Reconstruction of Monthly Mean Oceanic Sea Level Pressure Based on COADS and Station Data (1854–1997)

THOMAS M. SMITH AND RICHARD W. REYNOLDS

National Climatic Data Center, Asheville, North Carolina

(Manuscript received 19 November 2002, in final form 30 November 2003)

ABSTRACT

An extended reconstruction of monthly mean oceanic historical sea level pressure (SLP) based on Comprehensive Ocean–Atmosphere Data Set (COADS) release-2 observations is produced for the 1854–1997 period. The COADS data are first screened using an adaptive quality-control procedure. Land SLP data from coastal and island stations are used to supplement the COADS data. The SLP anomalies are analyzed monthly to a 2° grid using statistics based on 20 yr of assimilated atmospheric reanalysis. A first-order correction is applied to the reconstruction to minimize variations associated with spurious long-term changes in the atmospheric mass over the oceans.

In the nineteenth century, the reconstruction appears to underestimate the SLP-anomaly amplitudes, and error estimates for the reconstruction are largest. After 1900 the reconstruction variance is stronger, although there are periods in the first half of the twentieth century when sampling is poor and the variance decreases. Spatial correlations between the reconstruction and several comparison analyses are highest in the second half of the twentieth century, suggesting greater reconstruction reliability after 1950.

1. Introduction

Analyses of sea level pressure (SLP) are useful for climate studies, and variations in surface temperature should in some way be reflected by SLP variations. Several analyses of SLP have been computed for climate studies. Trenberth and Paolino (1980) reconstructed both land and ocean SLP over the Northern Hemisphere beginning in 1899. Their study found many SLP data-quality problems that they needed to account for and correct as part of their reconstruction, especially over land. Problems they found include discontinuities between datasets and land-data problems associated with high-altitude stations. Since we are here concerned mostly with oceanic SLP, we avoid problems with land data. We also have the advantage of using an updated version of the Comprehensive Ocean–Atmosphere Data Set (COADS; Slutz et al. 1985) with the necessary gravity corrections applied.

The analysis of Basnett and Parker (1997) is global, including land, for the 1871–1994 period, and blends together several datasets. Their analysis includes complicated quality-control methods, as well as data corrections where needed. Since their analysis is global and covers most of our reconstruction period, we use it for comparison to our results.

A COADS-based marine SLP analysis was produced by Kaplan et al. (2000) for most of the same period as our analysis period. Although the input data they use are mostly the same as ours, we supplement the COADS data with some coastal and island land stations, which may provide better continuity near land boundaries and islands. Also, they employ methods that require all variations to be analyzed using modes computed from a recent subset of the data itself. Thus, modes of variations not spanned by that data subset are filtered out of their analysis, and regions not covered by that subset can never be analyzed.

The assimilated National Centers for Environmental Prediction–National Center for Atmospheric Research (NCEP–NCAR) Reanalysis of Kalnay et al. (1996) gives global SLP data that cover approximately the last 50 yr. However, satellite data for assimilation are not available before 1979. Before then, the reanalysis depends more heavily on the base atmospheric model into which the data are assimilated. Here we use only the satellite-period (1980–99) SLP from the reanalysis, both to develop statistics needed for reconstruction and to help evaluate our results. After the reanalysis was completed, it was discovered that some SLP data were incorrectly assimilated into the reanalysis. However, the SLP monthly means were not affected.

Our analysis is an attempt to produce an improved

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

statistical historical SLP analyses. One difference in our analysis is the use of updated COADS SLP data, merged with some coastal and island station SLP data. These data are screened using adaptive quality-control methods. Another difference is our analysis method, which is explicitly designed to analyze large-scale signals while filtering out data noise. Since our method uses full global statistics from the reanalysis, our analysis is also fully global. Comparisons and error estimates are used to help evaluate our analysis. Validation against the post-1980 reanalysis SLP, and against station data through the analysis period, indicates that the new reconstruction is generally more reliable than the comparison historical analyses.

