

Improvements to NOAA's Historical Merged Land–Ocean Surface Temperature Analysis (1880–2006)

THOMAS M. SMITH

NOAA/NESDIS/STAR/SCSD, and CICS/ESSIC, University of Maryland, College Park, College Park, Maryland

RICHARD W. REYNOLDS, THOMAS C. PETERSON, AND JAY LAWRIMORE

NOAA/National Climatic Data Center, Asheville, North Carolina

(Manuscript received 15 June 2007, in final form 24 September 2007)

ABSTRACT

Observations of sea surface and land–near-surface merged temperature anomalies are used to monitor climate variations and to evaluate climate simulations; therefore, it is important to make analyses of these data as accurate as possible. Analysis uncertainty occurs because of data errors and incomplete sampling over the historical period. This manuscript documents recent improvements in NOAA's merged global surface temperature anomaly analysis, monthly, in spatial 5° grid boxes. These improvements allow better analysis of temperatures throughout the record, with the greatest improvements in the late nineteenth century and since 1985. Improvements in the late nineteenth century are due to improved tuning of the analysis methods. Beginning in 1985, improvements are due to the inclusion of bias-adjusted satellite data. The old analysis (version 2) was documented in 2005, and this improved analysis is called version 3.

1. Introduction

In recent years a number of extended historical observed temperature analyses have been produced for use in climate studies and climate monitoring (e.g., Solomon et al. 2007). The extended sea surface temperature (SST) studies include Smith et al. (1996), Kaplan et al. (1998), Rayner et al. (2003, 2006), and Smith and Reynolds (2003, 2004), to name a few. Some recent analyses of extended land–near-surface temperature (LST) include Peterson and Vose (1997), Hansen et al. (2001), and Jones and Moberg (2003), and references therein. In addition, in response to the need for merged SST and LST extended analyses with error estimates a number of analyses have been produced (e.g., Parker et al. 1994; Folland et al. 2001; Jones and Moberg 2003; Smith and Reynolds 2005, hereafter SR05; Brohan et al. 2006). This series of studies by different groups has gradually increased knowledge of temperature data and analysis methods. Collectively these studies have re-

sulted in more accurate analyses with better estimates of the analysis uncertainties.

The purpose of this paper is to document improvements in the merged extended temperature reconstruction of SR05. Using methods developed in earlier studies, that analysis separately reconstructed the SST anomalies and the LST anomalies using statistical methods. The SST and LST anomalies were merged to produce a global analysis with error estimates.

The SST and LST reconstructions were each produced separately, and each was in turn the sum of two analyses. For both SST and LST, first the low-frequency (LF) or decadal-scale component of the anomaly was analyzed using averaging and filtering of the available anomalies. This nonparametric LF analysis was done first because the climate-change part of the climate signal may not be stationary. Thus, it may be poorly represented by stationary statistics used to analyze the remaining signal. The analyzed LF signal was subtracted from the anomalies and the residual high-frequency (HF) signal was analyzed. The HF analysis was performed by fitting the observed HF anomalies to a set of large-scale spatial-covariance modes. A set of weights for the modes was computed to minimize the mean-squared error of the fit. For both SST and LST,

Corresponding author address: Thomas M. Smith, NOAA/NESDIS/STAR/SCSD and CICS/ESSIC, 4114 CSS Bldg., University of Maryland, College Park, College Park, MD 20742.
E-mail: tom.smith@noaa.gov

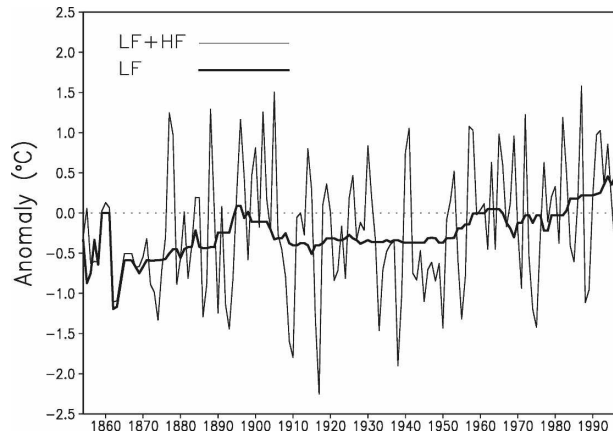


FIG. 1. LF and LF+HF SST anomaly analyses at 0°, 150°W.

the reconstruction was the sum of the LF and HF analyses.

To illustrate this, annual averages of the LF and LF + HF analysis anomalies are shown for a location in the tropical Pacific Ocean at 0°, 150°W (Fig. 1). In this location there are large interannual variations indicated by the LF + HF variations. The LF analysis gives the background climate-change variations that the interannual variations modulate. In this paper we discuss how these LF and HF analyses are produced.

The global SST and LST interdecadal variations are correlated (e.g., see SR05; Brohan et al. 2006). However, because oceans cover most of the earth's surface, roughly 70% oceans and 30% land, the SST analysis remains the more important component of the global analysis. To show the relative importance of SST to the global interdecadal variations, consider the separate SST and LST twentieth-century changes. From Trenberth et al. (2007), the 1901–2005 average trend in SST is roughly $0.067^{\circ}\text{C decade}^{-1}$, which is roughly the same SST trend from Smith and Reynolds (2004) data. For LST the Trenberth et al. (2007) trend is slightly larger at between 0.068 and $0.084^{\circ}\text{C decade}^{-1}$ for different estimates. The corresponding variance of the SST trend over this period is 0.04°C^2 , while for LST it is as much as 0.06°C^2 using the largest LST trend estimate. Because of its greater variance, the LST can affect the interdecadal signal more than may be assumed based only on the relative ocean-to-land area. However, when the trend variances are weighted by the relative area they represent, the SST-weighted variance is still 50% larger than the LST-weighted variance.

This SST analysis used here is an improved version of the Extended Reconstruction SST version 2 (ERSST.v2), developed by Smith and Reynolds (2004). The new SST analysis is referred to as ERSST.v3. Note that the ERSST.v3 analysis is monthly beginning in

1854, while the merged analysis begins in 1880. Improvements in methods used to compute ERSST.v3 are described in section 2, with some additional details given in the appendix.

The reconstructions of both SST and LST were designed to analyze signals supported by the historical sampling. Anomalies were damped toward a zero anomaly when sampling was insufficient to analyze the climate-scale signal. Here the 1971–2000 climate base period is used to form anomalies, and thus both the LF and HF analyses are damped toward this base when sampling is not adequate. Deciding how much sampling was sufficient was based on the data themselves and on estimates of spatial and temporal scales of the LF and HF components. In SR05 these decisions were conservative to ensure that data noise would not contaminate the analysis in sparse-sampling periods. A disadvantage of such conservative decisions is that they can lead to overly damped analyzed anomalies and large uncertainty early in the historical record when sampling tends to be most sparse.

Since the publication of SR05 several improvements to the analysis were developed and tested. In Smith et al. (2005) the global average was modified to exclude data from regions with sparse sampling to minimize damping of global-average anomalies. Other improvements were also considered and tested. Here a set of improvements and their effect on the reconstruction are evaluated. Of the improvements, the two that have the greatest influence on global averages are better tuning of the reconstruction method and inclusion of bias-adjusted satellite data since 1985. The following sections describe first the improvements one by one. Then the impacts of these improvements on the analysis and its uncertainty are discussed, including comparisons to the SR05 analysis.

