# Reconstructions Improvements Using Iteratively Adjusted Statistics, Demonstrated Using Model-Output Annual SST Anomalies and Historical Sampling

THOMAS M. SMITH

*NOAA/NESDIS/STAR, and Earth System Science Interdisciplinary Center, and Cooperative Institute for Climate and Satellites, University of Maryland, College Park, College Park, Maryland*

(Manuscript received 8 March 2016, in final form 28 June 2016)

ABSTRACT

Historical reconstructions of climate fields, such as sea surface temperature (SST), are important for climate studies and monitoring. Reconstructions use statistics from a well-sampled base period to analyze a sparsely sampled historical period. Here a method is shown for adjusting the base-period statistics using the available historical data so that statistics better represent historical variations. The method is demonstrated using annual SST anomalies from a coupled GCM historical run, 1861–2005, forced by greenhouse gases and aerosols. Simulated data are constructed from the model's SST using observed historical SST sampling with error estimates added. Reconstructions are performed using the simulated data, and the results are compared to the full model SST without added errors. The results from applying other reconstruction methods to the simulated data are compared. The tests show that the method improves annual SST reconstructions, especially in the early years, when sampling is most sparse and in the extratropics. In particular, the 1881–1900 correlation averaged over 30°–60°S and over 30°–60°N improves from about 0.4 using noniterative reconstruction to about 0.6 using iterative reconstruction. The correlations of annual values in the tropics are about 0.7 with both methods. Incorporating those improvements into an SST reconstruction could better represent extratropical climate variations in the nineteenth and early twentieth centuries, and improve the value of the reconstruction for long-period climate studies and for validating climate models.

## 1. Introduction

Reconstructions of historical climate fields are analyses that use statistics to compute complete fields based on limited historical observational data. Reconstructions based on historical observations are useful for long-term climate studies and for monitoring climate variations. Typically, the statistics are based on modern analyses that include satellite data to define spatially complete statistics. Reconstructions of historical sea surface temperature (SST) anomalies have been developed at several centers (e.g., Smith et al. 1996; Kaplan et al. 1998; Rayner et al. 2003). At the National Oceanic and Atmospheric Administration (NOAA), they were developed by Smith et al. (1996) for the period beginning 1950, using statistics based on the merged in situ and satellite analysis of Reynolds and Smith (1994). An Extended Reconstruction of SST (ERSST), for the period beginning 1854, was later developed by Smith and Reynolds (2003). The NOAA ERSST analysis has been updated and improved periodically, and the current version is by Huang et al. (2015), but the basic methods are from Smith and Reynolds (2003). This study describes a potential improvement in ERSST methods that can improve the representation of multidecadal to interannual variations.

Improvements described here are most important in the extratropics and before the mid-twentieth century, when sampling is relatively sparse. Interannual to multidecadal components of climate modes can be better represented for that early period, including modes such as the Pacific decadal oscillation (PDO; Mantua et al. 1997), the North Atlantic Oscillation (NAO; Hurrell 1995), and Southern Hemisphere annular modes (Thompson and Wallace 2000). As discussed later, the improvements in the tropics, including the representation of El Niño–Southern Oscillation (ENSO), are

*Corresponding author address*: Thomas Smith, ESSIC, University of Maryland, College Park, 5825 University Research Court, Suite 4001, College Park, MD 20740.
E-mail: tom.smith@noaa.gov

DOI: 10.1175/JTECH-D-16-0059.1

smaller. Better representations of nineteenth-century to early twentieth-century SST variations are important for multidecadal climate studies, including studies that link oceanic and continental variations and climate model validation. Therefore, the improvements discussed here could lead to improved understanding of long-term climate variations.

ERSST is a combination of two analyses (Smith and Reynolds 2003). First, an annual low-frequency analysis is produced using a moving 15-yr window of observed SST anomalies to compute an annual low-frequency anomaly. An annual average of SST anomalies is computed for each of the 15 years in the moving window, and those annual averages are smoothed, filled, and spatially expanded using large-scale averaging. For regions far from any sampling, the anomaly is set to zero. The median of the 15 years is then used to define the low-frequency analysis for the center year, which is used as a first-guess analysis for ERSST. Monthly increments from the first guess are computed by fitting the available observed monthly increment anomalies to a set of spatial covariance modes. The spatial modes are based on the Reynolds and Smith (1994) monthly analysis. The two-part approach is used for ERSST because the satellite-based SST analysis may not represent multidecadal changes that occur before the satellite base period. However, there are potential problems with the ERSST low-frequency estimate due to sparse sampling in the nineteenth and early twentieth centuries. In addition, the need for spatial filtering to fill regions degrades its resolution.

An alternative method for analysis of multidecadal variations was developed by Rayner et al. (2003). They used combined in situ and satellite observations, when available, for 1901–97. The SST anomalies are spatially averaged to a 4° grid and are averaged to seasons [January–March (JFM), April–June (AMJ), July–September (JAS), and October–December (OND)]. Seasonal data are then low-pass filtered using a Chebyshev filter to remove variations with periods less than 8 years. For each season the leading empirical orthogonal function (EOF) of the filled and smoothed data is used to define the multidecadal signal for their SST analysis. As with the ERSST method, their method requires spatial and temporal smoothing due to sparse sampling, and even with the filtering there are regions where sampling is too sparse to a define multidecadal EOF.

Here a method is discussed for computing an annual average SST reconstruction that could be used as an improved first guess for an extended monthly or seasonal reconstruction. This method uses EOFs that are iteratively improved by statistical reinjection of the available historical observations. The incorporation historical data with incomplete sampling throughout the period of record gives the EOFs the ability to better resolve large-scale variations over the full record. Here the method is tested using coupled atmosphere–ocean general circulation model (CGCM) output SSTs, with observed sampling and error estimates. Results are validated against the full model SST. The new method improves both the annual averages and the low-frequency estimates before about 1950. It also gives a more spatially complete representation of multidecadal variations due to improvements in teleconnection patterns in the leading EOFs. Improvements are greatest in the extratropics. In the next sections, the model output, observed SST sampling, and simulated errors are described. Following that is a description of the new reconstruction method, a discussion of results, and a summary and conclusions.

## 2. SST and sampling

The SST anomalies used for testing are output from a CGCM. For the annual analysis, this is subsampled using a historical sampling grid. Realistic historical errors are estimated and added to the subsampled data before analysis. Several historical error estimates are tested to evaluate their influence on the results. One estimate is of only random errors, and the other is of combined random and correlated errors. A comparison to the fully sampled SST anomalies is used to show the relative quality of different methods.

### *a. Model-output SSTs*

The CGCM output was downloaded from NOAA's Geophysical Fluid Dynamics Laboratory (GFDL) data portal (http://nomads.gfdl.noaa.gov/). Output from the GFDL-ESM2M historical run was used, produced as part of the CMIP5 experiment discussed by Taylor et al. (2012). The historical run CGCM forcing includes greenhouse gases, ozone, and aerosols. Further documentation of the model is available online at the GFDL data portal. Monthly surface temperatures for 1861–2005 were obtained on an irregular model grid and averaged to a regular 5° grid. Land regions were masked out for estimates of model SST. The 1971–2000 annual cycle was removed to form monthly anomalies, which were then averaged annually.