2. Data

a. COADS data

The SLP data used for this analysis are derived from the latest version of the COADS release 2, with updates through 1997 (Slutz et al. 1985; Woodruff et al. 1998). In this version of COADS, gravity corrections have been applied to data sources that require them. The earth's apparent gravity changes slightly from the equator to the Pole, creating the need for corrections. These corrections adjust observed raw observations to the gravity at 45° latitude, and can amount to approximately −2.6 mb at the equator to 2.6 mb at the Poles (see Basnett and Parker 1997 for a discussion of gravity corrections). We first quality-control (QC) the SLP, as in Smith and Reynolds (2002). A QC step is needed to ensure that isolated bad observations do not contaminate the reconstruction. The QC method compares SLP anomalies to a local smooth analysis of SLP anomalies, which resolves seasonal-to-interannual variations. Thus, large anomalies may pass the QC if they are associated with sampled local variations on those time scales. This local, smoothed analysis is an optimum interpolation (OI) analysis. The local OI uses only local SLP anomalies but excludes anomalies more than five standard deviations from the mean in order to keep outliers from distorting the local analysis. Statistics used for the SLP QC method are computed here using the reanalysis data, as discussed below. Since the reanalysis is spatially complete for all times, it can yield spatially complete statistics without noise induced by data gaps. The reanalysis does not reflect all possible natural variations, so its use for QC statistics could cause screening of some observations reflecting real variations. However, because it is spatially complete and because most variations important to a monthly analysis should be represented in the reanalysis, we use it to compute these statistics. We discard data more than three standard deviations from the local OI analysis. Details of the QC method are given by Smith and Reynolds (2002).

As an example we show the frequency distribution of all COADS observations and those observations that pass the QC, for two representative decades in the Southern Hemisphere extratropics (Fig. 1). Here individual observations are considered. For both the 1890s and the 1970s, anomalies of ±5 mb are minimally affected, and the frequencies of larger anomalies are only slightly reduced. Note that in the 1970s there are some extreme negative outlier observations that are off the scale. Those outliers are all screened out by the QC. In other times and regions results are similar.

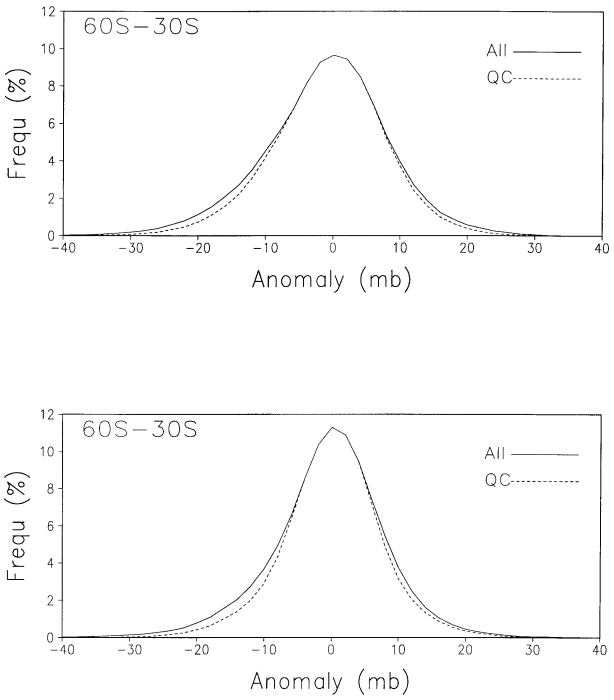

FIG. 1. Frequency distributions as a percent of the total number of observations within the region for 60°–30°S and the (upper) decade of the 1890s and the (lower) decade of the 1970s. The solid line is the distribution of raw COADS observations, the dashed line is for QC screened observations. The horizontal axes shows the SLP anomaly (in mb).

The global and annual number of individual observations and the number of observations that pass QC are shown in Fig. 2 (upper panel). The number of observations is low before the middle of the twentieth century, especially before 1900 and associated with the two world wars. For example, there are about 3 times as many observations per year in the last 30 yr compared to the 1930s. Anomalies that pass quality control are averaged together into superobservations, defined as monthly averages on a 2° grid, with grid centers on 0°, 2°E, . . . , 2°W by 88°S, 86°S, . . . , 88°N. The number of superobservations both without and with QC are shown in Fig. 2 (lower panel). The relative change in the number of superobservations between the early and late twentieth century is less than with individual observations, but there are still relatively low numbers before 1900 and during the world wars. In addition, in the post-1950 period a large number of superobserva-

