maryland, College Park, College Park, MD 20742-2465.
E-mail: tom.smith@noaa.gov

merge them to form a global analysis. The merged analysis is spatially complete, although anomalies are damped in regions with sparse sampling. For example, at very high latitudes there is little sampling and anomalies are usually damped in the Arctic and Antarctic regions. However, when sufficient data are available they are analyzed in those regions, and the analysis is formally global (90°S–90°N). In addition, the uncertainty estimates for the reconstruction are produced.

This merged reconstruction contains a number of improvements over many earlier studies, including the following. 1) The merged analysis is globally complete. Spatial covariance modes are used to interpolate anomalies in undersampled regions. Covariance modes are computed using the relatively dense modern sampling. However, we limit our covariance functions to be either entirely over land or entirely over water, and we also limit their spatial size, to minimize overinterpolation of anomalies. 2) The analysis incorporates the latest updates in the International Comprehensive Ocean–Atmosphere Data Set (ICOADS; Woodruff et al. 1998) and the Global Historical Climate Network (GHCN) data. 3) The analysis variance is found to have less dependence on sampling compared to some earlier analyses. This is because the covariance modes incorporated in the analysis only need to be partially sampled at any time in order to contribute to the analysis. 4) Uncertainty estimates indicate when and where the analysis is most reliable. The uncertainty estimates can also be computed for spatial and/or temporal averages of the analysis, to indicate how the averaging affects uncertainty.

This merged analysis provides another tool to help define long-term temperature variations over the twentieth century, and to indicate when and where the analyzed temperature variations are significant. This analysis is also being operationally extended at the National Climatic Data Center, so that future variations may be compared to historical variations. For some applications there may be advantages to having a spatially complete analysis such as this. For example, modeling studies using surface conditions as a boundary may be easier to conduct. In addition, the uncertainty estimates help to better define the possible range of climate change over the twentieth century, as discussed below.

2. Data

The primary SST data used for this study are the ICOADS SST observations release 2, with updates through 1997 (Slutz et al. 1985: Woodruff et al. 1998). The individual observations are screened using a quality control test, and those that pass are averaged to monthly and 2° spatial superobservations. Screening is done by comparing individual anomalies to a spatial–temporal local analysis of anomalies. Values close to the local analysis value are retained [see Smith and Reynolds (2003, 2004) for details about the quality-control methods]. Superobservations are defined as the average of all input data over a given grid box for a given month. For SSTs prior to 1942, the bias corrections of Smith and Reynolds (2002) are applied to correct for changes in measurement techniques.

Before 1942 most SST measurements were from ships that used buckets to bring samples of seawater onto the deck, where temperature measurements were made. Afterward it became more common to measure the temperature of the engine-intake seawater. Adjustments to the pre-1942 SSTs were developed by Folland and Parker (1995), and by Smith and Reynolds (2002). However, further analysis of the ICOADS release 2 SST by C. K. Folland (2003, personal communication) indicates that for 1939–41, the release 2 data are biased warm relative to the data used by Folland and Parker (1995). The release 2 data contain additional data, from different sources and with different bias characteristics compared to the Folland and Parker (1995) SST data. Those differences are minimized if the bias correction is reduced linearly to zero beginning in January 1939 (no adjustment to the bias correction) and ending in December 1941 (zero bias correction). Because we use the release 2 data, we apply this linear reduction to the Smith and Reynolds (2002) bias adjustment from 1939 to 1941.

Anomalies of the SST superobservations are formed by subtracting off the monthly Smith and Reynolds (1998) 1961–90 SST climatology. Because of our desire to eliminate noise from the reconstruction, those anomalies were subjected to a second quality-control screening to remove extreme anomalies (Smith and Reynolds 2004). In addition to ICOADS, we use the combined satellite and in situ SST analysis of Reynolds et al. (2002) to compute analysis statistics. The SST and their analysis are more fully described by Smith and Reynolds (2003, 2004).

The LST data used for reconstruction are the GHCN temperatures (Peterson and Vose 1997). To develop the gridded GHCN, the anomalies of individual GHCN stations are formed with respect to 1961–90. Those anomalies are then averaged to monthly and 5° spatial superobservations. The GHCN data are homogeneity adjusted to minimize artificial variations, such as from moving the location of a station (Peterson et al. 1998). These LST data are sufficient for defining large-scale temperature variations, although many details can be lost because of the relatively coarse grid. The LST

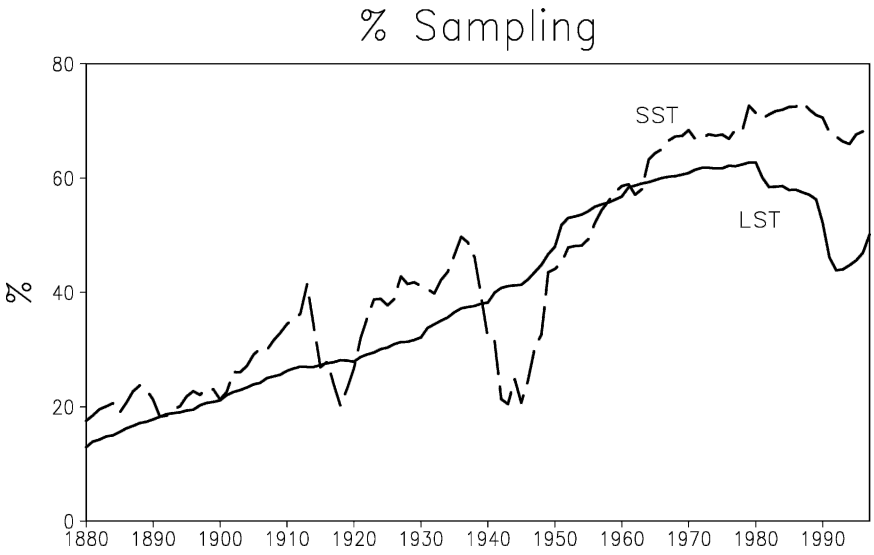

FIG. 1. Percent of the global reconstruction region sampled for SST (dashed) and LST (solid). Annual averages are given.

analysis was performed on this 5° grid, and the 2° SST analysis was averaged to this same 5° grid. In addition, the unanalyzed 2° superobservation anomalies were averaged to the 5° GHCN grid for comparisons, as discussed below. The unanalyzed data are the monthly superobservations that are used as input for the reconstructions.

Sampling of SST tends to increase over the analysis period, although there are dips in the sampling associated with the two world wars (Fig. 1). The SST sampling is consistently best after 1950. For LST, the sampling increase with time is monotonic up to about 1980, when it begins to decrease slightly. That recent decrease in LST sampling results from a time lag for the inclusion of some data into the GHCN and station dropouts. For the same reasons a similar but smaller decrease is evident in the ICOADS SST sampling. The LST percent of in situ sampling is often lower than for SST and even in recent years is not much over 60% because of persistent undersampled land regions, notably Antarctica, central Africa, and central South America.

3. Reconstruction

The historical SST reconstruction method used in this study is described in detail by Smith and Reynolds (2004). Reconstruction methods are also described in some detail in appendix A, and they are briefly reviewed here.

The reconstruction is separated into separate low- and high-frequency components, which are added for the total reconstruction. First the low frequency is reconstructed using spatial and temporal filtering of the available data. The time filter of 15 yr defines the low frequency as approximately decadal scale or longer. That low-frequency component is subtracted from the data before reconstruction of the high-frequency component using spatial covariance modes, which includes interannual and shorter-period variations. The covariance modes are spatially complete. However, dense enough data for computing spatially complete modes is limited to the recent period. Interannual or shorter-period variations, such as ENSO, can be represented using data from the recent densely sampled period. Lower-frequency variations may not be well represented by data from the most recent period. This led us to develop a method that separately reconstructs the low- and high-frequency anomaly.