The CGCM SST anomalies are used to simulate observed SST anomalies. For the testing to be valid, the model should have variations similar to the observed variations. Over the 1982–2005 overlap period, the model's overall pattern of annual SST standard deviation is similar to the optimum interpolation (OI; Reynolds and Smith 1994) observed standard deviation (Fig. 1). The OI analysis includes both satellite and

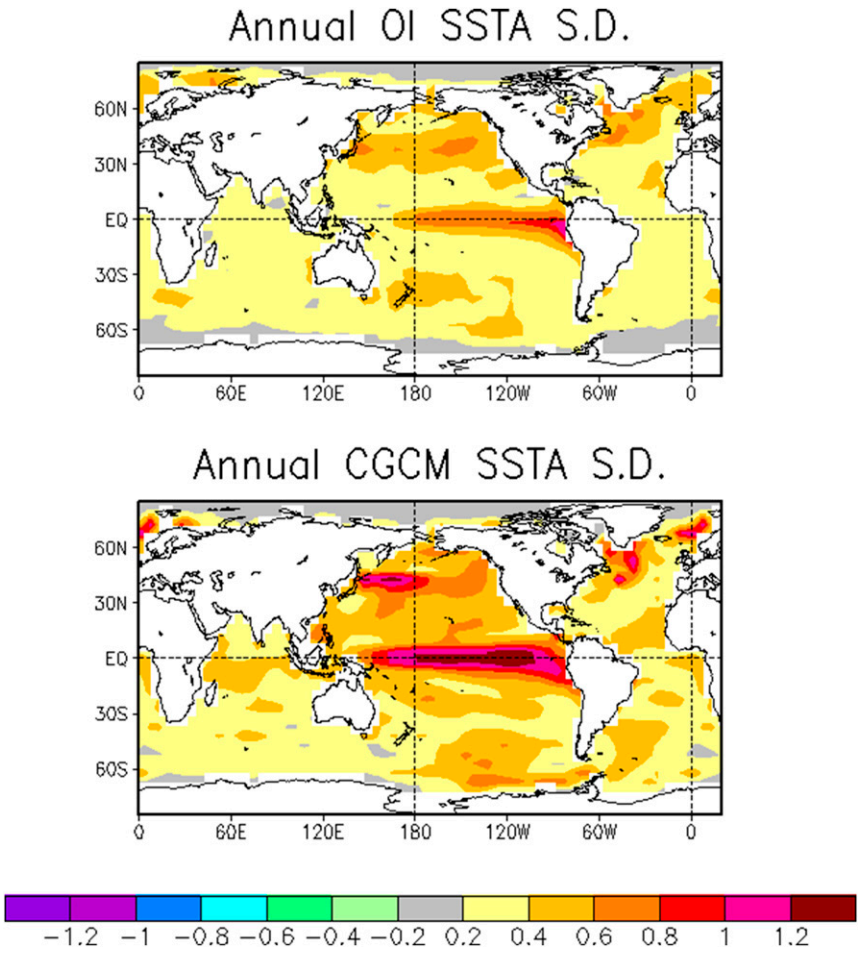


FIG. 1. Standard deviation of the 1982–2005 annual average SST anomalies from the (top) OI analysis and (bottom) CGCM.

in situ observations, and is here considered a good estimate of the true annual SST variations. This figure also shows the oceanic reconstruction regions tested. In both the model's and the observed standard deviations, there are higher values in the tropical Pacific and extratropical North Pacific and North Atlantic in similar regions. The model's standard deviation is higher than the observed OI standard deviation. But since the regions of high and low standard deviation are similar, the influence of regional sampling changes should be similar for both. There are slight differences in the standard deviation patterns. For example, the model's tropical and North Pacific standard deviation maxima are shifted slightly west of the observed value. However, the overall patterns are close and testing using the model output should give a useful measure of the relative quality of different reconstruction methods.

### *b. Observed historical sampling*

Historical 5° monthly SST anomalies and sampling are obtained from the Hadley Centre SST analysis, version 3 (HadSST; Kennedy et al. 2011a,b). The HadSST sampling contains the number of individual observations averaged in each 5° monthly region and is available beginning in 1850 with periodic updates.

Model annual average anomalies are used only in the tests for years at locations where the number of months sampled is at least four. The fraction of annual sampling is lowest before 1880, with a gradual increase in global sampling, except for interruptions associated with the two world wars in the first half of the twentieth century (Fig. 2). For the area north of 45°N, the annual sampling is lower but otherwise similar to the global sampling. In the area south of 45°S, there is modest sampling before 1914, when the Panama Canal opened and ships no longer needed to take the long southern route. The loss of sampling south of 45°S in the 1910s is 100% for several years, suggesting that the loss is largely caused by changes in shipping routes associated with the opening of the canal. The reduction in sampling in the 1910s is less severe north of 45°S, and is about 50% from 45°S to

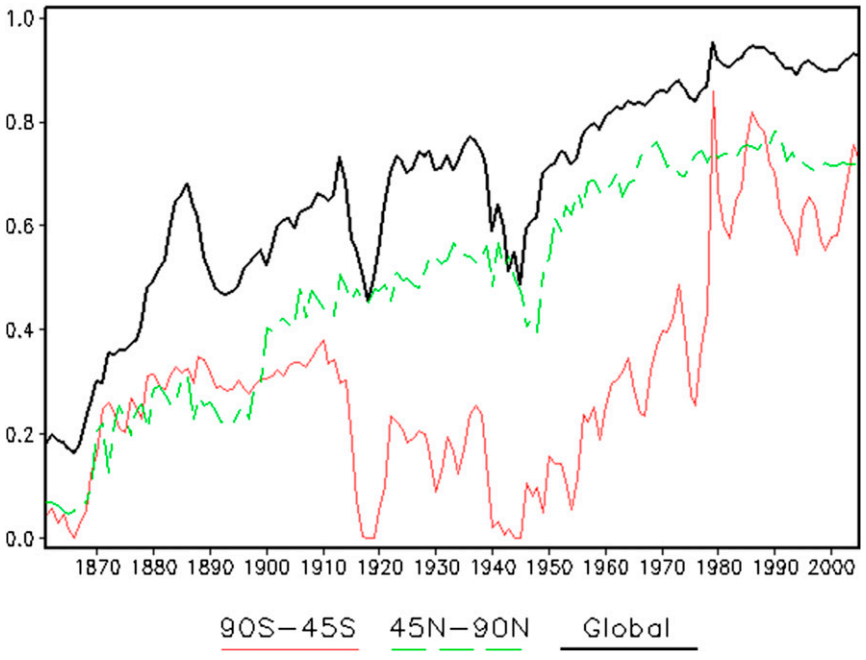


FIG. 2. Fraction of the indicated area sampled with 5° annual averages from HadSST observations.

the equator and 10% from the equator to 45°N. South of 45°S the sampling is relatively low from the 1910s to the 1940s, and the loss is again 100% in the early 1940s, apparently due to the Second World War. In the 1940s sampling decreases in all latitude bands, although the decreases are most severe in the Southern Hemisphere. After the 1950s sampling begins increasing. South of 45°S sampling is relatively high since the early 1980s, in large part due to more buoy sampling in the region.

Spatial patterns of sampling are shown for two years, 1896 and 1946 (Fig. 3). The 1896 sampling is typical for the late nineteenth century, and that year is also used for comparisons discussed later. For 1896 there is good annual sampling for most regions between about 50°S and 60°N, except for the western Pacific, which has a large gap. The 1946 sampling shows how sampling progressed over 50 years. The influence of the Panama Canal is apparent, and there is much more cross-Pacific Northern Hemisphere sampling. However, in some parts of the Southern Hemisphere sampling is less in 1946, as suggested by Fig. 2.

### c. *Random error estimates*

The noise-to-signal variance ratio for individual ship observations $\eta_0^2$ was measured by Reynolds and Smith (1994) and was found to be approximately 15, which is used here. Taking the square root and multiplying by a typical standard deviation estimate of 0.3°C gives the standard random error for an individual ship observation, roughly 1.2°C, which is similar to the midlatitude estimate of Kent et al. (1999). For each 5° monthly average the noise-to-signal variance is

$$\eta_{5\mathrm{d}}^2(N) = \eta_0^2/N, \tag{1}$$

where $N$ is the sum of the number of individual observations for all months averaged for the annual average, which is used to estimate the random standard error, $\varepsilon_R = R\sigma\eta_{5\mathrm{d}}$, where $R$ is a pseudorandom number with a zero mean and a standard deviation of 1, and $\sigma$ is the signal standard deviation, estimated using the model base standard deviation (shown in Fig. 1).

Error estimates are added to the simulated data before reconstruction to test their influence on the results. Both the random estimate and a partly correlated error estimate, described below, are tested.