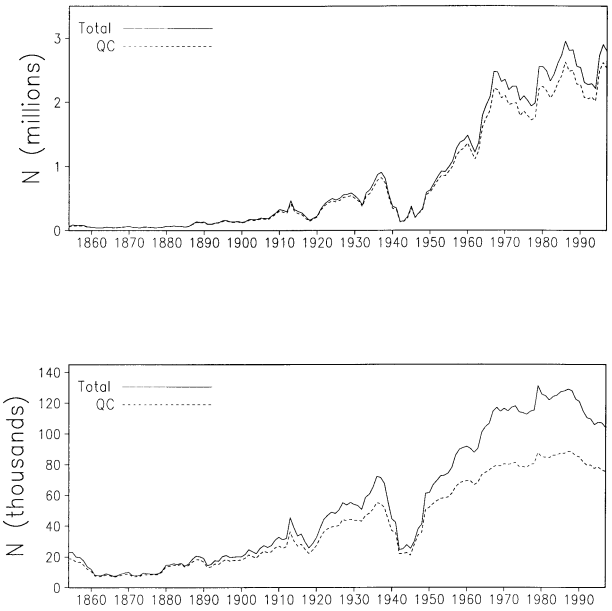

FIG. 2. (upper) The number of annual individual COADS SLP observations (in millions) and (lower) the number of annual 2° superobservations (in thousands). The solid line is the number of raw COADS observations and the dashed line is the number of QC screened observations.

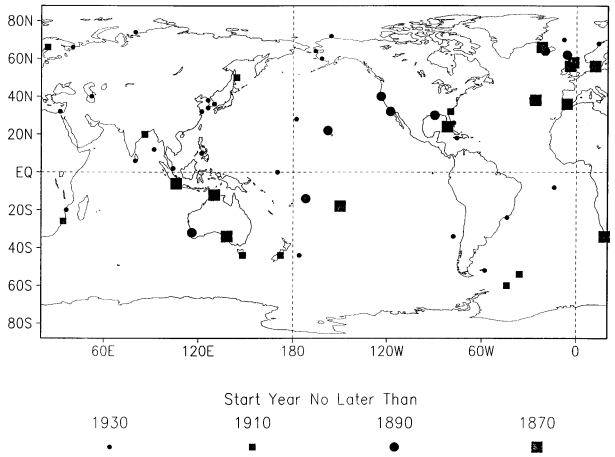

FIG. 3. Locations of the 58 supplemental coastal and island SLP stations. The size and shape of the mark shows the approximate length of each record, as indicated.

tions are lost due to quality control. As Fig. 1 and similar frequency diagrams indicate, the screening does not change greatly from one period to another. Some of the increased screening since the 1950s may be due to screening limits becoming tighter, due to an increasing number of observations in the local OI analysis used for QC. However, much of the increased screening in this period is due to more extreme outlier observations (absolute values of the anomalies exceeding 100 mb).

b. Supplemental station SLP data

Coastal and island SLP stations provide a number of fixed long time series, some extending over 100 yr or longer. These station data are less noisy than COADS data, as discussed below. Thus, the land-station data can help to anchor the reconstruction at islands and along coasts where they are available. The land-station SLP data we use are derived mostly from Vose et al. (1992), with updates. A few additional stations and extensions of the SLP data records were obtained using the data of Allan et al. (1991), Können et al. (1998), Hurrell (1995), and Jones et al. (1997).

To select the best stations for anchoring our reconstruction, we impose the following requirements. The station time series length must be at least 50 yr long and be at least 70% full on a monthly basis. Stations must begin no later than 1930. They must have at least 48 months of data within the 1980–99 period, since that is our base period. Geographically, stations must be within 2° of our ocean-analysis region and their reported altitude must be no more than 200 m.

A total of 58 stations pass these tests. These and the station data are recentered so that the average anomaly is zero over the 1980–99 period, and their locations are shifted to the center of the nearest 2° ocean grid square for use in the reconstruction. Locations of the station data (Fig. 3) indicate that these stations are most useful in the North Atlantic, tropical Pacific, and Indian Ocean regions. But there are also stations that can contribute to reconstruction in the Southern Hemisphere, the North Pacific, and the Arctic Ocean regions, especially in the period since 1930.

c. Reanalysis SLP data

Monthly reanalysis SLP data (Kalnay et al. 1996) from 1980–99 are used to compute reconstruction statistics. In this period the reanalysis incorporates satellite data, which influences the reanalysis SLP. Thus, the data for this period are more homogeneous with better coverage than over the longer reanalysis period.