Our SST anomaly base period data are the 1982–2002 Reynolds et al. (2002) in situ and satellite SST analysis. There are 130 spatial modes used for the SST high-frequency analysis. Sea ice information from Rayner et al. (2003) is merged with the SST reconstruction to adjust the high-latitude temperatures. The low- and high-frequency variations are separately reconstructed. Localized spatial covariance modes are computed from the recent base-period anomalies, and are used for the high-frequency reconstruction. Simple methods, described below, are used to reconstruct the low-frequency variations.

Because SST observations are often sparse, we include SST anomalies for three consecutive months, centered on the analysis month, in our high-frequency

TABLE 1. The numbers of 1-month lag autocorrelations in the given range for SST and for LST modes.

	(0.00, 0.25]	(0.25, 0.50]	(0.50, 0.75]	(0.75, 1.00]	Total
SST	0	2	122	6	130
LST	52	7	1	0	60

analysis. When there is a superobservation for a 2° square for the analysis month, we use that anomaly. When there is no superobservation available, we use the average of the anomalies from the previous and following month. This is justified because SST anomalies on the scales we are resolving typically persist for more than a month, as discussed below.

The historical LST reconstruction uses essentially the same methods, outlined in appendix A. The base-period data used to compute the high-frequency modes over land are the GHCN data for 1982–91. That period overlaps the SST base period and has good overall sampling (Fig. 1). After 1991 there are slightly fewer GHCN data, as discussed above. To make the modes continuous over all land regions, we use optimum interpolation (OI; e.g., Reynolds and Smith 1994) to fill in land regions without data. The LST spatial covariance modes are based on these OI-filled fields of GHCN data. In the future, we hope to have an improved modern-period LST analysis that blends in situ and satellite data. We considered using the National Centers for Environmental Prediction–National Center for Atmospheric Prediction (NCEP–NCAR) reanalysis data to define the LST modes. Those data are blended and assimilated into an atmospheric model, and anomalies based on those data are generally consistent with observed anomalies. However, because the filled recent-period LST anomalies adequately span the subdecadal variance, and also because we plan to update the LST analysis using blended in situ and satellite LST data, we decided against introducing an additional dataset at this time.

Because LST observations are fixed and not moving (in contrast to most SST observation platforms), there is little need for using 3 months of pooled data, as we did with the SST reconstruction. In addition, the LST anomalies are less persistent than the SST anomalies, discussed below (see Table 1). Therefore, for the LST high-frequency reconstruction we only use the LST anomalies for the analysis month.

The SST base period is both longer and better sampled than the LST base period because of the availability of more than 20 yr of satellite SST retrievals. In the future we hope to improve the LST database by incorporating more observations, including satellite-based LST estimates. However, we are able to compute 60 LST modes using the available data, and it is unlikely that additional base-period data will change our major results.

Anomaly persistence is stronger in SST anomalies than in LST anomalies, as shown by the autocorrelation of the covariance modes. The 1-month autocorrelation for each mode is computed from the time series of weights for each mode, reconstructed using the data since 1950 when all modes are defined most of the time. Table 1 shows that the SST modes typically have an autocorrelation between 0.5 and 0.75 (corresponding to *e*-folding times of 1.5–3.5 months), while for LST the autocorrelation is typically below 0.25 (corresponding to *e*-folding times less than a month). Thus, the LST analysis can make little use of persistence when a station disappears.

The global ocean area is roughly twice as large as the global land area, and there are roughly twice as many possible SST modes compared to LST modes. Land variations are often more complicated than sea surface variations because of different elevations and land types. However, this reconstruction has a coarse spatial resolution, and many variations associated with changes in land types have smaller scales. Thus, an LST reconstruction with finer resolution may require more spatial modes to resolve variations associated with changes in elevation or land type. For each month, modes that have 25% or more of their variance sampled (as defined in Smith et al. 1998 and Smith and Reynolds 2003) are selected for the high-frequency reconstruction. Using modes with less of their variance sampled could introduce noise into the analysis. When a mode is adequately sampled, it will filter out random noise because random noise will not fit the spatial pattern of any physically based mode. However, if a mode is undersampled then noise could influence the weight computed for that mode. By using testing we found that 25% sampling is adequate for filtering out nearly all of the data noise.

For the 1880–1997 period, the number of LST modes selected increases gradually until about 1950 (Fig. 2). After 1950 near the maximum number of LST modes are selected for most of the remaining period, with a slight decrease in the 1990s. For SST modes, nearly all modes are selected for most of twentieth century, except during the two world wars when sampling was interrupted.

The separate historical SST and LST anomaly reconstructions are merged to form the global-anomaly reconstruction. Over most regions, the two are merged by weighting the SST and LST reconstruction by the percentage of ocean and land area in the 5° region. How-

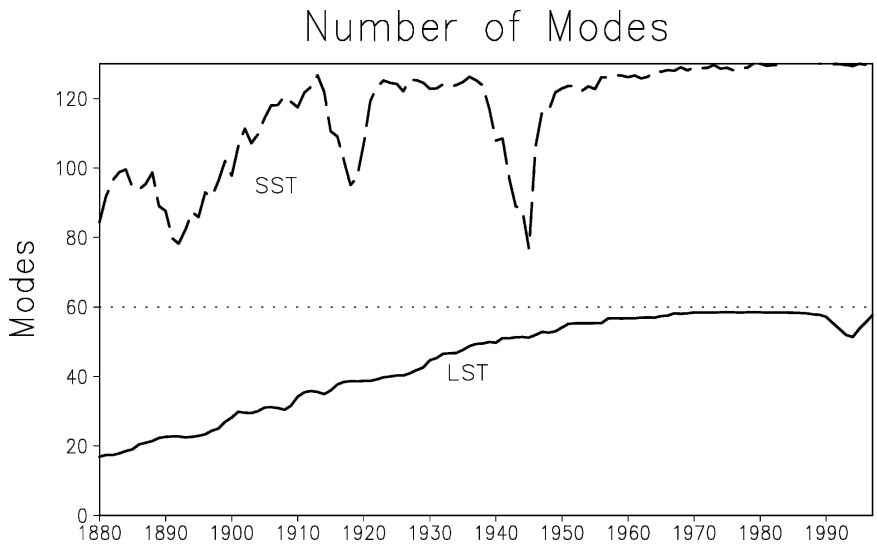

FIG. 2. Number of reconstruction modes selected for SST (dashed, maximum = 130) and LST (solid, maximum = 60). The annual-average number is given.

ever, since we are most interested in surface air temperature anomalies, it is not appropriate to use SST anomalies in regions covered by ice. Therefore, the historical fractional ice cover developed by Rayner et al. (2003) is used to further adjust the SST anomalies before merging with LST anomalies. If there is no ice then no further adjustment is done (Smith and Reynolds 2004). As the fraction of ice cover increases from 0 to 1, the merged SST anomaly is linearly damped toward 0. This treats ice-covered ocean regions the same as land regions with no sampling. In practice, there is usually little or no sampling in regions with ice, so those regions have reconstructed anomalies near zero to begin with. Thus, this adjustment has little effect on the merged temperature anomaly.

The method for estimating errors in the merged analysis is described in Smith and Reynolds (2004), and summarized in appendix B. Errors account for historical gaps in sampling and for bias uncertainties in the data. For the first half of the analysis period, the largest component of the reconstruction error is from the sampling error of the low-frequency component. This is because there is not enough sampling in that period to resolve a trend over much of the globe. Comparisons between averages of our merged analysis and simple averages of the data, discussed below, show that our error estimates generally bracket the range of comparison estimates.