### d. *Partially correlated errors*

Besides random errors, data can have correlated data errors. While random errors are greatly damped by averaging, correlated errors are not reduced by averaging and therefore can more strongly influence an analysis. In this subsection the amount of correlation in errors is estimated to evaluate how it influences the analysis. This partly correlated error estimate contains both a random and a correlated error component, and is intended to

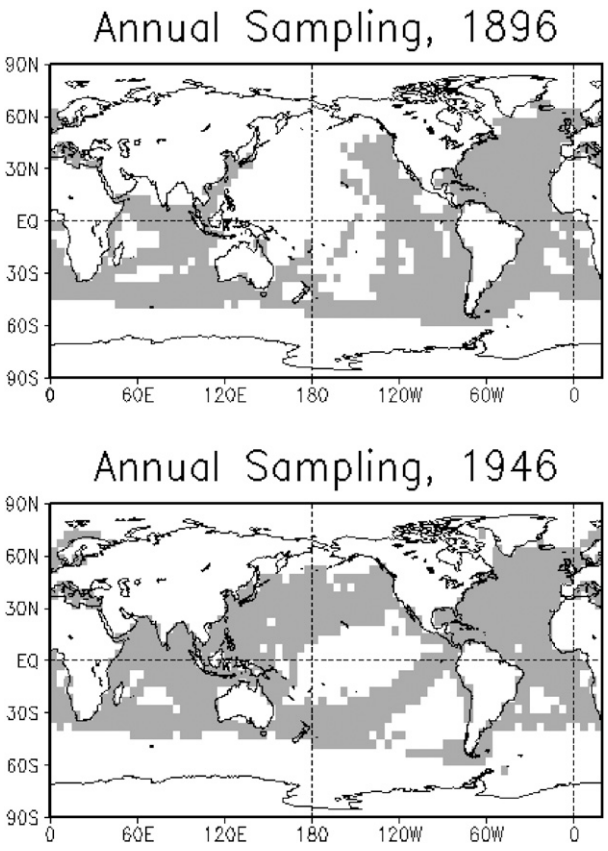


FIG. 3. Maps of annual sampling from HadSST 5° observations for the indicated years.

estimate the type of errors that may occur in HadSST or similar datasets used to represent monthly averages on a grid.

To estimate partly correlated errors, the HadSST errors relative to OI are evaluated. The estimates are computed using OI data averaged to the 5° grid. Although the OI also contains errors, its errors are here assumed to be much smaller than HadSST errors because of the much denser OI sampling from combined satellite and in situ observations. By contrast, HadSST uses only in situ observations. For this estimate HadSST and OI are used for 1982–2006, their overlap period when all inputs for HadSST are from ICOADS. After 2006 HadSST uses SST updates that may be noisier and thus may not be representative of the historical period. The average 1982–2006 HadSST–OI anomaly difference is removed, since the bias could be caused by climatology differences. The remaining differences are used to estimate partly correlated estimate errors. This gives an estimate of the typical errors that could be expected in historical monthly averages. Causes of the errors may include the representation of 5° monthly averages from incomplete sampling and instrument biases, as well as random errors. The representativeness error for 5° monthly averages has the potential to be large because HadSST observations are from different times of the day and month, as well as from different parts of the 5° square. Because of the satellite sampling, the OI representativeness error should be much smaller.

The monthly average error variance is computed as a function of the number of HadSST 5° observations $N$. At each location the error variance with $N$ is estimated using anomaly differences from all months. The error variance is normalized by the OI SST anomaly variance computed using all months, to remove differences due to spatial changes in SST variance. Spatial averages of the normalized error variances are used to evaluate normalized changes with $N$. If all errors have the same variance and they are random, then the error variance should change with $N$ as shown in Eq. (1).

If errors are partly correlated, then they will get smaller with $N$ more slowly than in Eq. (1). This is essentially a problem of finding the variance of a mean given the variance of individual correlated values being averaged (e.g., Smith et al. 1994). If the error variances averaged are constant and the correlations between errors are also constant, then the normalized error of the mean is

$$\eta_C^2(N) = \frac{\eta_0^2}{N}[1 + (N-1)r]. \tag{2}$$

If the correlation $r$ is zero, then there is no correlated error and Eq. (2) reduces to the random error estimate. If errors are perfectly correlated, then $r = 1$ and there is no reduction in error variance with sampling.

The global averages of normalized error variance estimates with $N$ are clearly not random, which would be the case for $r = 0$ (Table 1). The observed relationship does not always show reduced error with larger $N$, perhaps due to insufficient sampling for each value of $N$. However, the overall fit to Eq. (2) is best with $r = 0.7$. This estimate also shows that for the 5° monthly regions, the normalized error for individual observations is less than the Reynolds and Smith (1994) estimate, which was used to estimate random error $\varepsilon_R$. Here the average $N = 1$ normalized error is found to be about 0.5.

This relationship is used to estimate partly correlated errors over the historical period as follows. For each month from 1982 to 2006, the partly correlated error is computed from the difference between HadISST and OI SST anomalies. When HadISST $N > 1$, the error estimate is adjusted to the $N = 1$ error by scaling it by

$$\frac{\eta_0}{\eta_C(N)} = \frac{N}{1 + (N-1)r}. \tag{3}$$

TABLE 1. Global averages of the observed normalized error variance as a function of the number of observations $N$ listed under Obs. Also shown are estimates using Eq. (2) with different values of $r$ and with the $N = 1$ value set to 0.5.

| $N$ | Obs | $r = 0.0$ | 0.1 | 0.3 | 0.5 | 0.7 | 0.9 |
|---|---|---|---|---|---|---|---|
| 1 | 0.49 | 0.50 | 0.50 | 0.50 | 0.50 | 0.50 | 0.50 |
| 2 | 0.45 | 0.25 | 0.28 | 0.32 | 0.38 | 0.43 | 0.47 |
| 3 | 0.41 | 0.17 | 0.20 | 0.27 | 0.33 | 0.40 | 0.47 |
| 4 | 0.31 | 0.12 | 0.16 | 0.24 | 0.31 | 0.39 | 0.46 |
| 5 | 0.41 | 0.10 | 0.14 | 0.22 | 0.30 | 0.38 | 0.46 |
| 6 | 0.41 | 0.08 | 0.12 | 0.21 | 0.29 | 0.38 | 0.46 |
| 7 | 0.54 | 0.07 | 0.11 | 0.20 | 0.29 | 0.37 | 0.46 |
| 8 | 0.47 | 0.06 | 0.11 | 0.19 | 0.28 | 0.37 | 0.46 |
| 9 | 0.46 | 0.06 | 0.10 | 0.19 | 0.28 | 0.37 | 0.46 |
| 10 | 0.36 | 0.05 | 0.10 | 0.19 | 0.28 | 0.36 | 0.45 |

For large $N$ this scaling factor approaches $1/r$, which is about 1.4 for $r = 0.7$. The scaling allows maps of $N = 1$ errors to be computed for this value of $r$.

There are a few locations where the error could not be computed, mostly in the Arctic and Southern Ocean. Those remaining blanks in the error maps are filled so that the maps may be used to estimate errors in historical periods when sampling may be different. Simple interpolation filling may not be appropriate, since interpolated errors are spatially correlated perfectly. Those regions are filled with a combination of half random and half interpolated $N = 1$ error. The random part is chosen from random on the map and the spatial interpolation of errors is from surrounding regions. Since these fill areas that have few data in the historical period, the filling should minimally influence the results discussed later.

The monthly partly correlated errors for each year from 1982 to 2006 are used to estimate historical annual correlated errors. For each historical year, one of the years in the period 1982–2006 is randomly selected. For each month and each location, the $N = 1$ partly correlated error is scaled by the inverse of Eq. (3), $[1 + (N - 1)r]/N$, where $N$ is the number of observations and $r = 0.7$. These monthly correlated errors are averaged to give the annual partly correlated error estimate. In a later section, the influence on reconstructions of both random and correlated error estimates is tested.

## 3. Reconstruction methods

The reconstructions are computed by fitting the historical simulated sparse grid data a set of 10 global EOF modes. The basic method was described by Smith et al. (1996), and the iterative method is described in some detail in the appendix. How the method is used is described here.

A limited number of global EOFs are used for the reconstruction. Global EOFs are used because we wish to maximize all reasonable teleconnections because of the sparse sampling over much of the analysis period. Here the first 10 EOFs are used. Higher EOFs often describe processes with small space–time scales and resolving them is difficult or impossible using the sparse historical sampling. In addition, since the first 10 EOFs explain 89% of the annual anomaly base-period variance, the ability of higher modes to improve analysis is limited. More tuning could be performed to find the optimal number of modes to use, but the results indicate that this is a good number for testing the method. To further avoid including undersampled modes in historical periods with sparse sampling, those with less than 5% of their variance sampled for a given year are screened out of the analysis. The screening method is described in Smith and Reynolds (2003) and in the appendix.