The monthly SLP climatology is computed for the 1980–99 period using the reanalysis data. This climatology is subtracted from the reanalysis data to form anomalies. The reanalysis SLP anomaly standard deviation (Fig. 4) shows that strongest variations are in high latitudes, especially the North Atlantic, North Pacific, and the South Pacific. Although tropical variations are important to climate variations, their magnitude is much less than extratropical variations. Assuming that similar variations occur over the historical period, our SLP reconstruction should produce a similar standard deviation field.

The same reanalysis climatology used to form reanalysis anomalies is also used to form anomalies of COADS observations and station data. To compute the observation anomaly, the climatology is temporally in-

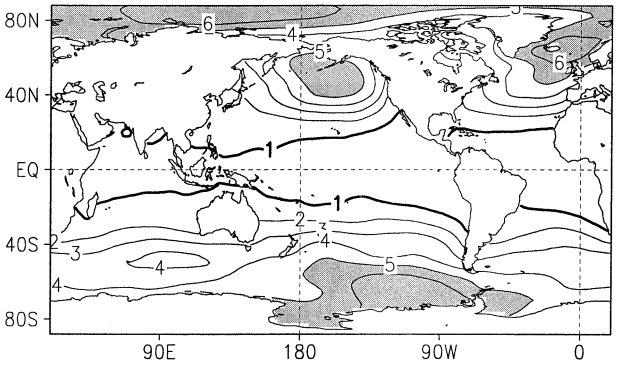

FIG. 4. Std dev (mb) of the 1980–99 monthly reanalysis SLP anomalies. Contour interval is 1 mb, the 1-mb contour is heavy, and values greater than 5 mb are shaded.

terpolated to the day of the observation. The station SLP anomalies are further adjusted by removing the mean 1980–99 anomaly, which is the SLP base period. We remove the base-period mean so that the stations are not biased with respect to that period.

d. Relative noise of COADS and station data

Some statistics of the SLP data are needed for merging the COADS data with the station data. The statistics are computed using the reanalysis data and the COADS and station observations, as discussed below. Jones et al. (2001) develop similar statistics for analysis of surface temperatures.

If the reanalysis SLP data are assumed to be the noise-free truth, then estimates of the noise in the station data and the COADS data can be estimated by comparison to the reanalysis data in their overlap period. Beginning with the station data, if we let P be the reanalysis SLP anomalies, assumed to be error free, and let P_s be the station SLP anomalies, then we can write

$$P_s = P + \varepsilon_s, \tag{1}$$

where ε_s is the error in the station data, which we will assume to be both random and unbiased. Since all error is assumed random and unbiased, we can use (1) to show that cov(P, P_s) = the variance of $P = \sigma^2$. Also with this error, the variance of P_s can be written as

$$\sigma_s^2 = \sigma^2 + \langle \varepsilon_s^2 \rangle,$$

where $\langle \varepsilon_s^2 \rangle$ is the error variance of the station data. Substituting these identities into the definition of correlation between P and P_s, which we call r_s, we can obtain an expression for the error variance in terms of the correlation and the variance,

$$\langle \varepsilon_s^2 \rangle = \sigma^2(1/r_s^2 - 1). \tag{2}$$

Thus, if the correlation is one, there is no error variance, and the error variance increases as the correlation decreases. Equation (2) assumes that the correlation is not zero. An estimate similar to (2) can be used to produce the COADS SLP error variance, $\langle \varepsilon_c^2 \rangle$, using the correlation with COADS SLP anomalies.

Since all errors are here assumed to be random, we can find the number of COADS observations equivalent to one station observation, N, using

$$\langle \varepsilon_s^2 \rangle = \langle \varepsilon_c^2 \rangle / N, \quad \text{or} \quad N = \langle \varepsilon_c^2 \rangle / \langle \varepsilon_s^2 \rangle. \tag{3}$$

From (2), this number can be computed using the correlations. Here the COADS 2° superobservations are used to compute typical $\langle \varepsilon_c^2 \rangle$ values. Using global values of this $\langle \varepsilon_c^2 \rangle$ estimate and all stations, Eq. (3) shows that COADS 2° superobservations are typically 9 times more noisy than station observations ($N \approx 9$).