As an example of the input anomalies and the reconstruction, we show the merged input data and the merged reconstruction averaged for 1900–09 (Fig. 3). For the unanalyzed anomalies, we require that at least 24 months be sampled over the decade. With less sampling the region is left blank in the unanalyzed average. The average reconstruction is broadly consistent with the average of the input data, but the pattern is smoother and the anomalies tend to be damped in the reconstruction. Over regions with little or no input data the reconstruction anomaly is near zero. Note that the analysis uses interpolation to smoothly fill those data-sparse regions. Modest size data-void regions surrounded by relatively large anomalies, such as the North Pacific midlatitudes and over Greenland, are filled with relatively strong anomalies, by interpolation using the spatial-covariance modes.

The unanalyzed anomalies are sometimes stronger than the analyzed anomalies. For example, in the 1900–09 average of unanalyzed anomalies, there are strong local anomalies off the east coast of Asia and North America. Averaging the analyzed anomalies damps the magnitude of those anomalies. One reason for stronger unanalyzed anomalies is that when data are sparse, the analyzed anomaly will damp toward zero. Thus, isolated anomalies inconsistent with neighboring anomalies are damped in the analysis. Of course, one purpose of the analysis is to reduce noise on monthly scales. Thus, there is a balance between reducing monthly noise and producing a smoothed analysis.

Averaging the damped anomalies weakens the decadal average. However, even with complete sampling the analysis filters the data. Since part of the analysis is performed by fitting data to a set of large-scale modes (generally covering spatial regions of 20° or larger), variations on smaller scales tend to be filtered. In ad-

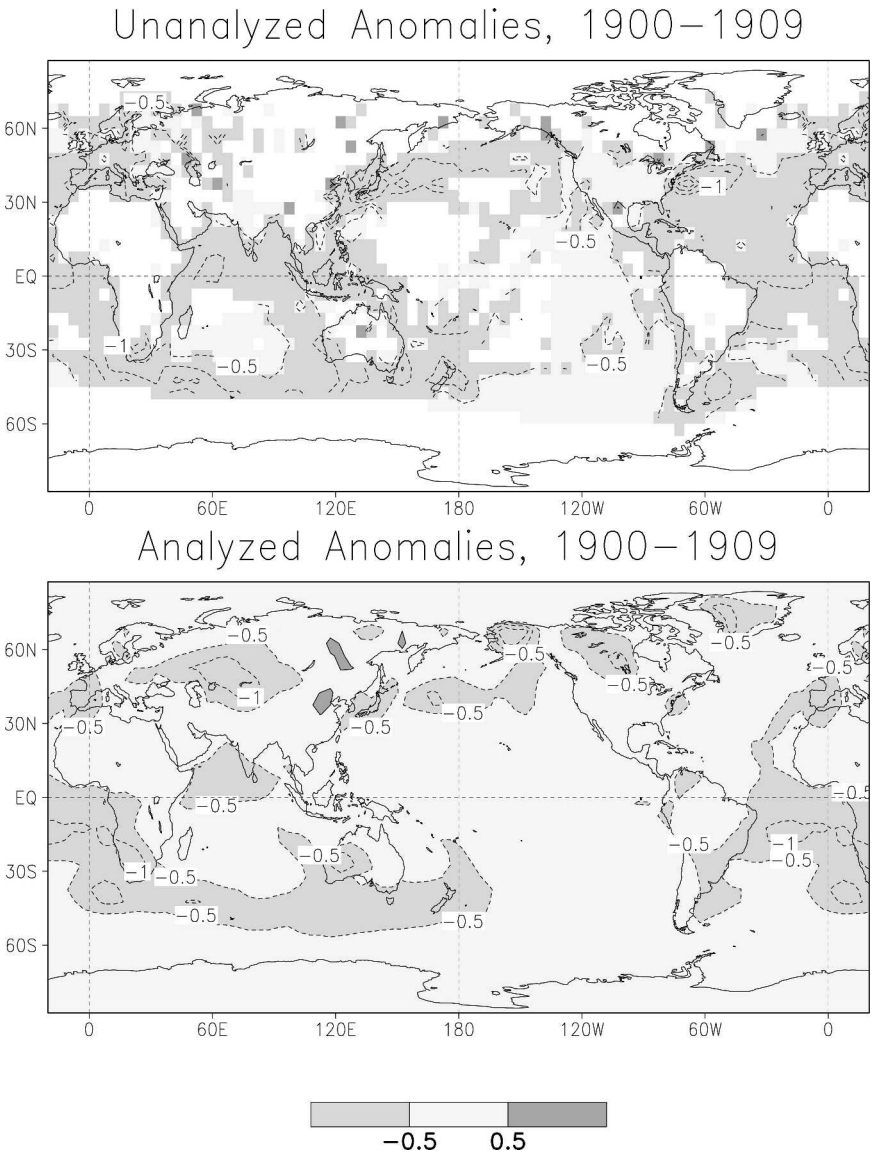

FIG. 3. Average surface temperature anomaly for 1900–09 (top) for the input data (with shading to indicate sampling) and (bottom) for the merged reconstruction.

dition, variations not represented by the modes will also be filtered.

4. Results

To illustrate historical temperature variations in the reconstructions along with their uncertainty, time series of averages with 95% confidence intervals are shown. Annual and 60°S–60°N averages are computed for the SST anomalies, the LST anomalies, and the merged anomalies. In addition, simple averages are computed of the unanalyzed data used for the reconstructions, and of comparable merged data produced by the Met Office Hadley Centre and at the Climate Research Unit (CRU) of the University of East Anglia (Jones and Moberg 2003). The Hadley Centre–CRU temperature (HadCRUT, version 2) simple averages (as updated by Jones and Moberg 2003) are displayed for comparison with the reconstruction averages.

The simple averages are area-weighted averages of the 5° superobservations. Since the superobservations are only defined where there is sampling, which is most dense in the Northern Hemisphere, area-weighted averages for the Northern and Southern Hemispheres are first computed separately, and then they are then averaged. Each hemispheric average is weighted by the to-

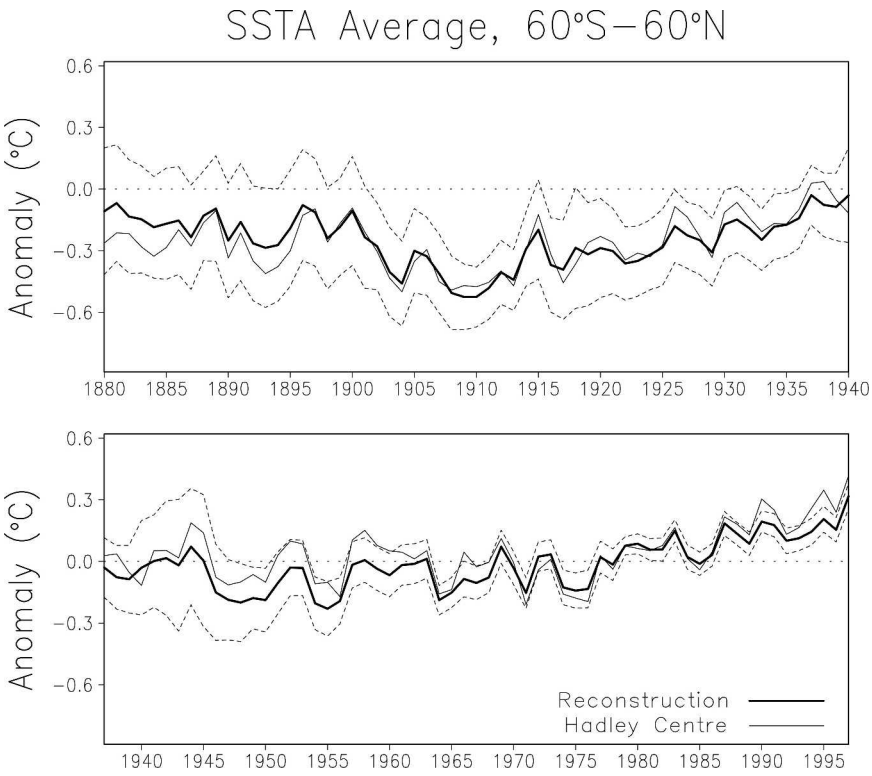

FIG. 4. Reconstructed SST anomaly averaged annually and between 60°S and 60°N (solid), with its 95% confidence intervals (dashed). Also shown is the simple average of the comparable SST anomalies produced by the Hadley Centre.

tal area in each hemisphere (of land or sea or merged). Averaging the hemispheres separately first and then area weighting the results keeps the better-sampled Northern Hemisphere from artificially dominating the averages.