Note that annual anomalies are analyzed. Before subannual anomalies could be analyzed using this method, they would need to be tested and tuned for the different time and space scales, and for the different sampling available, for subannual anomalies. Subannual variations typically have smaller time and space scales so more modes may be needed. However, sampling for individual months is typically less than annual sampling, and the larger number of modes may not be supported by the sparser subannual sampling.

The monthly anomaly increment analysis of Huang et al. (2015) does not use EOFs, but instead uses empirical orthogonal teleconnection (EOT) patterns. The EOTs are useful for that application because it is easy to limit the spatial scales of each EOT, to limit teleconnections that are not directly supported by observations. Monthly anomaly increments tend to have smaller scales than annual anomalies, which is desirable for their application. Here we wish to maximize teleconnections and therefore use a set of global EOFs.

Earlier reconstructions using EOF modes employ a stationary set of modes (e.g., Smith et al. 1996). Here, after an initial reconstruction, the modes are updated to include historical information, and then another reconstruction using the updated modes is computed. The process is repeated until the variance of the reconstruction stabilizes. This allows for a global set of EOF modes to be computed, and it allows for variations in those modes to be influenced by data from the entire reconstruction period. Variations that are only partially sampled by historical data cannot be fully represented in the adjusted EOFs, but the results show that a low number of adjusted modes can resolve many variations that are not resolved by the base-period modes, without

TABLE 2. Time averages of spatial correlation with the full model annual (ann) average and low-frequency estimate over 1861–2005. Values are shown using filtered data without simulated errors and using data with either random or correlated errors. For filtered data 10 EOFs are computed using 1861–2005 data (filter) and the 1982–2005 satellite period, FSP. Reconstructions use EOFs from the satellite period, RSP and iteratively-adjusted EOFs, Iter. Correlation with the model low-frequency include the ERSST method low-frequency estimate, LF(G), and the low-frequency estimate of the Iter reconstruction, LF(R).

| | Random | Correlated |
|---|---|---|
| Ann filter | 0.82 | 0.82 |
| Ann FSP | 0.68 | 0.68 |
| Ann RSP | 0.64 | 0.63 |
| Ann Iter | 0.79 | 0.74 |
| LF(G) | 0.72 | 0.66 |
| LF(R) | 0.83 | 0.77 |

producing physically unreasonable features associated with sampling.

Here the initial reconstruction uses the 1982–2005 base-period modes. For SST reconstructions the satellite period begins in 1982, so we use that base period to simulate practical conditions. The reconstruction EOF modes for the next iteration are adjusted as follows. First, the reconstructed anomalies over the entire 1861–2005 period are adjusted using anomaly increments at sampling locations using the statistical adjustment method described in the appendix. Then a new set of EOFs is computed using the adjusted 1861–2005 anomalies. These new EOFs are used to compute an updated reconstruction.

The anomaly increment analysis is done using an OI of the simulated observed anomaly with errors minus the reconstruction. The OI noise-to-signal variance ratio is defined by the number of observations, as shown by Eq. (1), and uses the random error estimate discussed in section 2b. The OI spatial correlation scales are set to 1000 km zonally $\times$ 500 km meridionally (roughly two grid boxes by one grid box at the equator). In addition, the OI analysis at each grid box only uses local data, within two grid boxes zonally and within one grid box meridionally. A minimum of two defined values within that range is required for computing an analysis. If there are too few data, then the increment analysis is set to 0, and if an analysis is computed with sparse and noisy data, then it will also be damped toward 0. The reinjection is therefore not a replacement of the previous reconstruction but statistical blending using analyzed increments, with stronger blending where data are more reliable.

Changes in the variance of the analysis are used to determine whether an additional iteration is justified. Here the global spatial variance is computed and averaged over the entire reanalysis period for this measure. If this average variance changes by less than 5% between iterations, then additional iterations are not computed.

## 4. Results

Reconstructions based on historically sampled data with simulated errors are validated against the full data. First, the influence of random and partly correlated errors is evaluated, both for annual and low-frequency signals. The low-frequency signal is the 15-yr median. It is computed using the annual reconstructions and by using ERSST methods of data averaging, for comparisons. That testing shows that the partly correlated errors create larger reconstruction errors than the purely random errors. The reconstruction that uses simulated data with partly correlated errors is further evaluated, since that may better represent the quality that can be expected from real historical data.

### *a. Testing different error estimates*

Different error estimates are tested to evaluate their sensitivity to random error estimates, described in section 2c, and the correlated error estimates, described in section 2d. Comparisons are made against EOF-filtered data. The EOF-filtered SST anomalies are computed by filtering using 10 EOFs computed from spatially complete and noise-free model output over the full period, 1861–2005. This method defines the best possible fit that can be expected from a 10-EOF reconstruction. The mean spatial correlation of the unfiltered and filtered model output is 0.82 (Table 2). Filtering using 10 EOFs computed from the satellite period, 1982–2005, gives the best reconstruction that can be obtained using the unadjusted EOFs, and is labeled the filter and satellite period (FSP). The FSP is lower than filtering using modes computed for the full period because the FSP is not long enough to resolve some of the historical variations.

Next, reconstructions are considered using historically sampled model output with simulated errors. The reconstruction using satellite-period modes, and historical sampling with simulated errors [the reconstruction satellite period (RSP)] gives an average correlation slightly lower than the FSP correlation. The different error estimates have little influence on the overall RSP correlation. The reconstruction using iteratively adjusted EOFs gives much better correlation, approaching the filter correlation for the random error test but lower for the correlated error test. The comparison shows that the historical sampling is adequate for annual reconstructions over most of the period, and that the

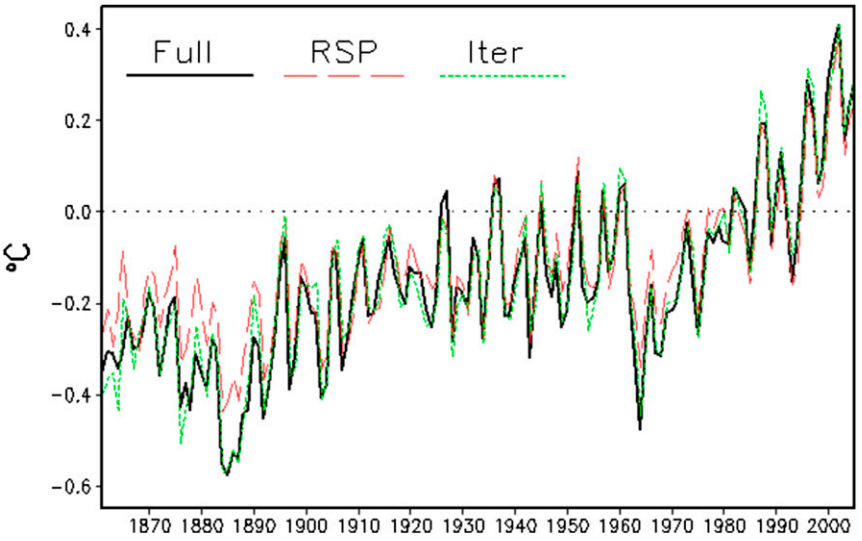


FIG. 4. Global averages of annual anomalies from the full model output, the RSP analysis, and the Iter analysis.

reconstruction method can remove most random errors. It also shows that correlated errors influence reconstructions more than random errors, although the iterative method still improves the analysis compared to RSP.

Correlations of the model low-frequency signal against estimates using the ERSST methods to fill gapped data [LF(G)] and the low-frequency signal of the iteratively adjusted EOF reconstruction [LF(R)] indicate that correlated errors most strongly influence both, and that the LF(R) is more reliable. These comparisons show that the iterative-adjusted EOF is better compared to RSP. The worst realistic case is the reconstruction using the correlated error estimates. In the following sections, further evaluation is done for the reconstruction using iteratively adjusted EOFs and correlated error estimates.