2. Improvements to the reconstruction

Most of the improvements are justified by testing with simulated data. The simulated data are a combination of model output and observed data along with the historical sampling grid. Testing is done using data averaged to monthly 5° grid boxes, to simulate the historical data. The model output used is from the Geophysical Fluid Dynamics Laboratory (GFDL) Climate Model 2.1 (CM2.1), which was produced for Intergovernmental Panel on Climate Change (IPCC) Fourth Assessment Report (AR4; Solomon et al. 2007). This coupled general circulation model (CGCM) simulates the large-scale climate signal using variations in forcing by greenhouse gases, aerosols, and the best available estimates of solar radiation changes (Delworth et al. 2006). Surface and near-surface temperatures from an

ensemble of five runs are used. The model SST is used over the oceans, and its near-surface temperature is used to simulate station LST. This model simulates interdecadal signals with characteristics similar to the observations where data are available. However, shorter-period signals are not simulated as well by this climate model. Therefore, output is filtered to extract the model LF component. This is done using the 15-yr LF filtering described by SR05. Briefly, the LF is computed by first averaging anomalies spatially over 15° latitude–longitude moving areas and then annually. The smoothed annual averages are then median filtered using 15 annual averages to produce the LF anomaly analysis. Because the model outputs are complete, there is no damping of this test LF output. These model LF anomalies are used for the 1860–2000 test period.

To make our study more realistic, HF variations from observations are added. The HF observations are from a combination of the optimum interpolation (OI) SST over oceans and from the Global Historical Climate Network (GHCN) over land, for the recent period (Reynolds et al. 2002; Peterson and Vose 1997). To form the merged complete data, the OI SST anomalies are averaged to the monthly 5° grid boxes for 1982–2001 and merged with the GHCN 5° monthly LST anomalies. These data fill nearly all monthly 5° grid squares within 1982–2001. The remaining unfilled grid squares are filled using linear spatial interpolation of the anomalies from their nearest neighbors. Linear trends in these data are removed for the HF simulation. The 20-yr record of HF variations is split into two 10-yr periods. One 10-yr period is used to compute test reconstruction statistics, and the other independent period is used to simulate HF variations. This split ensures that the simulated data are independent of the statistics used in the test reconstructions. The HF anomalies are added to the model LF anomalies, repeated over the length of the simulated data record, 1860–2000. Thus, the same HF anomalies are repeated with a 10-yr cycle over the 1860–2000 period.

To simulate random errors in the test data, the variance from the base period is scaled by random noise-to-signal variance ratio estimates. Those random-error variance estimates at each point are reduced by dividing by the number of observations available for each 5° monthly square. Historical sampling is obtained from the International Comprehensive Ocean–Atmosphere Data Set (ICOADS) for SST and from the GHCN for LST. For SST, the noise-to-signal variance ratio estimate for ships is used, as computed by Reynolds and Smith (1994). For LST, the noise-to-signal variance ratio for an individual station was estimated by assuming a ratio of 1 for an individual observation. This is similar

to the ratio for satellite and buoy SSTs, and it may be an overestimate for station data. For example, Brohan et al. (2006) estimate that the standard error for individual station observations is 0.2°C , and typically the signal standard deviation is at least that magnitude. However, because of the large number of monthly individual LST observations that error component is greatly reduced and this crude estimate is adequate for these tests. Monthly LST values are obtained from twice-daily observations averaged over the month, so the monthly LST test noise-to-signal variance ratio is 1:60 for each station. As with SST, the LST random error variance is reduced by the number of stations in each 5° square. In each square, for each time, the random error estimate is simulated by scaling the standard error by a random number. These are produced using a random-number generator that creates normally distributed numbers with a zero mean and a standard deviation of one.

The simulated data are subsampled using the historical sampling grid and random errors are added to each square with historical sampling. These data are used to produce test temperature-anomaly reconstructions, which are validated against the full simulated data with no random errors. The global mean-squared error (MSE) of the test analyses is used to evaluate the various analysis tuning parameters, one by one.

a. Low-frequency (LF) tuning

The LF defines how much of the interdecadal climate signal may be extracted from the data, given the historical sampling. The LF is constructed by averaging and filtering data over a spatial–temporal region, defined as 10° – 15° latitude–longitude areas spatially and 15 yr in SR05. There are several cutoffs in the LF analysis to prevent averaging and filtering if there are too few data. This ensures that data noise will be filtered out of the signal. When data are too sparse the LF analysis is set to a zero anomaly. Note that all analyses are of anomalies relative to the 1971–2000 base. Here the simulated anomalies are used to find settings that minimize the global MSE relative to the fully sampled anomalies.

The settings optimized (with a short name in parentheses) include the following:

- 1) The minimum number of months with data needed to define an annual average (months per yr).
- 2) The minimum number of defined annual averages needed to define a LF (min years per LF).
- 3) The size of the spatial averaging area for the LF (area).
- 4) The maximum number of years used for median filtering in the LF estimate (max years per LF).

TABLE 1. The LF analysis parameter default settings and optimized settings (see text for definitions).

	Default	Optimized
1) Months per yr	5	2
2) Min yr per LF	5	2
3) Area	15°	25°
4) Max yr per LF	15	15

The SR05 LF settings are here referred to as the default settings. The default settings are shown in Table 1, along with the optimized settings computed here. Note that the simulated SST and LST LF anomalies were constructed separately and then combined before the global LF was optimized by testing a range of values for each of the parameters.

For both months per year and min years per LF, the optimized settings are much lower than the default settings. This shows that the LF signal can be extracted with much less sampling than had been assumed by the default analysis. The analysis is also improved when the averaging area is increased from 15° to 25°. The sensitivity to max years per LF was initially evaluated for SST by Smith and Reynolds (2003). They found that using periods of 11–25 yr yields LF analyses with similar variance, while using fewer years did not sufficiently filter out data noise and using more years damps the LF analysis. Here a similar result is obtained, with the analysis optimized when 15 yr is used. These results also show that the random errors do not overly contaminate the LF signal using these adjusted settings.

The influence of changing these parameters is shown in the merged LF average between 60°S and 60°N using the simulated data (Fig. 2). With full sampling, the influence of the 10-yr repeat cycle of the HF data is apparent. After about 1930 there is little damping in either of the test estimates. In the earlier period, most damping occurs using the default parameter settings. The damping is greatly reduced using the optimized parameters. With the improved tuning there is little damping after about 1875, but before that year there is still damping due to sparse sampling. Since the historical merged analysis will begin in 1880, these new parameter settings eliminate most LF damping error in the improved (version 3) analysis.

An example of the test LF estimates for 1890 is shown using the default settings and the optimized settings, compared to fully sampled LF analysis (Fig. 3). Note that with the optimized settings much more of the anomaly is analyzed without excessive contamination by random noise. Using the default settings, damping is excessive over much of the Pacific and over Africa, where there are few observations for this year.