3. Analysis methods

As discussed above, a number of SLP reconstructions have been computed by others (Trenberth and Paolino 1980; Kalnay et al. 1996; Basnett and Parker 1997; Kaplan et al 2000). Here, we discuss our reconstruction methods to be used with updated COADS data and land stations to produce a statistical reconstruction of the oceanic SLP. In designing our methods we make use of global spatial covariance information provided by the Kalnay et al. (1996) reanalysis of the satellite period to produce a globally complete analysis. In regions with insufficient data the analysis is damped toward zero anomaly (i.e., toward the 1980–99 reanalysis climatology).

a. The statistical analysis method

The statistical analysis method is similar to the method used by Smith and Reynolds (2002) for sea surface temperature (SST), with modifications as discussed below. We separate the low- and high-frequency SLP anomaly variations and analyze them separately. The low-frequency anomaly variations are analyzed by first forming 10° latitude by 10° longitude monthly averages from all monthly anomaly 2° superobservations with absolute values within one standard deviation of the mean. Here the mean and anomaly standard deviation are defined using the 1980–99 reanalysis SLP. We exclude values more than one standard deviation from the mean in this step since we are here only interested in analyzing the relatively small low-frequency variations, and we wish to avoid having sparse large anomalies bias the estimate. Those monthly 10° values are then averaged annually for all years with at least 6 months defined. The annual averages are used to estimate the low-frequency anomalies using a 15-yr median filter centered on each year, provided that at least 4 of the 15 yr are defined. A median filter gives the same result as a running-mean filter if data are normally distributed. However, median filtering removes outliers and, therefore, it is used here to better estimate the low-frequency anomalies. Parts of the filter extending beyond the end points are truncated. Binomial spatial filters are used to

fill in any remaining blanks and to smooth the field. These 10° low-frequency fields, $L(x)$, are interpolated to the 2° grid and subtracted from the 2° superobservation anomalies for every month in the appropriate year. The residual high-frequency anomalies are then analyzed.

The high-frequency analysis is computed by projecting the residual anomaly observations onto a set of anomaly SLP modes. In Smith and Reynolds (2002) anomaly increments were analyzed, and the 1-month autocorrelation was used to damp variations associated with unsampled modes. However, although SLP anomalies tend to have larger spatial scales than SST anomalies, they also tend to have shorter temporal scales (e.g., Davis 1976). In a test we found that the SLP modes have *e*-folding times of a month or less. Since we are producing a monthly analysis, we cannot make use of temporal correlations to damp the modes, as we did for SST.

The high-frequency anomalies are projected onto a set of empirical orthogonal teleconnections (EOTs; Van den Dool et al. 2000) modes, computed as in Smith and Reynolds (2002). To compute the EOTs, first the point that explains the most variance over the map is found, defining a base point. The spatial covariance pattern associated with that base point is computed. The EOTs used here are then localized, by damping the covariance pattern at distances of more than 5000 km from the base point. The pattern is damped by a factor ranging linearly from 1 at 5000 km to 0 at 8000 km. The variance associated with this localized covariance mode is removed from the data, and then the next EOT mode is found in the same way. We found that the first 50 EOT modes span nearly all of the base-period reanalysis SLP anomaly variance, so we use that set of modes for our analysis. We analyze the data using only those modes that are adequately sampled by the data, as shown by the fraction of the mode's variance sampled,

$$f(m) = \frac{\sum_{x \in A} \delta(x)\psi_m^2(x)a(x)F(x)}{\sum_{x \in A} \psi_m^2(x)a(x)F(x)}, \tag{4}$$

where $\psi_m(x)$ is the value of mode m at point x; $\delta(x) = 1$ if point x is sufficiently sampled (as defined below), and zero otherwise; $a(x)$ is the relative area represented by point x; and the summation is over the reconstruction region A. The variable $F(x)$ is a weighting factor, defined below. We define a mode as sufficiently sampled if $f(m) \geq 0.15$, which is the same criteria used by Smith and Reynolds (2002).

The high-frequency analysis is thus computed by

$$H(x) = \sum_{m=1}^{M} \psi_m(x)\Delta_m w_m, \tag{5}$$

where $\Delta_M = 1$ if mode m is sufficiently sampled and zero otherwise, and M is the total number of modes used for analysis. For each mode selected, the set of weights, w_m, where $m = 1, 2, \ldots, M$, minimizes the mean-square error of $H(x)$ compared to the data over the entire region (as in Smith et al. 1996).