### *b. Annual results using correlated errors*

First consider global annual averages, comparing the full model output to reconstructions using satellite-period EOFs (RSP) and using the iteratively adjusted EOFs (Iter; Fig. 4). The satellite-period analysis is the first iteration, before historical data are reinjected. Here the Iter analysis is the second iteration. Additional iterations made little change to the reconstruction variance.

The full model output and the two reconstructions all have similar global interannual and multidecadal variations, but the RSP tends to be weaker over much of the historical period, especially before 1900. The global Iter analysis more closely follows the global average of the full model output.

The global spatial correlation with the full model output is next considered to show how well the spatial patterns of the reconstruction match the full model patterns. Here, in addition to RSP and Iter, the filtered data are also correlated with the full data. The Iter analysis correlation is only slightly lower than the filtered model correlation over most of the analysis period (Fig. 5). Before 1870 the Iter correlation is much lower due to a greater reduction in sampling for that period. The RSP correlation is slightly higher in the RSP base period, 1982–2005, because its modes are computed from that period and therefore can resolve more details for that brief base period. However, RSP correlation is noticeably lower in all other periods. This shows that before the satellite period an RSP analysis will have larger errors, and that sampling may be insufficient for a reliable annual analysis before 1870.

Spatial patterns for 1896 are considered next. This model year has a warm ENSO episode; it is also a year when the RSP shows much lower spatial correlation. Maps of anomalies (Fig. 6) show that the RSP analysis resolves the warm tropical Pacific anomaly and much of the North Pacific and tropical Atlantic anomalies. But it fails to resolve anomalies in the North Atlantic, parts of the North Pacific, and in the southern oceans. While there are errors in the Iter analysis, it is better at resolving anomalies in most regions, accounting for its higher spatial correlation.

The RSP analysis resolves low latitudes reasonably well, but it has more difficulty in mid- and high latitudes, which is typical for the late nineteenth century, as shown by maps of temporal correlation over 1881–1900 (Fig. 7). Correlation maps are of annual RSP and Iter analyses

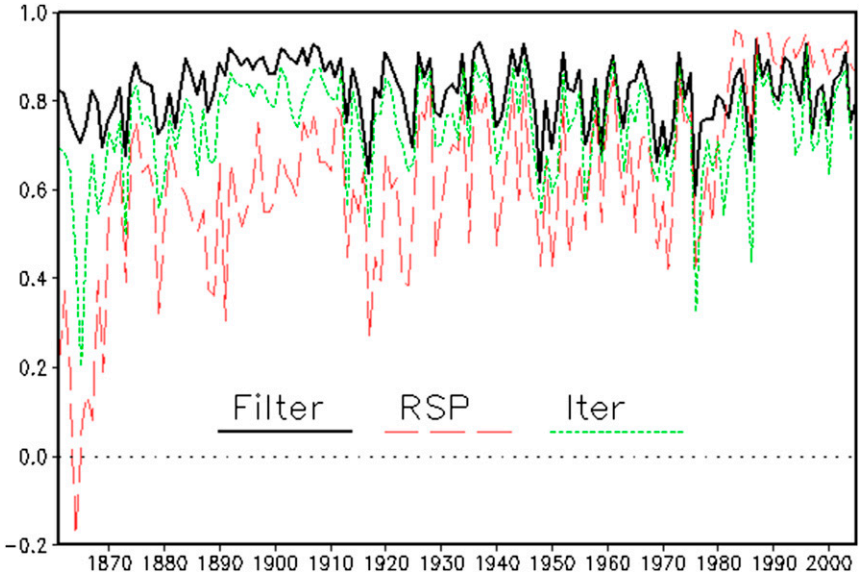


FIG. 5. Global spatial correlation with full data anomalies of the filtered data, the RSP analysis, and the Iter analysis.

against the full model annual average. The Iter analysis improvements are clearest south of 30°S and north of 45°N. In low latitudes spatial patterns are dominated by ENSO, which has large spatial scales. Because of those large spatial patterns, less sampling is needed. The ENSO patterns may also be more stable than extratropical patterns, making the iterative adjustment of modes less critical at low latitudes. At higher latitudes spatial patterns tend to have smaller scales and need either more sampling or better representation in the reconstruction modes. Since those high-latitude regions tend to be poorly sampled in the late nineteenth century, the RSP analysis is not able to reliable resolve variations with the available sampling. With the improved modes in the Iter analysis, the available sampling can better resolve the high-latitude variations.

How well sampled and unsampled regions are represented is evaluated using the spatial correlations over either sampled or unsampled areas. As before, the correlations are compared to the fully sampled model-output SST. The annual RSP correlation changes greatly depending on whether a region is sampled. However, the annual Iter reconstruction correlation is more consistent over sampled and unsampled regions, indicating that the teleconnections are improved by the iterative adjustment of the EOFs (shown by the annual comparisons in Table 3).

### c. *Low-frequency results using correlated errors*

The comparisons shown above are for annual averages, which could be used as a first guess for a monthly analysis. The current version of ERSST uses a 15-yr low-frequency analysis for a first guess. To show low-frequency changes associated with the improved annual analysis, the ERSST low-frequency method is applied to the sparse grid CGCM data and compared to low-frequency estimates of both the full SST and the Iter reconstruction. For the ERSST low-frequency anomalies, annual averages need to be first smoothed and filled using large-scale spatial filters, and then the centered 15-yr median is taken to define the low-frequency anomaly. For the full data and reconstruction, the spatial smoothing is not needed and the low-frequency is defined by the 15-yr median.

Global averages of the full model low-frequency signal [LF(F); full data], LF(R), and LF(G) indicate that at the global scale there is little difference between the estimates for most years (Fig. 8). They show that the ERSST global low-frequency signal is reliable, as is reflected in ERSST global error estimates (Liu et al. 2015).

Although the low-frequency global means are consistent throughout most of the record, spatial correlations with LF(F) indicate that the spatial patterns are more variable (Fig. 9). For most of the presatellite period, LF(G) has a correlation of 0.6–0.8, while the LF(R) correlation is systematically higher, typically 0.8–0.9. The correlations are lowest in 1986, which is in the middle of the climate base period when the low-frequency anomalies are weak. The relatively low correlation indicates low spatial variance and not larger errors.

Maps of the low-frequency anomalies for 1896 (Fig. 10) show how the ERSST method for gapped data (middle panel) represents the large-scale anomaly