The average reconstruction of SST anomalies from 60°S–60°N (Fig. 4) indicates warming through most of the twentieth century. The SST warming occurs in two parts, the first before 1940 and the second after 1970, with a roughly stationary period between 1940 and 1970. The uncertainty estimates indicate that the nineteenth-century anomalies should be used with caution. Part of the nineteenth-century uncertainty is due to bias uncertainty, which could be reduced using future bias corrections that incorporate a better understanding of the historical bias. However, much of the nineteenth-century uncertainty is due to the effect of sparse sampling on the low-frequency error estimate. In any case, the twentieth-century warming is significant. Simple averages of the ICOADS SST anomalies are consistent with the reconstruction, although the reconstruction anomaly is sometimes slightly weaker due to damping. The Hadley Centre SST anomaly simple average is also generally consistent with the reconstruction. Hadley Centre SSTs are slightly warmer than ICOADS in the 1990s. Most 1990s differences are in western boundary currents off the east coasts of North and South America, between about 30° and 60° latitude, with the largest differences off South America. Differences in data quality control and screening of suspect observations in those regions could account for much of the difference. For most of the period the simple averages are within the reconstruction 95% confidence interval. The 60°S–60°N average SST anomaly and error estimates are similar to those computed by Smith and Reynolds (2003). For the recent period, the uncertainty is similar to the ship–buoy SST bias computed by Reynolds et al. (2002).

The average reconstruction of LST anomalies (Fig. 5) also indicates warming through the twentieth century. There is gradual LST warming until about 1940, when the trend flattens out, and a second strong warming trend begins about 1970. These LST trends are consistent with the SST trends, and this similarity was also noted by Folland et al. (2001b). However, there is more uncertainty in the LST average, especially before 1940 when much of the land areas are undersampled. Since we analyzed LST and SST anomalies separately, the

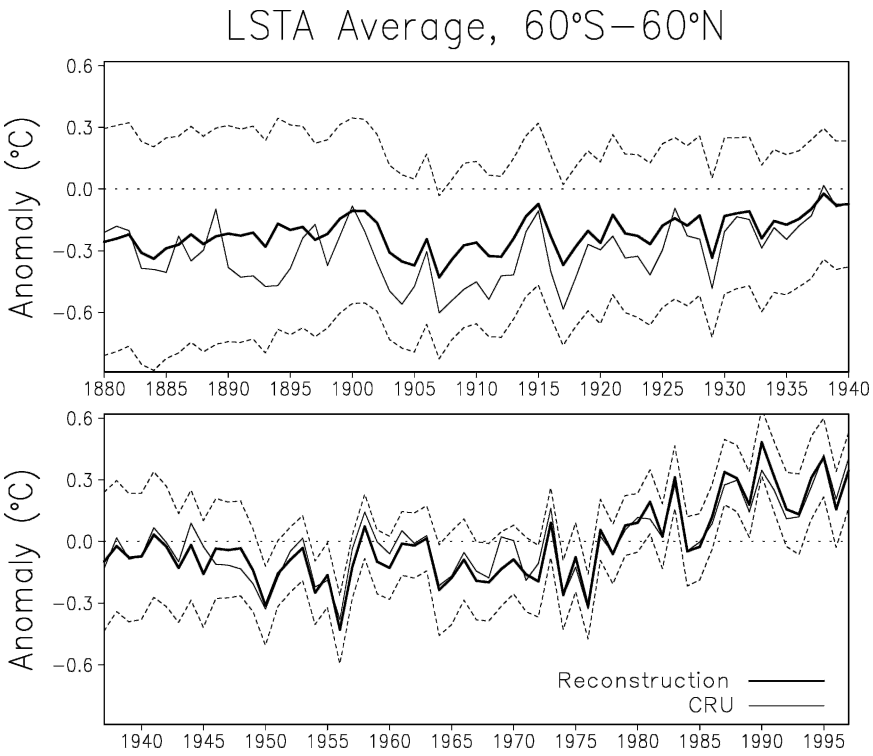

FIG. 5. Reconstructed LST anomaly averaged annually and between 60°S and 60°N (solid), with its 95% confidence intervals (dashed). Also shown is the simple average of the comparable data produced by the CRU (Jones and Moberg 2003).

similarity between the two is derived from the data, and is not an artifact of the analysis method. Variations in the LST simple averages in the nineteenth and early twentieth century indicate that these large uncertainty estimates are justified. The Jones–CRU LST anomalies are cooler than the reconstruction early in the record, when sampling is sparse and our reconstruction is more damped. After 1930, those simple averages are more consistent with the reconstruction.

The merged reconstructed anomalies (Fig. 6) indicate a blend of the SST and LST anomalies, weighted heavily toward the SST. Note that here the 60°S–60°N average is displayed. This is similar to the full global mean, both because this includes most of the earth and because there are few data outside this region. Including the polar regions would, however, increase the uncertainty estimates. As with the SST, the simple averages are generally consistent with the reconstruction average. The twentieth-century warming is about 0.6°C. Because of the uncertainty estimates, the warming can only be confidently established between 0.3° and 0.9°C. Compared to the global average of Folland et al. (2001, 2001b), our average is similar over most of the analysis period. However, our nineteenth-century negative anomalies are slightly weaker than theirs, because their analysis is not damped when sampling is sparse. Both this study and Folland et al. (2001, 2001b) indicate warming of about 0.6°C over the twentieth century. However, our uncertainty estimate for the warming is ±0.3°C, slightly large than their estimate of ±0.2°C.

Our uncertainty estimates are larger than the Folland et al. (2001) estimates because of our large low-frequency error estimate. In their estimate, large-scale spatial modes are used to estimate both the high- and low-frequency spatial covariance, making their low-frequency sampling error much smaller than for our more conservative method of estimating the low-frequency analysis. The global error estimates of Jones et al. (1997) are also larger than the Folland et al. (2001) estimates, but smaller than the estimates of this study.

Some idea of where the reconstruction is most and least reliable for different periods can be obtained by correlations with comparable analyses. Here we correlate reconstruction anomalies with those from the Hadley Centre Sea Ice and SST (HadISST) analysis over the oceans (Rayner et al. 2003) and Jones and Moberg (2003) over land. The HadISST analysis is spatially complete, as is the reconstruction, but the Jones land data have gaps where there are no stations. Thus, there