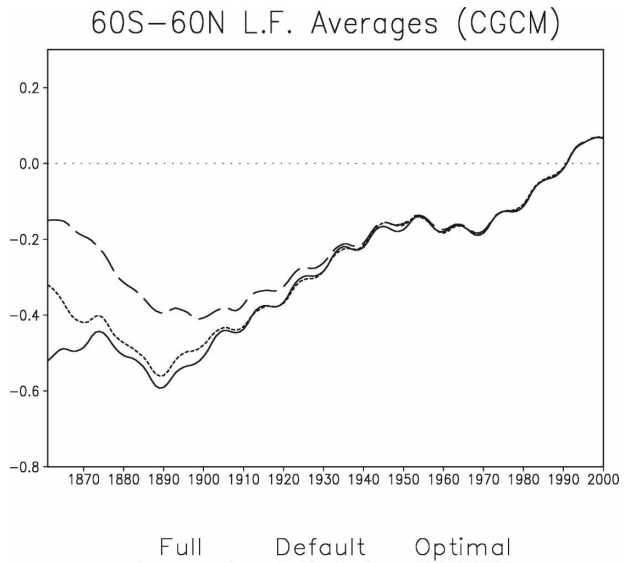


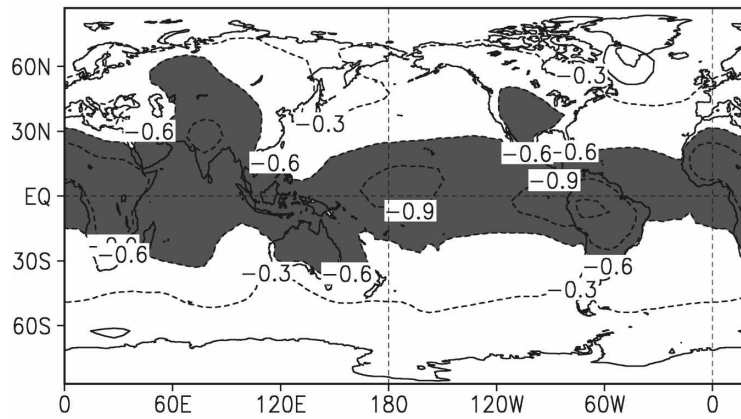
FIG. 2. The LF analysis using the fully sampled simulated data (full), and the historically sampled data reconstructed using the default SR05 parameter settings (default), and the optimized settings (optimal). Analysis damping is greatly reduced using optimal settings compared to the default settings.

Because of how the LF analysis is computed damping is due to sparse data and the effect of random errors is filtered out of the analysis. To test this, the analysis averages over only sampled regions were compared to averages using the full data. The results show that the two averages are almost indistinguishable. This is true for both the LF analysis and the HF analysis, discussed in section 2b.

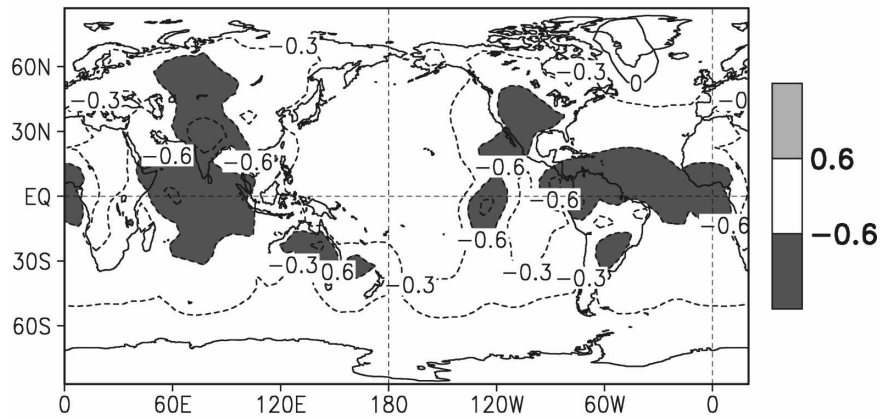
b. High-frequency (HF) tuning

As mentioned above, the reconstruction consists of first analyzing the LF anomalies, and then removing that signal from the data and analyzing the residual HF anomalies. The reconstructed anomaly is the sum of the LF and HF anomalies. The HF component is analyzed by fitting HF anomalies to a set of spatial modes. Tuning is done to find the optimum sampling needed to use a mode in the HF analysis. Each mode represents an anomaly spatial covariance pattern, and sampling is adequate when a critical percentage of the mode's variance is sampled by the available observations (Smith and Reynolds 2003). We test the critical sampling percentage by varying it between 10% and 35%. This range of critical sampling shows the minimum MSE for this parameter (Table 2). We evaluate as before to minimize the global MSE averaged in time. These tests show that between 20% and 25% sampling yields the lowest MSE, similar to the SR05 default value of 25%.

Full Grid LF, 1890



Default



Optimal

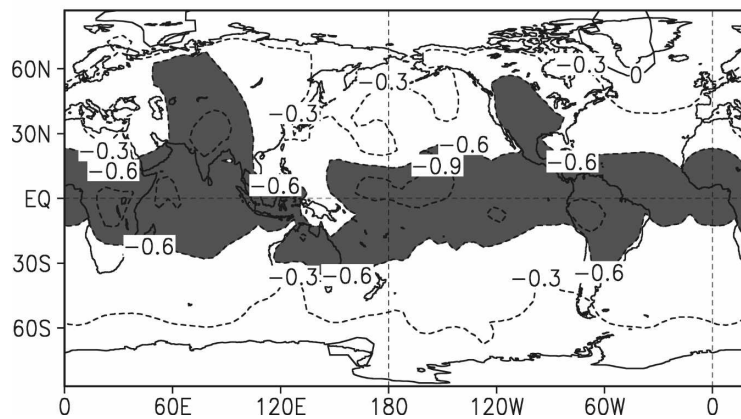


FIG. 3. (top) The LF full-grid data. Analyses using the historically sampled data, with the (middle) default and (bottom) optimized parameter settings. This example year shows the improved spatial resolution of the signal using optimal settings. Anomalies are in $^{\circ}\text{C}$, with values less than -0.6°C shaded dark and values greater than 0.6°C shaded light.

TABLE 2. Global- and time-average MSE for the HF tests, in $^{\circ}\text{C}^2$. The critical sampling (Crit) is evaluated using the global MSE averaged for 1861–99 (nineteenth century) and for 1861–2000 (full period).

Nineteenth century	MSE	
	Full	Crit (%)
0.668	0.431	10
0.615	0.409	15
0.603	0.400	20
0.608	0.399	25
0.635	0.406	30
0.653	0.412	35

Thus, most overall improvements in the historical reconstruction are from the LF analysis.

A potential deficiency in the SR05 analysis is its tendency to underrepresent interannual variations when sampling is sparse. That is because the HF analysis will damp anomalies in regions where too few modes are chosen for the analysis. For example, the Niño-3.4 (5°S – 5°N , 120° – 170°W) area SSTs may be slightly damped early in the twentieth century, as discussed in section 2c (V. Kousky 2006, personal communication; see Fig. 8). To minimize this potential HF damping the new analysis uses a 20% sampling cutoff for modes, which is the lowest global value justified by these tests. When we compute the MSE for the Niño-3.4 area only, and for the nineteenth century, the MSE is also minimized using 20% sampling.

c. Sampling cutoffs for large-scale averaging

The above results show that the reconstructions can be improved in periods with sparse sampling. However, there can still be damping errors in periods with sparse sampling. Damping of large-scale averages may be reduced by eliminating poorly sampled regions because anomalies in those regions may be greatly damped. In Smith et al. (2005) error estimates were used to show that most Arctic and Antarctic anomalies are unreliable and those regions were removed from the global-average computation. Here testing using the simulated data is done to find objectively when regions should be eliminated from the global average to minimize the MSE of the average compared to the full data.

To define spatial sampling for each reconstructed 5° latitude–longitude area, the surrounding 25° latitude–longitude region is examined. This larger area is used because both the LF and the HF analyses use surrounding data to analyze the central area. The percentage of the 25° area that is sampled by the historical sampling grid is computed for each month. If the percentage falls

below a given value, then the central 5° anomaly is excluded from the average. Tuning defines the critical sampling value.