We account for the greater accuracy of the coastal and island land SLP data, compared to the COADS super observations, in both the weighting of superobservations and in the analysis. To form combined superobservations at locations with a land station and a COADS superobservation, the combined superobservation value is set equal to $P = (P_c + NP_s)/(1 + N)$, where P_c is the COADS SLP superobservation monthly anomaly, P_s is the station monthly anomaly, and N is defined in Eq. (3). We use $N = 9$ globally based on our estimates. If there is no COADS superobservation, $P = P_s$. These merged COADS and land-station SLP data are then analyzed.

In order to compute the best-fit weights for Eq. (5), we solve the equation

$$\sum_{n=1}^{K} w_n \sum_{x \in A} \psi_n(x)\psi_m(x)\delta(x)a(x)F(x) = \sum_{x \in A} P_h\psi_m(x)\delta(x)F(x), \tag{6}$$

where $m = 1, 2, \ldots, K$ are the modes found to be adequately sampled ($K \leq M$), and P_h is the high-frequency merged COADS and land-station SLP anomalies. Here and in Eq. (4),

$$F(x) = \begin{cases} 1 & \text{COADS only} \\ N & \text{land station only} \\ 1 + N & \text{both} \end{cases}$$

is a weight that accounts for the greater accuracy of the land-station data. Thus, land-station observations are weighted N times more than COADS superobservations. The total reconstruction for a particular time is thus

$$R(x) = H(x) + L(x), \tag{7}$$

where $L(x)$ is the low-frequency analysis, discussed above.

b. Removal of artificial trends

As noted above, gravity corrections were applied to some of the SLP data. Careful research was used to identify data requiring these corrections, but it is possible that future work will show the need for a slightly different distribution of corrections. There may also be other unknown instrument or recording problems that could induce an artificial trend in the data. In addition, the early decades have more sparse and variable sampling compared to the most recent decades. Sampling of SLP may also change over time as ships evolved from sailing to steam- and motor-powered vessels, and as better forecasts allowed ships to avoid storms. Such changes in sampling can contribute to artificial trends in the SLP data. Real trends in SLP can occur due to

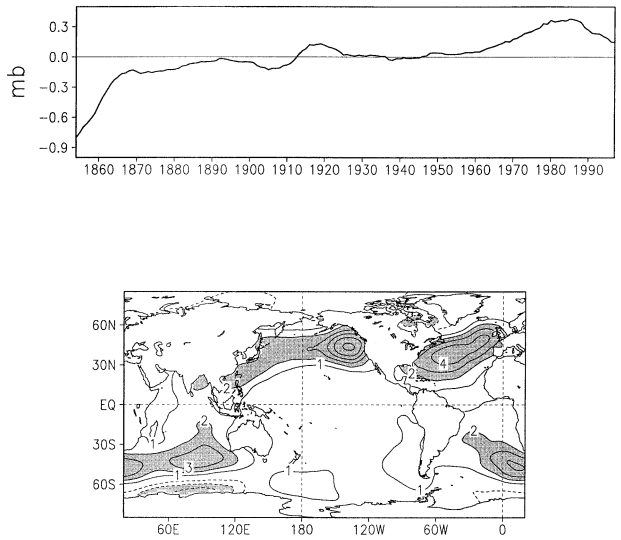

FIG. 5. Correction for the reconstructed SLP anomalies. (upper) The 11-yr and global-average SLP anomaly (in mb) and the (lower) regression of that time series on the uncorrected annual-average reconstructed SLP anomalies (in nondimensional units).

climate variations, but since the atmospheric mass is approximately constant, any change in one place must be balanced elsewhere. To minimize artificial trends we apply a first-order correction to the analyzed data so that the long-term oceanic SLP approximately maintains this balance.

Since we wish to produce a first-order correction for trends, we base the correction on averages of reconstructed SLP anomalies. We first compute the global and 11-yr running average of the SLP anomalies. If there are no artificial trends, then this time series should be constant because the global average SLP over the oceans should be approximately constant. This assumes that there is no long-term net atmospheric mass exchange from the continents to the oceans. The 11-yr average is used to remove high-frequency noise. Any changes in this time series are assumed to be artificial. This time series is regressed against the annual-average SLP anomalies at each point on the map to find the spatial pattern associated with the artificial variations (Fig. 5). The spatial pattern is computed similar to an EOT mode, except that here the time series is specified as the artificial global trend.