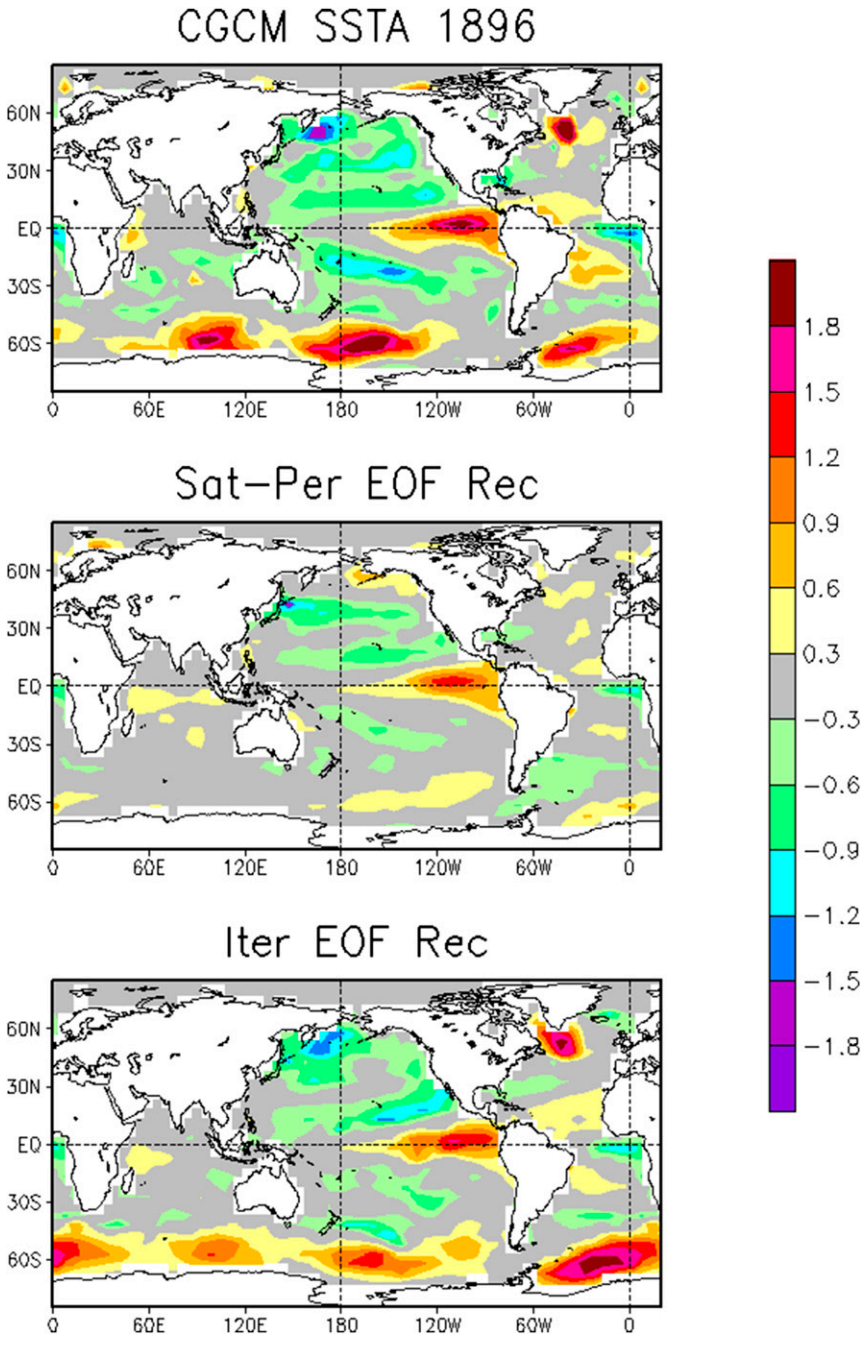


FIG. 6. Annual SST anomalies for model year 1896 from the (top) full model, (middle) RSP analysis, and (bottom) the Iter analysis.

reasonably well in tropical and northern midlatitudes, where sampling is reasonable. However, in the southern oceans and high latitudes of the Northern Hemisphere, the ERSST method strongly damps the low-frequency anomaly toward zero, where the full low frequency indicates strong anomalies (upper panel). The Iter reconstruction low-frequency analysis map (lower panel) closely resembles the full model-output low-frequency map for the year. Thus, the improvements it gives can reduce regional errors in the low-frequency component of an ERSST-type analysis that uses a low-frequency first guess.

The low-frequency spatial correlations are also evaluated for sampled and unsampled regions. Correlations for the low-frequency components are separated using the sampling of the center year of the 15-yr smoothing

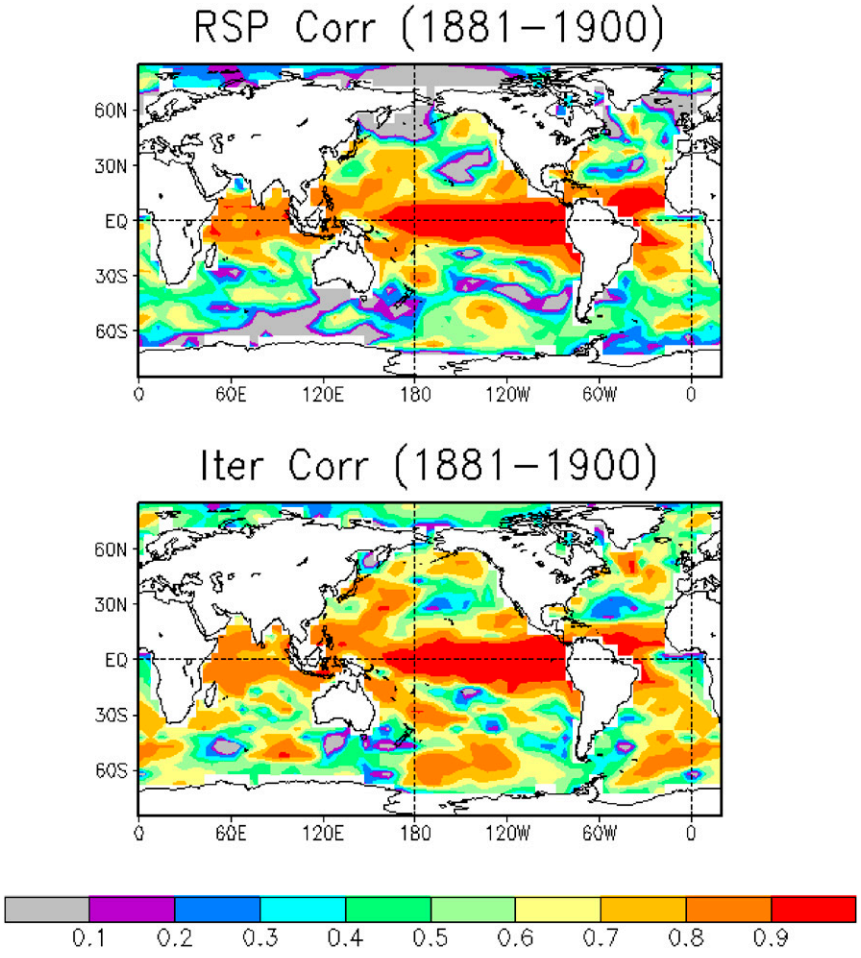


FIG. 7. Maps of temporal correlation against the annual model SST over 1881–1900 for the (top) RSP and (bottom) Iter analyses.

period. This estimate is used because for most periods sampling tends to change slowly. The LF(G) has relatively low but consistent correlation over both sampled and unsampled regions. However, the LF(R) correlation is actually higher for unsampled regions (Table 3, low-frequency comparisons), which may be due to high variance in high latitudes, where sampling is sparse, and from the ability of the Iter reconstruction to resolve those variations by better resolution of teleconnections. These comparisons indicate that much of the improvement in the Iter reconstruction occurs in unsampled regions.

## 5. Summary and conclusions

A method for improving historical annual average reconstruction is presented and tested using CGCM SST output and observed sampling grids. The method begins with an EOF analysis based on SST from the satellite period, since in practice an SST reconstruction needs spatially complete statistics typically obtained from satellite sampling. Historical SSTs are reconstructed using the initial modes, and then the historical anomaly increments at sampling locations are statistically analyzed and used to adjust the reconstruction. A new EOF analysis is then computed using the full period adjusted

TABLE 3. Time averages of spatial correlation with the full model annual average and low-frequency estimate, over 1861–2005. Correlations are over regions with historical sampling (Sample) and without sampling (No Sample). Reconstructions use EOFs from the satellite period, RSP, and iteratively-adjusted EOFs, Iter. Correlations with the model low-frequency include the ERSST-method low-frequency estimate, LF(G) and the low-frequency estimate of the Iter reconstruction, LF(R). For the LF correlations, sampling of the center year is used to define regions.

| | Sample | No Sample |
|---|---|---|
| Ann RSP | 0.64 | 0.51 |
| Ann Iter | 0.73 | 0.71 |
| LF(G) | 0.66 | 0.65 |
| LF(R) | 0.73 | 0.80 |

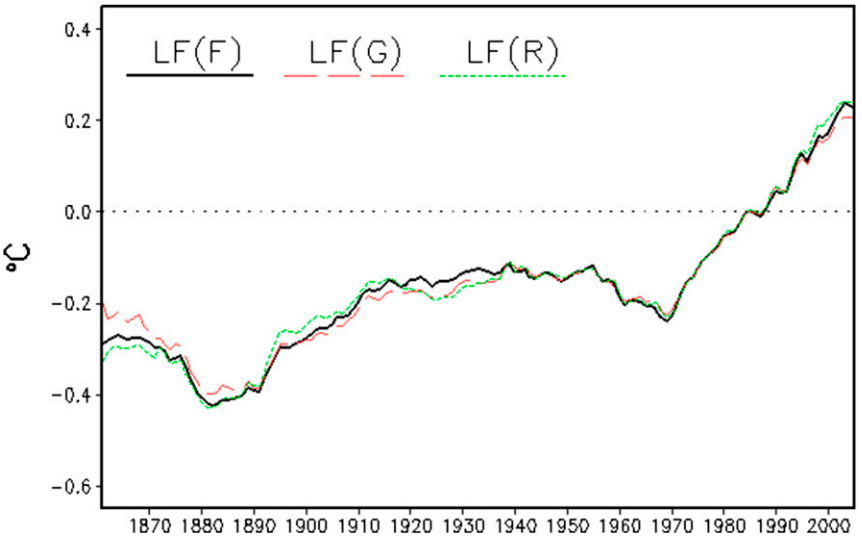


FIG. 8. Global averages of annual low-frequency anomalies from the full data, LF(F), the ERSST method applied to the gapped data, LF(G), and the Iter reconstruction, LF(R).

reconstruction and the new EOF modes are used to produce an updated reconstruction. The process is repeated until the reconstruction variance stabilizes.