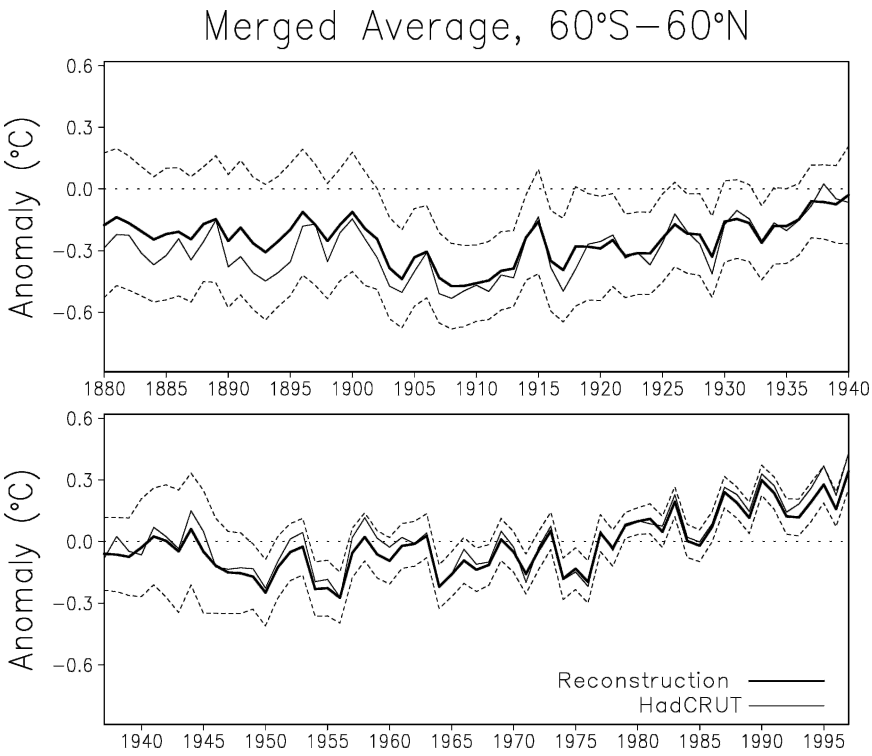

FIG. 6. Merged SST and LST reconstructed anomaly averaged annually and between 60°S and 60°N (solid), with its 95% confidence intervals (dashed). Also shown is the simple average of the comparable HadCRUT (Jones and Moberg 2003) data.

are some gaps in the land correlations. Correlations are computed using anomalies from all months for approximately 30-yr periods: 1880–1909, 1910–39, 1940–69, and 1970–97. Although not a direct estimate of error in either analysis, these correlations indicate where and when sampling is sufficient for the different analyses to converge. Such convergence may increase user confidence in the reliability of the reconstruction.

As expected, correlations generally increase with sampling from the earliest to the most recent period (Fig. 7). In all periods the correlations are high over the eastern tropical Pacific and tropical Atlantic, and over North America and Europe. In the earliest period there is low correlation over the western tropical Pacific, where variance is always much lower than in the eastern tropical Pacific. For most of the reconstruction period the correlation is also low south of 40°S, where dense SST sampling has only become available in recent decades from drifting buoys and satellites (Reynolds et al. 2002). Correlations also tend to be low or absent in all periods for most of Africa, northern South America, and Greenland, which have persistently sparse sampling. For the 1910–39 period, most of the Northern Hemisphere land area has a high correlation, and oceanic correlations are slightly larger compared to 1880–1909. Larger increases over oceans are evident when going from 1910–39 to 1940–69, due to increased SST sampling after about 1950. The change in correlations from one period to another is broadly consistent with changes in error for the averages (see Figs. 4–6). Both are heavily influenced by the available sampling.

To better identify a few climate signals, the correlation is computed between the reconstruction and a one-predictor linear model based on the reconstruction. We construct a one-predictor linear regression model for surface temperature anomalies. The given climate-signal index is used as the predictor in this statistical model. Correlations between the predicted temperature anomalies and the reconstruction are similar to linear correlations between the climate index and the reconstruction.

For these comparisons only boreal winter (December–March) averaged anomalies are used. Comparisons are over the 1900–97 period (the nominal year is the year of January). Climate indices used are the Southern Oscillation index (SOI) provided by the NCEP/CPC (Chelliah 1990); the North Atlantic Oscillation (NAO) as defined by Hurrell (1995); and the Pacific decadal oscillation as defined by Mantua et al. (1997). For these indices, climate signals are strongest in boreal winter, so

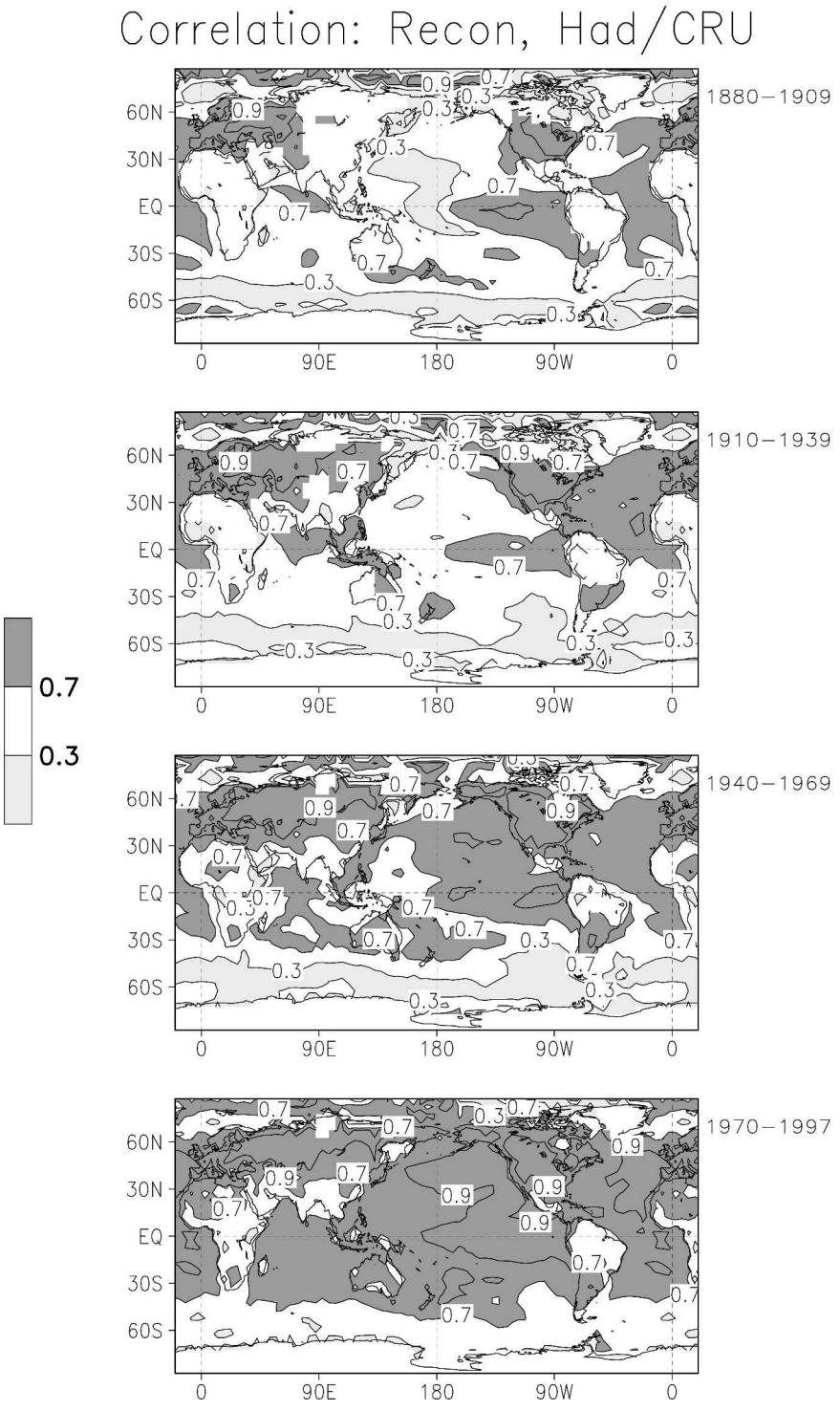

FIG. 7. Correlation between the reconstruction anomaly and the merged HadISST–CRU anomaly, for the periods 1880–1909, 1910–39, 1940–69, and 1970–97, as indicated. Values of 0.3, 0.7, and 0.9 are contoured. Light shading indicates correlations less than 0.3, and dark shading indicates correlations greater than 0.7.