Much of the artificial change occurs before 1870, when sampling is extremely sparse. However, a gradual change continues over the analysis period. The regions where the artificial changes are largest are high-variance regions, including the midlatitudes of the Northern and Southern Hemispheres, where there is a positive trend over the analysis period. Over the Tropics, this regression indicates that there is little variation. This pattern suggests that improvement of the midlatitude storm avoidance by ships measuring SLP may by responsible for the trend. This artificial-variation mode is almost identical to the leading empirical orthogonal function of the low-frequency filtered reconstruction, indicating the need to remove it before searching for physical modes of variation. Note that physical changes in atmospheric vapor pressure can produce systematic changes in SLP of perhaps as much as 1 or 2 mb, but the trend shown here is often larger than that and it is associated with typical ship tracks, so we assume that it is spurious.

We remove this estimate of the artificial variance from the reconstruction. If $\tau(t)$ is the spatial and 11-yr average for a given time t, and $\gamma(x)$ is the regression value at a spatial location x (as shown by Fig. 5), then the corrected reconstruction is produced by adjusting Eq. (7),

$$R_c(x, t) = H(x, t) + L(x, t) - \tau(t)\gamma(x). \quad (8)$$

c. Error estimates

It is useful to have an analysis error estimate to help indicate when and where the analysis is reliable. The mean-square error (mse) of the analysis is by definition

$$E^2 = \langle (P - P_a)^2 \rangle,$$
$$= \langle P^2 \rangle + \langle P_a^2 \rangle - 2\langle P P_a \rangle, \quad (9)$$

where P is the actual SLP anomaly, P_a is the analysis estimate, and the brackets denote averaging. We may define the analysis as a linear function of the actual value plus some error,

$$P_a = \alpha P + \beta + \varepsilon,$$

where α and β are constants and ε is an uncorrelated residual with zero mean. The constant α may be less than one if the analysis is damped, which occurs when data are sparse. Using this functional definition, the variance of P_a is

$$\sigma_a^2 = \langle P_a^2 \rangle - \langle P_a \rangle^2 = \alpha^2 \sigma^2 + \langle \varepsilon^2 \rangle,$$

where $\sigma^2 = \langle P^2 \rangle - \langle P \rangle^2$ is the variance of P, and the covariance between them is

$$\mathrm{cov}(P, P_a) = \langle P P_a \rangle - \langle P \rangle \langle P_a \rangle = \alpha^2 \sigma^2.$$

Substituting these definitions into (9) and factoring of terms allows us to rewrite the error as

$$E^2 = \sigma^2 (1 - \alpha)^2 + \langle \varepsilon^2 \rangle + (\langle P \rangle - \langle P_a \rangle)^2. \quad (10)$$

On the right-hand side of Eq. (10), the first term, $\sigma^2(1 - \alpha)^2$, is the error caused by the analysis being too weak or too strong. This amplitude error is largest in our analysis when the sampling is sparse and the variance is damped. The term $\langle \varepsilon^2 \rangle$ is error caused by uncorrelated noise in the analysis. Because ε is uncorrelated the mean of its cross products with other terms is zero. Since we project data onto physical modes, screening out all poorly sampled modes, this term should be small in our analysis. Therefore, we disregard it. The final term on the right-hand side is error caused

by bias in the analysis. As noted above we constrain our analysis to minimize the global-average trend, which should greatly reduce the bias error. Therefore, we also disregard this term. It is unlikely that either of these terms are exactly equal to zero in our analysis, but they should generally be less than the amplitude error.

With $\langle \varepsilon^2 \rangle = 0$ the analysis variance is $\sigma_a^2 = \alpha^2 \sigma^2$. Using that and the assumption of zero bias error, we may estimate our analysis mse as

$$E_a^2 = \sigma^2(1 - \alpha)^2 = (\sigma - \sigma_a)^2. \qquad (11)$$

In general both σ and σ_a are time dependent. Here we assume that the actual standard deviation σ is approximately stationary and can be estimated using the 1980–99 reanalysis data (Fig. 4). We estimate the analysis standard deviation σ_a using 15 yr of data centered on the time of interest, which allows us to estimate the sampling error using Eq. (11).

This may underestimate error because we disregard noise and bias error. However, if the variance is not stationary, then this estimate may also overestimate the error. The error estimation using Eq. (11) should be used to give the user guidance about when and where the analysis is more or less reliable. Error estimates are further discussed in section 4.