Using the default SR05 parameters, the merged global MSE is minimized when the 25° region has at least 35% sampling. Using the improved tuning discussed above, the MSE is minimized when there is at least 20% sampling. The improved parameters yield a lower optimal sampling for global averages because they produce a less-damped analysis in the presence of sparse sampling. However, even with the improved parameters the MSE for global averages can be reduced by omitting some sparsely sampled regions. When these tests are repeated using sampling regions larger than 25° , the minimum MSE is still found using a 20% sampling cutoff. However, for larger sampling regions that minimum MSE is larger than for 25° sampling regions.

Because of the larger relative ocean area, the SSTs are most strongly affected by the tuning improvements discussed in sections 2a–c. Next we specifically focus on changes in the SST analysis in sections 2d–e, and on changes in the LST analysis in section 2f.

d. Bias-adjusted satellite SSTs

Since 1985 the Advanced Very High Resolution Radiometer (AVHRR) Pathfinder day and night satellite SST observations are available (Kilpatrick et al. 2001). These data improve SST sampling, especially in the Southern Ocean, and they are incorporated into the improved ERSST.v3 analysis. However, before they may be used, the satellite SSTs require bias adjustments, as discussed by Reynolds and Smith (1994).

The satellite data biases are usually associated with aerosols and clouds, both of which cause a cool bias. For example, both day and night satellite observations show strong tropical biases associated with the Mount Pinatubo eruption in 1991. Compared to the operational AVHRR SSTs used by Reynolds et al. (2002), initial Pathfinder bias variability is lower because of the more careful processing of this delayed-time product.

The satellite SSTs are bias adjusted relative to the merged ship and buoy SSTs. Adjustments are produced using analyses similar to the HF SST analyses. Separate analyses of the in situ and satellite SSTs are produced using only spatial modes adequately sampled by both data types. The difference between the analyses defines the satellite bias. Using only modes sampled by both data types removes the sampling bias from the separate analyses, and ensures that their differences are caused by data biases.

Separate adjustments are performed for day and night satellite SSTs. After adjustment, all data types are

TABLE 3. Relative weights of different types of SST data used in analysis.

Ship	Buoy	Day satellite	Night satellite
1	7	6	19

merged to form the adjusted merged data used in the statistical analysis. In merging the SSTs the relative weights for ships, buoys, day satellites, and night satellites are given in Table 3. These weights are based on the relative noise of the different data types, as estimated by Reynolds and Smith (1994). All available data types are used to form the merged data. The weighted sum of the available data types is computed using these weights normalized by the sum of the weights. That normalization ensures that there is no damping or inflation of the merged SST.

Since most of the oceans are adequately sampled by in situ data, the influence of satellite data is greatest in the Southern Ocean. South of about 45°S , the satellite data cause a slight cooling of the SSTs, which results in a slight reduction in the near-global (in situ + sats) average compared to the in situ analysis (Fig. 4). The difference in the average caused by including satellite data is only about 10% of the anomaly for the most recent years.

Because the Southern Ocean is sparsely sampled by in situ data, and most in situ data in that region are buoy SSTs, we performed some more detailed testing of the influence of satellite data in that region. A test region was chosen where the in situ analysis sometimes differs greatly from the combined satellite and in situ analysis ($55^{\circ}\text{--}45^{\circ}\text{S}$, $160^{\circ}\text{--}170^{\circ}\text{W}$). Averages of both analyses in this region showed that they are usually

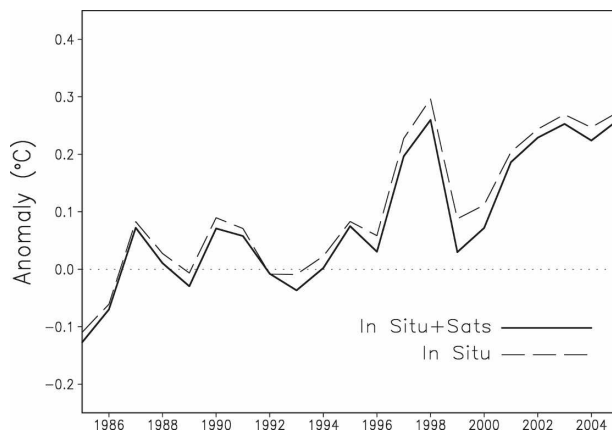


FIG. 4. Annual reconstructed SST anomaly averaged between 60°S and 60°N for analyses with and without bias-adjusted satellite data. Satellite data are bias adjusted relative to the in situ data in regions where both are available.

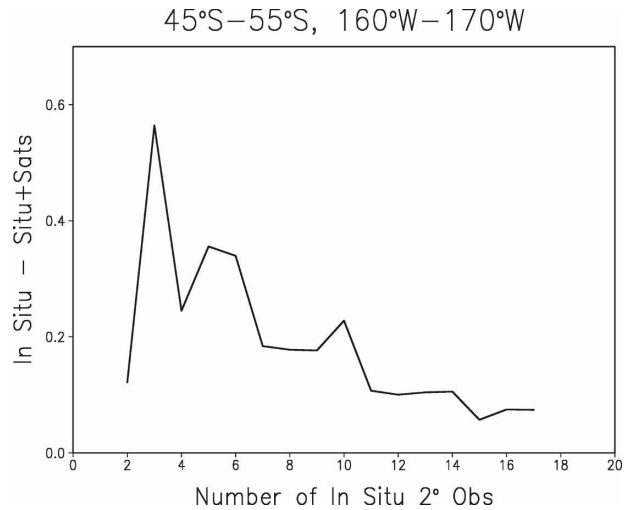


FIG. 5. Analyzed differences (in situ minus the in situ and satellite analysis) as a function of the number of in situ observations. The maximum possible number of in situ observations in the region is 25. The monthly differences are from the indicated test region and are averaged over 1985–2005.

similar, but in periods when in situ sampling is sparse they can have large differences. This difference occurs because when in situ sampling is sparse, the dominant analysis mode for the region is not sampled by the in situ data while satellite data always sampled the mode. When its sampling is too sparse to resolve that mode, the in situ-only analysis anomaly is damped toward a zero anomaly while the satellite anomaly is not damped. Differences are largest and somewhat erratic when few 2° squares within the test region are sampled, although even with in situ sampling available the satellite data tend to always cool the analysis slightly (Fig. 5). For low numbers of in situ data some of the difference may be due to in situ noise.

Some satellite bias adjustment may be computed when the local in situ sampling is sparse even if the dominant mode is missing. A residual adjustment may occur due to the influence of other modes. Thus, some bias adjustment may still be computed for the region based on more remote in situ and satellite data. However, these remotely based adjustments are weaker than more locally based adjustments. This increases the uncertainty in the analysis when local in situ data are not available, although satellite data should still improve the Southern Ocean analysis by resolving anomalies that would otherwise be greatly damped. However, as Fig. 5 indicates, the local bias uncertainty in those cases may be as large as 0.5°C .

In addition to biases in satellite data, there are other data biases. The most important additional data bias may be the ship–buoy bias (Kent and Taylor 2006;

Rayner et al. 2006). This relative bias is important because of the growing number of buoy SSTs since the mid-1980s (e.g., Reynolds et al. 2002). Before 1985 most in situ SSTs are ship measurements. Where both ship and buoy observations are available, the ships are typically about 0.1°C warmer. However, the bias is not constant in either space or time where both data types are available. In addition, before the mid-1980s there are few buoy observations so directly analyzing the bias from data over the full reconstruction period is not possible.