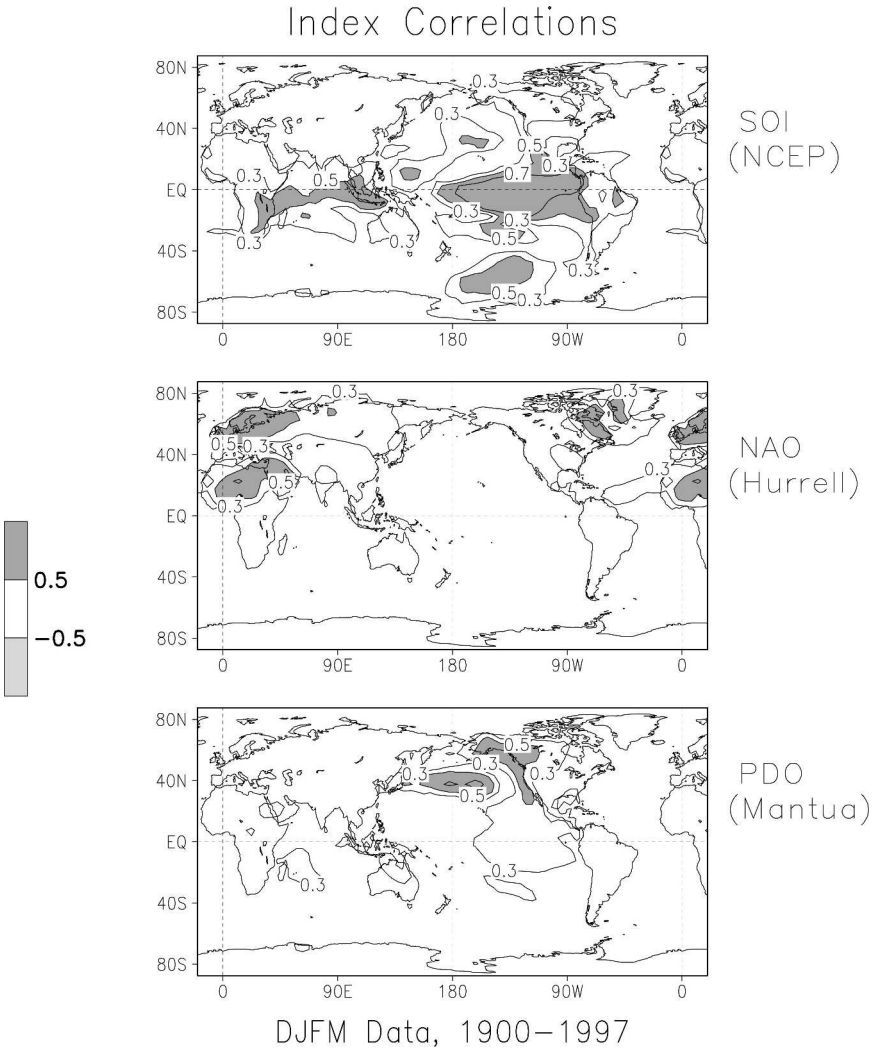

FIG. 8. Correlation of the full merged anomalies with the linear regression model using the given climate index as the predictor, and using data averaged for the Dec–Mar season, from 1900 to 1997.

these seasonal averages should emphasize the correlations.

Among the three (Fig. 8), correlations with the SOI are strongest. The NAO is related to Eurasian winter temperatures. Over the North Pacific the PDO correlations are somewhat similar to the SOI correlations. However, the PDO has a strong low-frequency component.

5. Summary and conclusions

A merged land and sea surface temperature reconstruction is developed. The reconstruction produces globally complete monthly surface temperature anomalies. These fields can be averaged spatially or temporally as desired. In addition, methods of estimating errors in the surface temperature reconstruction are produced, and examples are given.

For the near-global average, the average of the reconstruction is similar to that produced by other studies for most of the analysis period. In the nineteenth-century averages of the analyzed anomalies are weaker than in some other studies because of damping in the reconstruction when data are sparse.

Several possible improvements may be made to the merged reconstruction. Improvements in the SST bias corrections applied to data before 1942 could decrease uncertainty in that period. More data for anytime before 1950 could decrease sampling errors for both the land and sea analysis. In addition, an improved satellite

and in situ analysis of land surface air temperatures for the recent period could improve the statistics used for the land analysis.

Error estimates for the merged analysis are larger than those estimated by others. For example, Folland et al. (2001) estimate that the sampling error for the global average is less than in this study. The larger errors in this study are due to the simple methods used for the low-frequency analysis. Here the low-frequency analysis is only resolved if there are sufficient local data for a large-scale (10° or 15° spatial) average and enough data for low-frequency (15 yr) filtering. Otherwise the low-frequency analysis damps to zero anomaly. In Folland et al. (2001), both the low and high frequencies are analyzed together using a set of spatial modes with much larger scales (spatial scales generally >15°), and less sampling is needed to resolve their low-frequency variation. It is not clear which approach is better. It may be possible to span most low-frequency variations using a set of modes based on 50 yr. However, there is no guarantee that the more recent low-frequency variations are characteristic of the entire reconstruction period. In addition, differences between simple averages of the data and averages of the analysis are sometimes about as large as our 95% uncertainty estimates, indicating that the magnitude of the estimated errors may be justified. (The merged reconstructed data are available online at ftp://ftp.ncdc.noaa.gov/pub/data/er-ghch-sst/.)

Acknowledgments. We thank J. Lawrimore for providing the GHCN data, P. Jones for supplying the HadCRUT data, and S. Woodruff for supplying the ICOADS data. We also thank Y. Chen and J. Francis for helpful discussions of Arctic temperature variations. Climate signals were obtained from the JISAO Web site maintained by the University of Washington (http://www.jisao.washington.edu/data_sets/). D. Parker and C. Folland assisted with discussions about the SST bias correction. This study was also improved by discussions with T. Karl, S. LeDuc, and T. Peterson and reviews by V. Kousky, C. Ropelewski, R. Vose, and two anonymous reviewers. Funding for this study was provided by NOAA office of global programs.

APPENDIX A

Reconstruction Methods

The reconstruction is done in two separate parts: a low-frequency reconstruction and a high-frequency reconstruction. A low-frequency anomaly reconstruction is first produced using simple methods. Those simple methods do not depend on stationary statistics. This is done because the base period that statistics are computed from is relatively short compared to the analysis period, and thus may not be long enough to span all low-frequency variations. Simple methods may better preserve the historical low-frequency variations. Subtracting the low-frequency analysis from the data yields the high-frequency residuals. Those residuals are analyzed separately by fitting them to a set of spatial modes.

The high-frequency modes represent seasonal to interannual variations. The statistical structures of the interannual variations are assumed to be fully represented by the base-period data that the modes are computed from. Here our base periods are 10–20 yr long. The total reconstruction is the sum of the low- and high-frequency analyses. Details of the reconstruction methods are given in Smith and Reynolds (2003, 2004). Here a description is given of how the methods are applied in this study.

The low-frequency analysis is performed by averaging and filtering the temperature anomalies. First, the anomaly monthly superobservations are spatially averaged. For SST anomalies the 2° monthly superobservations are averaged to a 10° grid, provided that there are at least three 2° superobservations defined for the month. For LST anomalies the 5° monthly superobservations are averaged to a 15° grid. For each year, monthly values are used to produce an annual average, provided that there are at least 4 monthly averages defined. These minimum data requirements are to ensure that there are enough data to damp random noise and resolve the average. The annual averages are then smoothed and expanded slightly using spatial binomial filters. The resulting smoothed annual values are filtered using a 15-yr median filter. Where there are not enough annual averages to compute a median, an anomaly of zero is assigned. Spatial and temporal binomial filters are then applied to this to further smooth the filled low-frequency estimate.

These procedures produce a low-frequency anomaly analysis that retains large-scale variations supported by the network of observations, without imposing any stationary structures on the variations. However, the low-frequency analysis is limited by the available sampling, which is sparse throughout the early part of the historical record. Because an anomaly of zero is assigned where there are too few data, these procedures tend to damp the low-frequency anomaly early in the record.