This gives a way to combine fully sampled modern data with historical data to develop a set of spatially complete EOFs that better reflect variations over the historical period. Testing shows that the annual SST historical sampling is sufficient for this method. If historical data are too sparse, then the EOF modes cannot be reliably adjusted. How much sampling is needed depends on the spatial scales of the variations. Annual variations tend to have larger spatial scales than monthly variations, and annual sampling tends to be better than sampling for individual months. Those qualities help to make the method more useful for annual analysis.

The method was also tested for the analysis of monthly anomaly increments from the annual average and compared to an increment analysis using fixed satellite-period modes. The test showed that the method did not improve the monthly increment analysis. For a monthly SST analysis, the iterative method is most useful for giving a

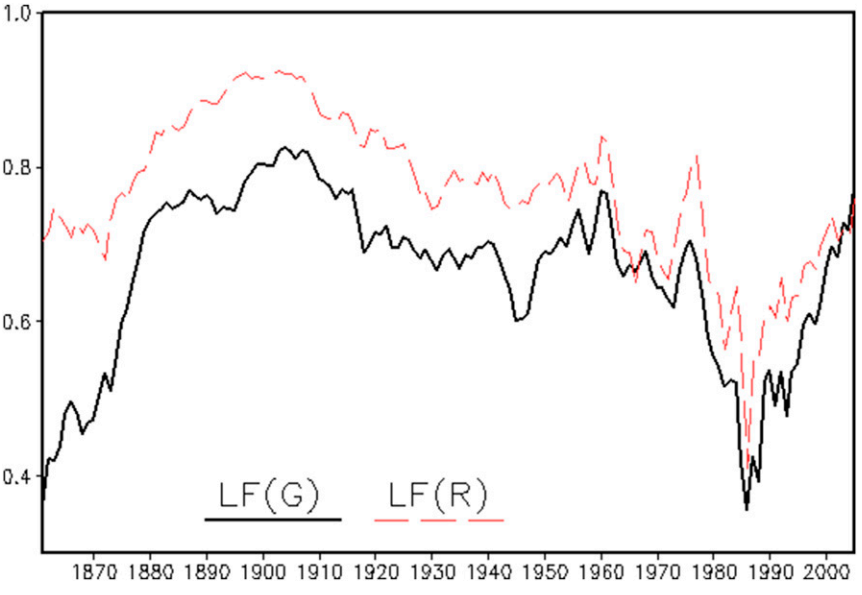


FIG. 9. Global spatial correlations of LF(F) with LF(G), and LF(R).

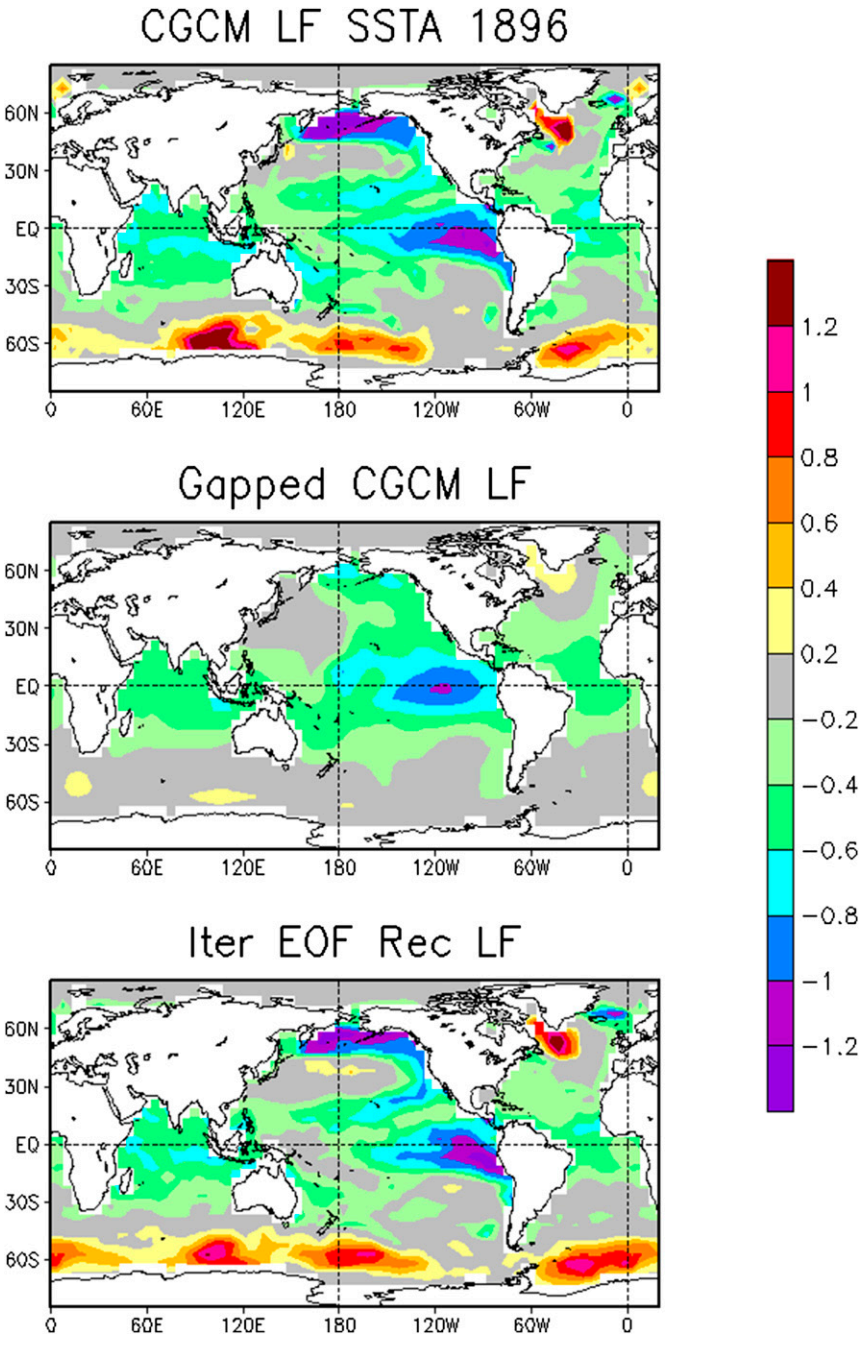


FIG. 10. Low-frequency anomalies for model year 1896 from (top) LF(F), (middle) LF(G), and (bottom) LF(R).

superior first-guess analysis that more fully accounts for multidecadal to interannual variations.

Although this study tests a global annual SST analysis, in principle the method could be applied on different space and time scales, and for different fields. However, the method would need to be tested and tuned for the different scales. The historical sampling needs to be dense enough to reliably adjust the leading reconstruction modes. If the leading modes represent only relatively small-scale features for a region, then historical sampling may not be dense enough to improve that region. The amount of historical variance that can be analyzed is dependent on the sampling relative to the spatial correlation scales for the variations of interest.

Results show that potential improvements in the SST reconstruction are largest in the extratropics and before

the mid-twentieth century. Improvements are reflected in better representation of large-scale multidecadal variations for that period. Variations that could be better represented by the improved method include the global change signal and extratropical climate modes, such as the PDO, the NAO, and Southern Hemisphere annular modes. An improved analysis could help to develop a better understanding of how those variations interact with and influence climate for long periods. In addition, the interactions between Southern Ocean variations and global warming may be more reliably studied with an improved analysis. An improved analysis can be useful for statistical studies of long-term climate variation and also to help validate models used for dynamic climate studies.