Because ships tend to be biased warm relative to buoys and because of the increase in the number of buoys and the decrease in the number of ships, the merged in situ data without bias adjustment can have a cool bias relative to data with no ship–buoy bias. As buoys become more important to the in situ record, that bias can increase. Since the 1980s the SST in most areas has been warming. The increasing negative bias due to the increase in buoys tends to reduce this recent warming. This change in observations makes the in situ temperatures up to about 0.1°C cooler than they would be without bias. At present, methods for removing the ship–buoy bias are being developed and tested.

Besides ship–buoy biases, there are also biases in the expendable bathythermograph (XBT) temperatures (Gouretski and Koltermann 2007). The XBTs are included in the ICOADS observations, so their biases affect the analysis from the 1950s to 1984 when ICOADS data are used. The XBTs constitute the major part of subsurface sampling, and correcting their biases reduces the warm subsurface upper-ocean temperature anomaly by about a third. However, the XBTs account for less than 5% of the ICOADS SST sampling when they are available. Thus, their biases should have a minor influence on the pre-1985 analysis and we do not adjust for them in this analysis. The much larger problem is the ship–buoy biases beginning 1985.

e. Sea ice impact on the SST analysis

In SR05 sea ice concentration of 0.6 (60%) and above, are used to linearly damp the analyzed SSTs toward the freezing temperature of seawater, -1.8°C , for concentrations between 0.6 and 0.9. Concentrations below 0.6 have no effect on the SST, and above 0.9 the SST is set to the freezing temperature. Variations in the historical sea ice coverage have little effect on global-average temperatures, but they can be important locally. Historical estimates of ice concentration are based on observations from ships and aircraft, and they are most suitable for seasonal and decadal variations in sea ice concentrations. Here SST–ice adjustments are

applied as in SR05. Differences from SR05 are from improved bias adjustments applied to satellite-based sea ice concentrations.

Beginning in the late 1970s, satellites have been used to estimate sea ice concentrations. However, satellite sea ice concentrations cannot separate summer melt ponds on the sea ice surface from the open ocean. This results in an ice concentration bias in warm seasons. Rayner et al. (2003) provided a method to correct these biases based on the 1979–99 data. The corrections were computed using the same method used for the sea ice concentrations in Reynolds et al. (2002) and extended through the present. These adjusted sea ice concentrations were used in SR05. The impacts of these adjustments in SR05 were relatively small because only ice concentrations ≥ 0.6 were used in SR05. In addition, the final monthly product of Reynolds et al. (2002) could be delayed up to 10 days after the first of the month. To eliminate this time delay in updates of the new analysis, new corrections were produced using satellite sea ice concentrations from Cavalieri et al. (1999) and the Reynolds et al. (2002) concentrations. The adjustment is a function of the satellite concentration, computed seasonally for each hemisphere for 2000–04. Each concentration is adjusted with a constant computed to minimize the error of the adjusted value. There is no adjustment for satellite concentrations less than 40% or equal to 100%. For in-between concentrations the adjustment removes most bias caused by ice ponds. Differences in SST caused by the adjustments are almost always less than 0.1°C , and typically much less. Differences are only important locally and they have almost no effect on hemispheric or larger-scale spatial averages. The adjustment was tested and found to be stable after 1995. Updates of ERSST.v3 will use this new bias-adjusted sea ice beginning in 2000. The sea ice concentration adjustment to SSTs is otherwise the same as in SR05.

The effect of ERSST.v3 sea ice adjustments on global-average SSTs is only about 0.01°C . However, ice adjustments can cool SSTs by 0.2°C or more locally in the marginal-ice zones. As noted above, the SST–ice adjustment is only applied for a sea ice concentration ≥ 0.6 . In the large-scale averages of merged LST and SST anomalies discussed in section 2f, SSTs with ice concentrations > 0.5 are treated as missing regions that do not contribute to the average. Thus, sea ice adjustments do not influence the large-scale averages of merged temperature anomalies discussed in section 2f, although the masking out of those regions increases the sampling-error estimates of those large-scale averages.

TABLE 4. Global- and time-average of merged temperature anomalies. All time series are recentered so that the 1971–2000 average is zero. Shown are time averages for the improved reconstruction (merged.v3), SR05, Quayle et al. (1999), and HadCRUT3v. All values are in °C.

	1880–1900	1901–50	1951–2000	2002–04
Merged.v3	−0.38	−0.32	−0.07	0.34
SR05	−0.36	−0.31	−0.08	0.35
Quayle et al. (1999)	−0.37	−0.31	−0.07	0.35
HadCRUT3v	−0.43	−0.36	−0.08	0.37

f. Reinjecting in situ land data

In 5° squares with GHCN sampling the LST is much more reliable than in regions filled by interpolation. The interpolation filters and smoothes the LST in all regions, including regions with sampling where the unfiltered GHCN should be more accurate. To minimize differences between the GHCN and the analysis where sampling is available, the GHCN anomalies are reinjected into the analysis after the statistical interpolation. In 5° squares with no GHCN sampling, no adjustment is done.

The number of GHCN stations in 5° latitude–longitude regions is used to determine how strongly the LST reinjection should be. With only one station in a 5° square the GHCN is less reliable and the statistical analysis is more heavily dependent on it in that situation. With more stations in the square, the GHCN is more reliable. The relative weight of the GHCN anomaly as a function of the number of stations in a 5° square, n , is computed by

$$W_G = 1 - e^{-n/3}.$$

The statistical reconstruction is assigned the weight $1 - W_G$. Thus, with $n = 2$ stations the GHCN and statistical reconstruction are about equally weighted, and with more stations the GHCN dominates. This adjustment increases the LST spatial variations, but it has almost no effect on hemispheric and larger-scale spatial averages.

3. Improved analysis

Here the new merged reconstruction (merged.v3) is discussed and compared to other analyses. Compared to Quayle et al. (1999) and the SR05 analysis, there is little difference in the global average temperatures for most of the analysis period (Table 4). Another analysis used for comparisons is the merged analysis produced jointly by the Met Office Hadley Centre and the University of East Anglia’s Climatic Research Unit [the Hadley Centre Climatic Research Unit Temperature dataset version 3 (HadCRUT3v); Brohan et al. (2006)].

TABLE 5. Global-annual standard error averaged for the given periods. Shown is the merged land–sea temperature standard errors, for the improved reconstruction (merged.v3) and SR05. All values are in °C.

	1880–1900	1901–50	1951–2000
Merged.v3	0.09	0.07	0.03
SR05	0.13	0.11	0.06

Those comparisons show slightly larger differences with HadCRUT3v giving slightly larger interdecadal changes. For the other comparisons in Table 4, the greatest differences from the merged.v3 are early in the analysis period when the merged.v3 produces stronger anomalies due to better tuning.

Compared to the SR05 error estimates, the merged.v3 analysis has lower error (Table 5). In addition, the total global error estimates of Brohan et al. (2006) are similar to the merged.v3 total global error estimates. Merged.v3 error estimates account for sampling and bias errors. Bias errors include uncertainty in historical SST bias adjustments, uncertainty from land shelter types, and from land-use changes at the station locations such as urbanization. Error estimation methods are described in detail in SR05. The only difference in the current study is that for recent years the urbanization uncertainty is assigned a maximum equal to its value for the year 2000 while in SR05 the urbanization uncertainty continues to grow linearly with time. As with other error components, the urbanization uncertainty widens the error estimates on both sides of the expected value. Given that recent research by Parker (2004, 2006) and Peterson and Owen (2005), indicates that “the effects of urbanization and land use change on land-based temperature record are negligible” (Trenberth et al. 2007), the urbanization error used in this analysis is likely overestimated.