We do not interpolate the low-frequency analysis to fill all unsampled regions because we cannot be certain that the correlation scales are always large enough to allow this. We chose not to impose a large-scale correlation structure on the low-frequency variations. However, the low-frequency correlation structure may have

very large scales, as indicated by studies using more recent data to estimate them (e.g., Folland et al. 2001).

The high-frequency anomaly is defined as the difference between the full anomaly and the low-frequency anomaly: $R(x, t) = \text{Anom}(x, t) - \text{LF}(x, t)$. Those residuals are fit to a set of spatial-covariance modes to find a weight for each mode. For the SST analysis the base-period data used to compute the modes are the in situ and satellite OI analysis of Reynolds et al. (2002) from 1982 to 2002. For LST the base-period data are the GHCN data for 1982–91. Because the in situ GHCN sampling becomes less dense in the 1990s, due to station dropouts and delays in obtaining data, we use this shorter period to keep the modes more spatially complete. The weighted sum of those modes defines the high-frequency analysis:

$$\text{HF}(x, t) = \sum_{m=1}^{N} w_m(t) E_m(x), \tag{A1}$$

where $E_m(x)$ is the spatial mode, defined for $m = 1, \ldots, N$ modes, and $w_m(t)$ is the weight for the mode, computed independently from the available data at each time. To compute the set of weights for the modes, the data are fit to the set of modes, to minimize the mean-squared difference between the fit and the data at locations where there are observations. This is done by solving the system of equations:

$$\sum_{m=1}^{N} w_m(t) \sum_{x=1}^{K} E_m(x) E_n(x) \delta(x) a(x) = \sum_{x=1}^{K} R(x, t) E_n(x) \delta(x) a(x), \; n = 1, \ldots, N. \tag{A2}$$

Here $\delta(x) = 1$ if there is an observation at spatial position x, and 0 otherwise, and $a(x)$ is the area weight for position x. There are K spatial positions over the reconstruction area.

Before computing the weights, sampling for each mode is tested. If less than a critical percentage of the variance associated with the mode is sampled, then that mode is not used for the month's reconstruction (see Smith et al. 1998). Using a mode not adequately sampled can cause instabilities because of data noise. We find that rejecting modes with less than 25% of their variance sampled produces a consistently stable analysis for both SST and LST. For the set of adequately sampled modes, a weight for each mode is found by solving (A2), and the high-frequency analysis is then computed using (A1).

Within periods when the low-frequency analysis is greatly damped because of sparse data, the available residuals may contain some low-frequency variance. To the extent that the low-frequency variance can be projected onto the available modes, that low-frequency variance in the residuals will be analyzed using the modes.

Sampling may change over time, so a mode may be adequately sampled one month and not the next. However, there is a certain amount of persistence associated with each of these modes. Some modes have greater persistence and some less, as indicated by their 1-month lag autocorrelation (C_1). We compute the autocorrelation for each mode using the weights in the well-sampled modern period. Those autocorrelations are used to define the weights for modes in the months when they are not adequately sampled.

For each mode, the weights for undersampled months are computed by damping the weights from the nearest adequately sampled months. The damping factor is C_1^k, where k is the number of months from the undersampled month to the nearest adequately sampled month. Table 1 indicates the magnitude of the correlations for both SST and LST. Typical SST correlations are about 0.7, while LST correlations are typically smaller, about 0.2. By looking for the nearest adequately sampled month in both the forward and backward directions, temporal covariance in both directions is included. If the autocorrelation is large then that mode's weight can persist from months with good sampling into months with poor sampling. Modes that are not adequately sampled for a long time will have weights that damp to zero. The directly computed weights from adequately sampled months are combined with the weights estimated from autocorrelation. Since the LST modes have low autocorrelation, only the SST weights were extended using their autocorrelation. This set of weights is used to compute the high-frequency analysis using (A1).

For both SST and LST high-frequency analyses, the modes used are defined using empirical orthogonal teleconnections (EOTs; Van den Dool et al. 2000). To compute EOTs first a base point is found. The base point is the point with the greatest spatial covariance with all other points on the field. The covariance pattern for that mode is computed and subtracted from the data. The process is then repeated to find the next mode. The EOTs are similar to rotated empirical orthogonal functions (EOFs). However, EOTs are easier to modify and control. In particular, we are concerned that there may be very large-scale teleconnections in the base period that may not be common in all details to all historical periods. Using larger-scale EOFs would give a better reconstruction for the base period, but in more remote decades there is a danger of over specification from sparse data. Therefore, we localize the

EOTs by damping the covariance pattern at distances of more than 5000 km from the base point. At distances of between 5000 and 8000 km, the pattern is damped linearly from full strength to zero. Teleconnections are not allowed at distances greater than 8000 km.

APPENDIX B

Reconstruction Error Estimates

Analysis errors can be separated into three independent types of error (e.g., Kagan 1979): random error, sampling error, and bias error. The total analysis error variance may be written as their sum

$$\varepsilon^2 = \varepsilon_R^2 + \varepsilon_S^2 + \varepsilon_B^2. \quad \text{(B1)}$$

The subscripts R, S, and B indicate random, sampling, and bias error variances, respectively. Random error variance in an analysis, ε_R^2, is from random errors in the input data. In this analysis, those errors are almost entirely filtered out. Tests have shown that for this analysis method, the signal-to-noise variance ratio for SST is about 30 (see Smith and Reynolds 2003, 2004; the ratios were computed over approximately 30-yr periods and averaged over 60°S–60°N). For monthly LST analysis the ratio should be even larger. This is because the continental LST signal tends to be larger than the marine SST signal, and the monthly average LST anomalies are constructed from averaging more individual observation than is typical for SST. The relatively small random error variance allows simplification of the error estimate, as discussed below. Sampling error variance, ε_S^2, reflects the density and distribution of observations. Bias error variance, ε_B^2, is due to systematic biases in the data or from the analysis method.

If we let T be the true anomaly and we let T_a be the reconstruction anomaly, then we may express the error variance as

$$\begin{aligned}\varepsilon^2 &= \langle (T - T_a)^2 \rangle \\ &= \sigma_T^2 + \sigma_{T_a}^2 - 2\sigma_T \sigma_{T_a} r_{T,T_a} + (\langle T \rangle - \langle T_a \rangle)^2, \end{aligned} \quad \text{(B2)}$$

where the angle brackets denote averaging. Here the reconstruction variance is $\sigma_{T_a}^2$ and the true anomaly variance is σ_T^2. The correlation between the true anomaly and the reconstruction anomaly is r_{T,T_a}. Furthermore, we may express the reconstruction anomaly, T_a, in terms of the true anomaly, T,

$$T_a = \alpha T + \beta + R, \quad \text{(B3)}$$

where α and β are constants, and R is the random noise. Thus, the reconstruction variance can be expressed as

$$\sigma_{Ta}^2 = \alpha^2 \sigma_T^2 + \varepsilon_R^2,$$

which is the sum of the signal and noise variance. This allows the correlation to be expressed as

$$r_{T,Ta} = \frac{\alpha \sigma_T}{\sqrt{\alpha^2 \sigma_T^2 + \varepsilon_R^2}}.$$

As discussed above, the noise error variance is much smaller than the signal variance for this analysis. Thus, the correlation is approximately 1 for this reconstruction technique. This allows us to simplify (B2) and approximate the error as

$$\langle (T - T_a)^2 \rangle \approx (\sigma_T - \sigma_{T_a})^2 + (\langle T \rangle - \langle T_a \rangle)^2. \quad \text{(B4)}$$

On the right-hand side of (B4), the first term is sampling error variance, $\varepsilon_S^2 = (\sigma_T - \sigma_{Ta})^2$, and the last term is bias error variance, $\varepsilon_B^2 = (\langle T \rangle - \langle T_a \rangle)^2$.