*Acknowledgments.* Thanks to B. Huang, J. Kennedy, and three anonymous reviewers for the many suggestions for improving earlier drafts of this paper. The contents of this paper are solely the opinions of the authors and do not constitute a statement of policy, decision, or position on behalf of NOAA or the U. S. government.

## APPENDIX

### The Iterative Reconstruction Method

The reconstruction method is based on the analyses of Smith et al. (1996, 1998). The method analyzes sparsely sampled historical anomalies by fitting them to a set of EOF spatial modes. Here covariance-based EOFs are used to analyze the simulated SST data. The EOFs for the first iteration are computed using spatially complete annual SST anomalies over the satellite period, 1982–2005. Error estimates are not added to the first iteration's simulated data, to simulate high-quality satellite-based data.

The historical annual anomalies with error estimates are fit to the set of modes to produce the set of weights that minimize the mean-square error (MSE) of the analysis as measured against the available simulated data at sampling locations. The reconstruction analysis is defined as

$$F(x,t) = \sum_{m=1}^{M} \psi_m(x) w_m(t). \tag{A1}$$

Here $\psi_m(x)$ is the EOF mode $m$ eigenvector at spatial location $x$, and $w_m(t)$ is the optimal mode weight for time $t$. The total number of modes used is $M$. The best-fit weights for the modes are computed separately for each time. The spatial MSE is defined at each time as

$$E^2 = \sum_{x=1}^{K} [D(x,t) - F(x,t)]^2 \delta(x) \cos\varphi(x). \tag{A2}$$

The data anomalies being fit to the modes are $D(x, t)$; $\delta(x)$ is a delta function equal to 1 if there is sampling at location $x$ and 0 otherwise; and $\varphi(x)$ is the latitude of location $x$, used as an area-weighting factor for the gridded data. The total number of spatial locations is $K$. The minimum MSE is found by solving

$$\frac{\partial E^2}{\partial w_m} = 0, \quad \text{for} \quad m = 1, 2, \ldots, M. \tag{A3}$$

By substituting Eqs. (A1) and (A2) into Eq. (A3), it can be shown that the weights are computed by solving the following system of equations:

$$\sum_{m=1}^{M} \left[ w_m \sum_{x=1}^{K} \psi_m(x) \psi_n(x) \delta(x) \cos\varphi(x) \right] = \sum_{x=1}^{K} D(x,t) \psi_n(x) \delta(x) \cos\varphi(x), \quad \text{for} \quad m = 1, 2, \ldots, M. \tag{A4}$$

Equations (A4) are solved numerically to yield the optimal set of weights for reconstruction.

Because of the sparse historical sampling, some modes may not always be adequately sampled. A badly sampled mode could inflate random errors in the historical data, so modes with poor sampling are screened out of the analysis. When modes are adequately sampled, the random errors are almost completely filtered out of the analysis, since random variations will not match the coherent variations represented by the modes. Sampling of each mode is described using the fraction of the mode variance sampled by the available data,

$$f_m = \frac{\sum_{x=1}^{K} \psi_m^2(x) \delta(x) \cos\varphi(x)}{\sum_{x=1}^{K} \psi_m^2(x) \cos\varphi(x)}. \tag{A5}$$

If this fraction falls below a critical fraction for the mode $m$, then that mode is excluded from the reconstruction. Here a critical fraction of 0.05 is used, which we have found to be sufficient to filter out almost all noise while minimizing analysis damping.

The first iteration uses EOFs, $\psi_{m,1}$, based on simulated data from 1982 to 2005 to simulate the satellite-period base data. To update the EOFs for the next iteration, the simulated historical data are statistically reinjected into the first iteration's reconstruction, $F_1(x, t)$. The data

reinjection is done for each year by forming increments between the simulated data with errors and the reconstruction, $I_1(x, t) = D(x, t) - F_1(x, t)$. An OI of those increments is done on the annual 5° grid using increments at sampling locations within 10° zonally and within 5° meridionally. The OI weights are damped so that if there are few data or large errors, the interpolated value will be damped toward zero, as in the OI analysis of Reynolds and Smith (1994). An updated annual analysis is then computed for 1861–2005, $T_1(x, t) = F_1(x, t) + \mathrm{OI}[I_1(x, t)]$.

For the second iteration a new set of EOFs is computed, $\psi_{m,2}$, based on $T_1(x, t)$ over 1861–2005. This new set of EOFs incorporates historical variations that have enough sampling to influence the leading modes of the analysis. Higher modes are not used because they may represent variations with small spatial or temporal scales, which could be caused by poor sampling or data errors. The updated set of EOFs is used to compute an updated reconstruction, $F_2(x, t)$.

The updated reconstruction is compared to the previous iteration using the average spatial variance over all areas and all times. If the average spatial variance change is less than 5%, the process stops; if not, then more iterations are computed until the change is less than 5%.

## REFERENCES


Huang, B., and Coauthors, 2015: Extended Reconstruction Sea Surface Temperature version 4 (ERSST.v4). Part I: Upgrades and intercomparisons. *J. Climate*, **28**, 911–930, doi:10.1175/JCLI-D-14-00006.1.

Hurrell, J. W., 1995: Decadal trends in the North Atlantic Oscillation: Regional temperatures and precipitation. *Science*, **269**, 676–679, doi:10.1126/science.269.5224.676.

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, doi:10.1029/97JC01736.

Kennedy, J. J., N. A. Rayner, R. O. Smith, D. E. Parker, and M. Saunby, 2011a: Reassessing biases and other uncertainties in sea surface temperature observations since 1850: 1. Measurement and sampling uncertainties. *J. Geophys. Res.*, **116**, D14103, doi:10.1029/2010JD015218.

——, ——, ——, ——, and ——, 2011b: Reassessing biases and other uncertainties in sea surface temperature observations measured in situ since 1850: 2. Biases and homogenization. *J. Geophys. Res.*, **116**, D14104, doi:10.1029/2010JD015220.

Kent, E. C., P. G. Challenor, and P. K. Taylor, 1999: A statistical determination of random observational errors present in voluntary observing ships meteorological reports. *J. Atmos. Oceanic Technol.*, **16**, 905–914, doi:10.1175/1520-0426(1999)016<0905:ASDOTR>2.0.CO;2.

Liu, W., and Coauthors, 2015: Extended Reconstructed Sea Surface Temperature version 4 (ERSST.v4): Part II. Parametric and structural uncertainty estimations. *J. Climate*, **28**, 931–951, doi:10.1175/JCLI-D-14-00007.1.

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, doi:10.1175/1520-0477(1997)078<1069:APICOW>2.0.CO;2.

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.

Reynolds, R. W., and T. M. Smith, 1994: Improved global sea surface temperature analysis using optimal interpolation. *J. Climate*, **7**, 929–948, doi:10.1175/1520-0442(1994)007<0929:IGSSTA>2.0.CO;2.

Smith, T. M., and R. W. Reynolds, 2003: Extended reconstruction of global sea surface temperatures based on COADS data (1854–1997). *J. Climate*, **16**, 1495–1510, doi:10.1175/1520-0442-16.10.1495.

——, ——, and C. F. Ropelewski, 1994: Optimal averaging of seasonal sea surface temperatures and associated confidence intervals (1860–1989). *J. Climate*, **7**, 949–964, doi:10.1175/1520-0442(1994)007<0949:OAOSSS>2.0.CO;2.

——, ——, R. E. Livezey, and D. C. Stokes, 1996: Reconstruction of historical sea surface temperatures using empirical orthogonal functions. *J. Climate*, **9**, 1403–1420, doi:10.1175/1520-0442(1996)009<1403:ROHSST>2.0.CO;2.

——, R. E. Livezey, and S. S. P. 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, doi:10.1175/1520-0442(1998)011<1717:AIMFAS>2.0.CO;2.

Taylor, K. E., R. J. Stouffer, and G. A. Meehl, 2012: An overview of CMIP5 and the experimental design. *Bull. Amer. Meteor. Soc.*, **93**, 485–498, doi:10.1175/BAMS-D-11-00094.1.

Thompson, D. W. J., and J. M. Wallace, 2000: Annular modes in the extratropical circulation. Part I: Month-to-month variability. *J. Climate*, **13**, 1000–1016, doi:10.1175/1520-0442(2000)013<1000:AMITEC>2.0.CO;2.