In SR05 the low-frequency sampling dominates the global-average error, while in the improved analysis that error component is much smaller because of the improved tuning and the bias error now dominates. For all years, the SR05 analysis is within the 95% confidence levels of the merged.v3 analysis. In an independent study Brohan et al. (2006) similarly separated and evaluated errors. They also found that the most important components to the global-average error are sampling error from limited spatial coverage and bias uncertainty, and computed total error estimates similar to those computed here. However, here the bias error is the largest component while in Brohan et al. (2006) error from the limited spatial coverage is larger. This is in part because Brohan et al. (2006) do not interpolate to fill all locations, so their sampling error for the

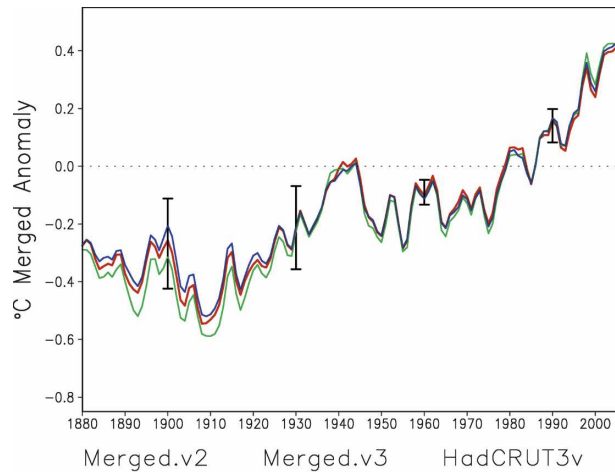


FIG. 6. Global and annual merged temperature anomalies from SR05 (merged.v2), the improved analysis (merged.v3), and from HadCRUT3v. The base periods are adjusted to match the 1971–2000 base used in the improved analysis. A 1–2–1 smoother has been applied to each time series of annual averages. The 95% confidence limits reflecting all errors estimated for merged.v3 are drawn for the years 1900, 1930, 1960, and 1990.

global average is larger. The difference is also in part because the SST bias uncertainty estimates used here are slightly larger than those used by Brohan et al. (2006).

Comparison of the merged.v3 global and annual-average temperature anomalies to SR05 (merged.v2) and Brohan et al. (2006; HadCRUT3v) anomalies (Fig. 6) shows that all have similar global variations throughout the analysis period. Differences are less than the 95% confidence limits of the merged.v3 analysis. As noted above, the overall errors for merged.v3 and Brohan et al. (2006) are also similar for most of the analysis period. For the recent period, since 1950, the merged.v3 errors are slightly smaller than the Brohan et al. (2006) estimates. Note that the confidence limits are wide early in the twentieth century due to insufficient sampling and bias uncertainty. They decrease greatly between 1930 and 1950 due to increased sampling, and they increase slightly after 1950 due to increasing urbanization bias uncertainty. As discussed above, we may overestimate the recent urbanization bias uncertainty in this analysis.

Differences are most noticeable before 1930, when anomaly damping causes SR05 to have a slightly weaker anomaly while HadCRUT3v has a slightly stronger anomaly in that period. The addition of satellite data also causes a slight cooling of the merged.v3 analysis after about 2000. That is due to the cooler satellite Southern Ocean SSTs, a region poorly sampled by ship data although there are Southern Ocean buoys.

TABLE 6. Rankings of the 10 warmest global and annual averages from 1880 to 2006 for SR05 (merged.v2) and the improved analysis (merged.v3).

Rank	Merged.v2		Merged.v3	
	Year	Anomaly	Year	Anomaly
1	2005	0.41	2005	0.40
2	1998	0.38	1998	0.37
3	2002	0.36	2003	0.36
4	2003	0.36	2002	0.35
5	2006	0.35	2006	0.33
6	2004	0.34	2004	0.32
7	2001	0.30	2001	0.29
8	1997	0.27	1997	0.25
9	1999	0.20	1995	0.18
10	1995	0.19	1999	0.17

Note that if the ship–buoy bias were also adjusted with respect to the ships, then the most recent years would be warmer, because the ship–buoy difference tends to be positive and because of the increasing number of buoy observations. However, as discussed above, these differences are well within the 95% confidence limits.

Although the correlations are high between SR05 and merged.v3, the additional bias-adjusted data for the recent years affect the rankings. The rankings of the warmest 10 yr are similar for both, with 2005 the warmest for both followed by 1998, a year with a strong warm ENSO episode (Table 6). The few changes in rankings occur where there is little difference between the years, when slight changes in the analyses and the input data can cause a shift. Note that the error estimates for the recent period (Table 5) indicate that there is no significant difference between the warmest years shown in Table 6. In preliminary testing of ship–buoy bias adjustments, the most recent 5 yr are warmed slightly when that bias adjustment is applied. The adjustments would also change the rankings slightly. The fact that the 10 warmest years occur in the last 12 yr of this 127-yr record is more significant than the precise order of the rankings.

The improved analysis shows similar variations when compared to other analyses, as expected. Correlations between merged.v3 and HadCRUT3v are used to show their similar variations over most regions (Fig. 7). All month anomaly correlations of the 5° monthly data are computed. The figure shows correlations for the full 1880–2006 period. Correlations are computed only for regions that have at least 30 pairs of merged.v3 and HadCRUT3v data, which omits some 5° squares because HadCRUT3v does not use interpolation to fill all regions. For the full 1880–2006 period, correlations are highest between 45°S and 70°N, with an average value of 0.74. In addition, correlations are computed using data from 1900 to 1949 and from 1950 to 1999 (not

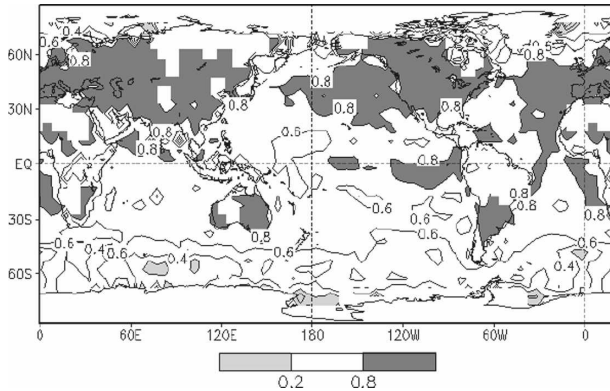


FIG. 7. Anomaly all month temperature correlation of merged.v3 and HadCRUT3v on the 5° grid for 1880–2006.

shown). For 1900–49 the average correlation for this region is 0.68, while for 1950–99 it is 0.77. This difference suggests that the better sampling in the second half of the twentieth century improves the comparisons. Outside of this region sampling is sparser and correlations are lower.

Similar comments may be made of comparison of ERSST.v3 to the Hadley Centre Sea Ice and SST analysis (HadISST; Rayner et al. 2003). One region of particular interest is the Niño-3.4 area (5°S–5°N, 120°–170W°). The all-month anomaly correlation of HadISST with ERSST.v3 in this region for 1880–1997 is 0.90. Both analyses are clearly producing consistent interannual variations. But there are important differences in this region in periods when sampling is sparse. In Niño-3.4 prior to 1950, HadISST is biased about 0.3°C warmer than ERSST.v3. Much of the bias is due to the use of different historical bias adjustments in the two analyses prior to 1942. Another important difference depends on the method used to compute low-frequency variations. In HadISST they are computed by fitting data to a global mode, while here simpler averaging and filtering is used, as discussed above.