If we could estimate the true anomaly variance for any time, then we could use (B4) to estimate the sampling error variance. However, because of changing low-frequency variations, the true variance may be difficult to estimate in all periods. The high-frequency component of the true variance may be approximately stationary. Therefore, we use the expression

$$(\varepsilon_S^2)_{\text{hf}} = (\sigma_T - \sigma_{T_a})_{\text{hf}}^2 \quad \text{(B5)}$$

to compute the high-frequency (hf) sampling error. The true hf variance is estimated from the detrended base-period data. The reconstruction hf variance is computed directly from the hf analysis. The way that the hf variations evolve is not important, so long as its variance is approximately stationary. However, there are not dense enough data through the historical period to know for certain that the hf variance is unchanged.

The low-frequency sampling error is computed by estimating how well the available sampling resolves a linear trend. A trend of 0.5°C $(100\ \text{yr})^{-1}$ is assigned everywhere on the globe. That is approximately the magnitude of the global-average temperature trend over the twentieth century. The actual temperature trend is not constant either spatially or temporally, and although the measured trend is usually positive, it is negative in some regions at some times (e.g., Folland et al. 2001b). However, for estimation of the error the actual value of the trend is not important. The magnitude of the trend needs to about right, so that the sampling error for a trend of that magnitude can be estimated. In times when the typical trend is weaker than the assumed value, as it may be before 1900 based on the available data, this may overestimate the sampling error. In periods when the trend is particularly strong

the error could be underestimated. In addition, since some parts of the low-frequency variance may fit the large-scale modes developed for the high-frequency analysis, some low-frequency variations missed by the simple analysis could be picked up by the high-frequency analysis. Thus, the error estimate for the low frequency may be too large.

For the SST only, the low-frequency sampling error computed using the method of this study was compared to the comparable estimate of Smith and Reynolds (2003). That estimate was based on how well the low-frequency SST in a climate model would be sampled using the historical network of SST observations. The climate model has increasing greenhouse gases, and the overall temperature change is similar to the observed changes (see Smith and Reynolds 2003 for more details). For global SST, the simpler low-frequency error estimate used here is nearly the same as the model-based error estimate of Smith and Reynolds (2003). This gives us confidence that our low-frequency error estimates are reasonable.

The SST bias error variance is computed from differences between the Folland and Parker (1995) and the Smith and Reynolds (2002) bias-correction estimates for the pre-1942 period. From 1942 on, the SST bias is given a minimum standard error of 0.015°C, based on typical differences between all observations and ship-intake temperatures in ICOADS data (Smith and Reynolds 2003). In addition, to account for the 1939–41 adjustment, we increase the difference in that period by a factor proportional to the magnitude of the adjustment. With maximum adjustment the factor is 2 in December 1941.

For LST, bias errors may be caused by urbanization over the twentieth century, and uncertainty due to the use of nonstandard thermometer shelters before 1950 (Jones et al. 1990; Parker 1994; Folland et al. 2001). Here we use the LST bias uncertainty estimates of Folland et al. (2001). Peterson et al. (1999) and Peterson (2003) suggests that the urbanization uncertainty may be less than that estimate. Thus, we may be overestimating that component of the error.

The SST error variance (E_S^2) and LST error variance (E_L^2) components are computed separately. They are merged using the relationship

$$E_m^2 = A_L^2 E_L^2 + A_S^2 E_S^2 + 2A_L A_S E_L E_S r_{\mathrm{LS}}, \quad \text{(B6)}$$

where A_L and A_S are the fractional areas of land and sea, respectively, and r_{LS} is the correlation between the land and sea errors. For both SST and LST, the sampling errors are largest early in the period and smallest late in the period, since sampling increases over the period for both. Thus, we may assume that sampling errors are correlated. Here we simplify the merged error variance estimate by assuming that the correlation between sampling errors is 1. However, bias errors on land and sea are caused by completely different processes, so we assume that for bias errors the correlation is 0. We use (B6) to merge the sampling and bias error variance separately, and add them for the total merged error variance.

REFERENCES

Bottomley, M., C. K. Folland, J. Hsiung, R. E. Newell, and D. E. Parker, 1990: *Global Ocean Surface Temperature Atlas "GOSTA."* Met Office and the Massachusetts Institute of Technology, 20 pp. and 313 plates.

Chelliah, M., 1990: The global climate for June–August 1989: A season of near normal conditions in the tropical Pacific. *J. Climate,* **3,** 138–160.

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

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

——, and Coauthors, 2001b: Observed climate variability and change. *Climate Change 2001: The Scientific Basis,* J. T. Houghton et al., Eds., Cambridge University Press, 99–181.

Hansen, J., R. Ruedy, J. Glascoe, and M. Sato, 1999: GISS analysis of surface temperature change. *J. Geophys. Res.,* **104,** 30 997–31 022.

Hurrell, J. W., 1995: Decadal trends in the North Atlantic Oscillation: Regional temperatures and precipitation. *Science,* **269,** 676–679.

Jones, P. D., and A. Moberg, 2003: Hemispheric and large-scale surface air temperature variations: An extensive revision and an update to 2001. *J. Climate,* **16,** 206–223.

——, P. Ya. Groisman, M. Coughlan, N. Plummer, W.-C. Wang, and T. R. Karl, 1990: Assessment of urbanization effects in time series of surface air temperature over land. *Nature,* **347,** 169–172.

——, T. J. Osborn, and K. R. Briffa, 1997: Estimating sampling errors in large-scale temperature averages. *J. Climate,* **10,** 2548–2568.

——, ——, ——, C. K. Folland, B. Horton, L. V. Alexander, D. E. Parker, and N. A. Rayner, 2001: Adjusting for sampling density in grid-box land and ocean surface temperature time series. *J. Geophys. Res.,* **106,** 3371–3380.

Kagan, R. L., 1979: *Averaging of Meteorological Fields* (in Russian). Gidrometeoizdat, 212 pp.

Kaplan, A., M. A. Cane, Y. Kushnir, A. C. Clement, M. B. Blumenthal, and B. Rajagopalan, 1998: Analyses of global sea surface temperature 1850–1991. *J. Geophys. Res.,* **103,** 18 567–18 589.

Mantua, N. J., S. R. Hare, Y. Zhang, J. M. Wallace, and R. C. Francis, 1997: A Pacific interdecadal climate oscillation with impacts on salmon production. *Bull. Amer. Meteor. Soc.,* **78,** 1069–1079.

Parker, D. E., 1994: Effects of changing exposure of thermometers at land stations. *Int. J. Climatol.,* **14,** 1–31.

——, 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.

——, L. V. Alexander, and J. Kennedy, 2004: Global and regional climate of 2003. *Weather,* **59,** 145–152.

Peterson, T. C., 2003: Assessment of urban versus rural in situ surface temperatures in the contiguous United States: No difference found. *J. Climate,* **16,** 2941–2959.

——, and R. S. Vose, 1997: An overview of the global historical climatology network temperature database. *Bull. Amer. Meteor. Soc.,* **78,** 2837–2849.

——, and Coauthors, 1998: Homogeneity adjustments of in situ atmospheric climatic data: A review. *Int. J. Climatol.,* **18,** 1493–1517.

——, K. P. Gallo, J. Lawrimore, T. W. Owen, A. Huang, and D. A. McKittrick, 1999: Global rural temperature trends. *Geophys. Res. Lett.,* **26,** 329–332.

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 sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *J. Geoghys. Res.,* **108,** 4407, doi:10.1029/2002JD002670.

Reynolds, R. W., and T. M. Smith, 1994: Improved global sea surface temperature analysis using optimal 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 for climate. *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.

——, and ——, 2003: Extended reconstruction of global sea surface temperatures based on COADS data (1854–1997). *J. Climate,* **16,** 1495–1510.

——, and ——, 2004: Improved extended reconstruction of SST (1854–1997). *J. Climate,* **17,** 2466–2477.

——, ——, 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.

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

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.