Changes in the Niño-3.4 SST anomalies between ERSST.v2 and ERSST.v3 are very small after 1950. Earlier in the record the two are also highly correlated, but there are times when the ERSST.v2 anomaly is greatly damped from a lack of sampling (Fig. 8). These times include years before 1880 and around 1918, shown more clearly in the difference (Fig. 8, bottom panel). The improved tuning used for ERSST.v3 allows these variations to be more reliably resolved. Because the merged.v3 analysis begins in 1880, SST damping from before that year does not affect the merged analysis. However, as the figure shows there are other times in the record when sparse sampling could have produced damping in the old (merged.v2) analysis.

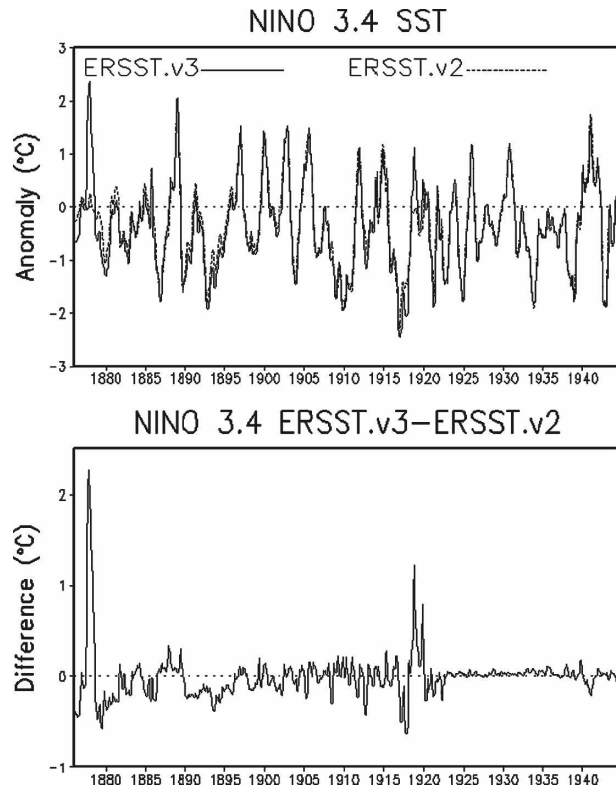


FIG. 8. Niño-3.4 SST anomalies from (top) ERSST.v2 and ERSST.v3, 1876–1944. (bottom) The ERSST.v3 – ERSST.v2 difference. Monthly values of each are smoothed using a 3-month running mean. Note ERSST.v2 damping is before 1880 and near 1918.

4. Summary

Simulated data from models and observations are used to improve the tuning of the National Oceanic and Atmospheric Administration (NOAA) operational surface temperature analysis. Errors from excessive damping are reduced in the improved analysis (merged.v3). This is especially important in the ocean component of the analysis (ERSST.v3). Compared to SR05, the greatest improvements occur in the nineteenth century. However, there are some sparsely sampled regions in all periods that are improved by the new tuning. In addition, global averaging of the analysis is optimally tuned to exclude undersampled regions responsible for excessive damping of global averages.

In addition to improvements from better tuning, other improvements are also incorporated. Bias-adjusted AVHRR satellite SSTs are added, which gives better resolution in the Southern Ocean. Because the satellite data are filtered by the analysis modes, including satellite data does not cause a large jump in the analyzed variance and has little effect outside the Southern Ocean. Improved sea ice analyses are avail-

able with near-real-time updates. These satellite-based ice analyses are bias adjusted to reduce the effect of melt ponds and incorporated to improve the high-latitude SST analysis.

Large-scale averages of the improved merged.v3 are similar to earlier analyses, although the improvements reduce damping errors and uncertainty estimates early in the record. This allows the climate-change signal to be measured more accurately. Regions with sparse sampling have the greatest analysis uncertainty, especially the Arctic and Antarctic regions where sampling is always sparse. However, even with these limitations the data are sufficient to indicate global twentieth-century warming of roughly $0.7^{\circ} \pm 0.2^{\circ}\text{C}$.

Although this analysis contains a number of improvements, more improvements are possible. Satellite analyses for land temperatures are beginning to be developed (e.g., see Jin 2004 and references therein). There are problems associated with LST analyses from satellite data due to contamination from snow and ice, but if these problems can be overcome the satellite data could contribute to improved recent-period LST analysis. A future LST historical reconstruction could thus be improved using land satellite temperature to improve reconstruction statistics over land and also to help fill sampling gaps in the recent period.

Several refinements to the error estimates are also possible. Uncertainty in the satellite bias adjustments could be incorporated in the error estimate. For the land temperatures it is assumed that urbanization causes an increasing bias error through the second half of the twentieth century, as in Folland et al. (2001). However, more recent studies indicate that urbanization-error estimate may be too large. For example, the IPCC's assessment of many recent studies indicates that the impact of urbanization is "negligible" (Trenberth et al. 2007). In addition, the studies of Peterson et al. (1999) and Peterson (2003) suggest that the urbanization error is much less than is assumed here. Additional refinements are unlikely to change the basic character of the large-scale variations, but they could improve the analysis locally and further reduce its uncertainty.

Acknowledgments. This study was encouraged and assisted by discussions and comments from many others. Contributors to the discussions include T. Karl, D. Easterling, R. Vose, J. Bates, D. Kim, J. Privette, S. Leduc, V. Kousky, and Y. Xue. This paper was also improved by the suggestions of two anonymous reviewers. The views and opinions, and findings contained in this report are those of the authors and should not be construed as an official NOAA or U.S. government position, policy, or decision.

APPENDIX

Improved Extended Reconstructed SST (ERSST.v3)

A major component of the improved merged land and ocean temperature analysis is the improved Extended Reconstructed SST version 3 (ERSST.v3). Improvements in analysis methods between ERSST.v2 and ERSST.v3 are discussed in section 2 and the impact of those improvements is discussed in section 3. The major differences between ERSST.v2 and ERSST.v3 are summarized here. This is intended to better describe ERSST.v3 for readers especially interested in the SST component of the analysis.

The period of record for both ERSST.v2 and ERSST.v3 is the same, monthly beginning 1854. Both historical analyses are based on ICOADS SST anomalies and both use the same historical bias adjustment (Smith and Reynolds 2002). For the historical analyses, roughly before 1980 when only in situ data are available, the major differences are caused by the improved tuning of ERSST.v3. As discussed in section 2a, the biggest change occurs due to the improved tuning are in the low-frequency (LF) analysis. However, as discussed in section 2b and in section 3, resolution of interannual variance is also improved by the high-frequency (HF) tuning. Changes in the sea ice to SST analysis produce only minor differences.

The ERSST.v3 is improved by explicitly including bias-adjusted satellite infrared SST estimates. In ERSST.v2 and ERSST.v3, information from satellites is indirectly included because the HF analyses are based on modes computed from the Reynolds et al. (2002) analysis, which includes the satellite data. In ERSST.v3 the Pathfinder infrared SST estimates are introduced in the analysis by combining those SST data with ship and buoy data. Satellite SSTs are bias adjusted relative to the ship and buoy data as previously discussed. The SST estimates from satellite, ships, and buoys are merged using a weighted sum of the different inputs, with weights inversely proportional to the noise estimate for each type (see section 2d). The merged SSTs are used in the ERSST.v3 analysis. In ERSST.v2 only in situ SSTs are used. The greatest influence of the satellite data is to produce greater variability south of 45°S beginning in 1985. In most other regions the influence of satellite data is small because of generally sufficient in situ monthly sampling in the recent period.

Overall improvements in ERSST.v3 variance can be seen by comparing the global spatial standard deviations of ERSST.v3, ERSST.v2, and OI.v2 since 1985 (Fig. A1, top panel). All three reflect the same interannual variations. The OI.v2 variations are strongest

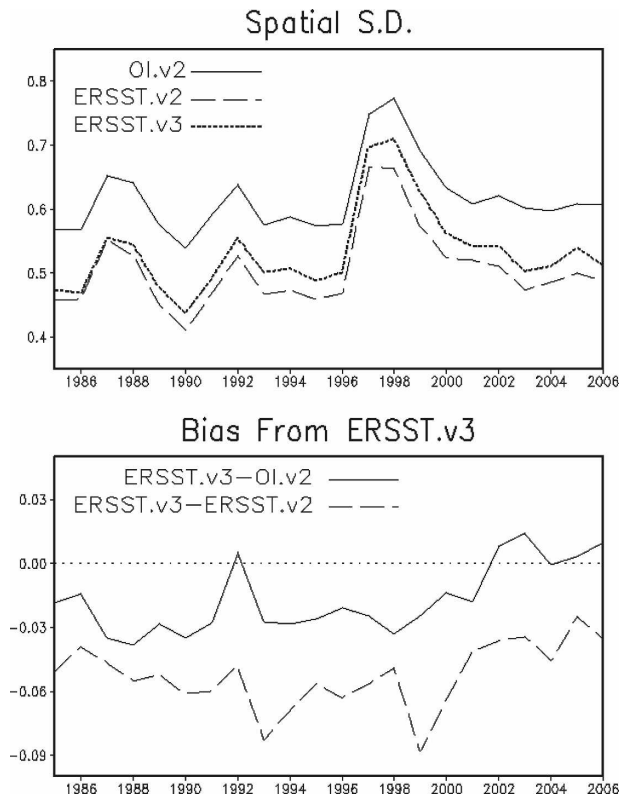


FIG. A1. Global spatial statistics ($^{\circ}\text{C}$). (top) Spatial std dev from OI.v2, ERSST.v2, and ERSST.v3. (bottom) ERSST.v3 – OI.v2 and ERSST.v3 – ERSST.v2 biases.

because that analysis does not filter satellite data through spatial modes, as the reconstructions do. However, even with the mode filtering, ERSST.v3 resolves more variance than ERSST.v2, mostly because of the better resolution of Southern Ocean variations. Relative to ERSST.v3, the bias of OI.v2 is lower than ERSST.v2 (Fig. A1, bottom panel). That is because both ERSST.v3 and OI.v2 incorporate bias-adjusted satellite data.

REFERENCES

- Brohan, P., J. J. Kennedy, I. Harris, S. F. B. Tett, and P. D. Jones, 2006: Uncertainty estimates in regional and global observed temperature changes: A new data set from 1850. *J. Geophys. Res.*, **111**, D12106, doi:10.1029/2005JD006548.
- Cavalieri, D. J., C. L. Parkinson, P. Gloersen, J. C. Comiso, and H. J. Zwally, 1999: Deriving long-term time series of sea ice cover from satellite passive-microwave multisensor data sets. *J. Geophys. Res.*, **104**, 15 803–15 814.
- Delworth, T. L., and Coauthors, 2006: GFDL's CM2 global coupled climate models. Part I: Formulation and simulation characteristics. *J. Climate*, **19**, 643–674.
- Folland, C. K., and Coauthors, 2001: Global temperature change and its uncertainties since 1861. *Geophys. Res. Lett.*, **28**, 2621–2624.
- Gouretski, V., and K. P. Koltermann, 2007: How much is the ocean really warming? *Geophys. Res. Lett.*, **34**, L01610, doi:10.1029/2006GL027834.
- Hansen, J., R. Ruedy, M. Stao, M. Imhoff, W. Lawrence, D. Easterling, T. Peterson, and T. Karl, 2001: A closer look at United States and global surface temperature change. *J. Geophys. Res.*, **106**, 23 947–23 963.
- Jin, M., 2004: Analysis of land skin temperature using AVHRR observations. *Bull. Amer. Meteor. Soc.*, **85**, 587–600.
- 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.
- Kaplan, A., 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.
- Kent, E. C., and P. K. Taylor, 2006: Toward estimating climatic trends in SST. Part I: Methods of measurement. *J. Atmos. Oceanic Technol.*, **23**, 464–475.
- Kilpatrick, K. A., G. P. Podesta, and R. Evans, 2001: Overview of the NOAA/NASA advanced very high resolution radiometer Pathfinder algorithm for sea surface temperature and associated matchup database. *J. Geophys. Res.*, **106**, 9179–9198.
- Parker, D. E., 2004: Large-scale warming is not urban. *Nature*, **432**, 290.
- , 2006: A demonstration that large-scale warming is not urban. *J. Climate*, **19**, 2882–2895.
- , P. D. Jones, A. Bevan, and C. K. Folland, 1994: Interdecadal changes of surface temperature since the late 19th century. *J. Geophys. Res.*, **99**, 14 373–14 399.
- Peterson, T. C., 2003: Assessment of urban versus rural in situ surface temperatures in the contiguous United States: No difference found. *J. Climate*, **18**, 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 T. W. Owen, 2005: Urban heat island assessment: Metadata are important. *J. Climate*, **18**, 2637–2646.
- , 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.
- Quayle, R. G., T. C. Peterson, A. N. Basist, and C. S. Godfrey, 1999: An operational near real time global temperature index. *Geophys. Res. Lett.*, **26**, 333–336.
- 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. Geophys. Res.*, **108**, 4407, doi:10.1029/2002JD002670.
- , P. Brohan, D. E. Parker, C. K. Folland, J. J. Kennedy, M. Vanicek, T. J. Ansell, and S. F. B. Tett, 2006: Improved analyses of changes and uncertainties in sea surface temperature measured in situ since the mid-nineteenth century: The HadSST2 dataset. *J. Climate*, **19**, 446–469.
- 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. *J. Climate*, **15**, 1609–1625.
- Smith, T. M., and R. W. Reynolds, 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.
- , and —, 2005: A global merged land air and sea surface temperature reconstruction based on historical observations (1880–1997). *J. Climate*, **18**, 2021–2036.
- , —, R. E. Livezey, and D. C. Stokes, 1996: Reconstruction of sea surface temperature using empirical orthogonal functions. *J. Climate*, **9**, 1403–1420.
- , T. C. Peterson, J. Lawrimore, and R. W. Reynolds, 2005: New surface temperature analyses for climate monitoring. *Geophys. Res. Lett.*, **32**, L14712, doi:10.1029/2005GL02342.
- Solomon, S., D. Qin, M. Manning, M. Marquis, K. Averyt, M. M. B. Tignor, H. L. Miller Jr., and Z. Chen, Eds., 2007: *Climate Change 2007: The Physical Science Basis*. Cambridge University Press, 996 pp.
- Trenberth, K. E., and Coauthors, 2007: Observations: Atmospheric surface and climate change. *Climate Change 2007: The Physical Science Basis*, S. Solomon et al., Eds., Cambridge University Press, 235–336.