The normalized standard error, $E_a/\sigma = 1 - \sigma_a/\sigma$, indicates where sampling for the analysis is good enough to produce a strong signal relative to the background state. A normalized error close to one indicates a heavily damped analysis due to insufficient sampling. Small normalized errors indicate good sampling and a strong analysis.

4. Results

Results of the corrected reconstruction, from Eq. (8), are discussed and compared to the Global Mean SLP Data Set 2, version-f analysis of Basnett and Parker (1997, hereafter BP97), and also compared to the Kaplan et al. (2000, hereafter KEA) analysis. Anomalies of both are interpolated to our 2° grid. We use these analyses for comparisons because they are global or near-global and they overlap most of our analysis period. Performing analyses on the 5° BP97 grid gives similar results, since our analysis does not resolve small-scale variations. The 2° grid is used for convenience.

Spatial statistics (Murphy and Epstein 1989) help to illustrate some overall characteristics of the analyses. Here we discuss annual averages of monthly spatial statistics. The spatial standard deviations, over the area 60°S–60°N for the reconstruction (Recon), BP97, the analysis of KEA, and Kalnay et al. (1996, hereafter Rean) indicate the spatial variability of each analysis over time (Fig. 6a). All of the datasets are most similar after 1950. Before 1950 the Recon strength is more variable than BP97, with larger standard deviation values around 1910 and the 1920s and 1930s and lower

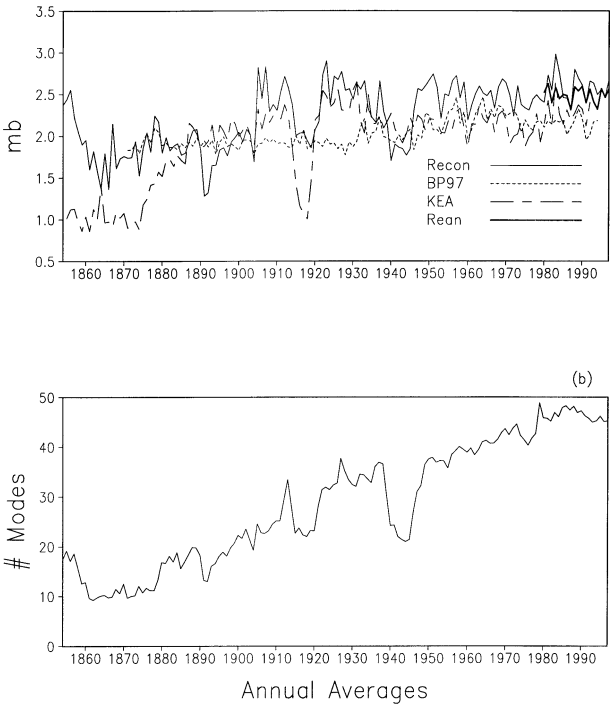

FIG. 6. (upper) Spatial std dev (60°S–60°N) of the reconstruction (Recon), BP97, KEA, and the Kalnay et al. (1996) reanalysis (Rean). (lower) The number of modes used in the reconstruction. For both the annual averages of monthly values are shown.

values in the late 1910s and the early 1940s. These times of lower reconstruction standard deviation values are also times when the number of modes used for reconstruction is relatively low (Fig. 6b), indicating that the decreased variance is due to damping when sampling is poor.

Compared to the Recon, the COADS-based KEA analysis generally has weaker spatial variability, and it shows greater reductions in the 1910s and before 1880. This suggests that Rean-based modes used for our reconstruction may better span SLPA variance than the COADS-based modes used by KEA. In addition, the BP97 spatial standard deviation values are more constant than with the Recon. After 1950, temporal changes in the BP97 standard deviation are similar to changes in the Recon and KEA. Before 1950, when sampling is more sparse, the BP97 standard deviation has little change over time. This suggests that BP97 in this sparse-sampling period is more damped toward a base state.

Regional spatial standard deviations (Fig. 7) show that dips in the Recon standard deviations in the 1910s and 1940s are due to large reductions in the Southern Hemisphere extratropics. The KEA analysis shows a similar reduction in the 1910s in the Southern Hemisphere as well as a reduction in the Tropics for that period. However, KEA has less Southern Hemisphere reduction than the Recon in the 1940s. The lesser KEA reduction in the 1940s may be due to the fact that the KEA analysis does not cover much of the Southern

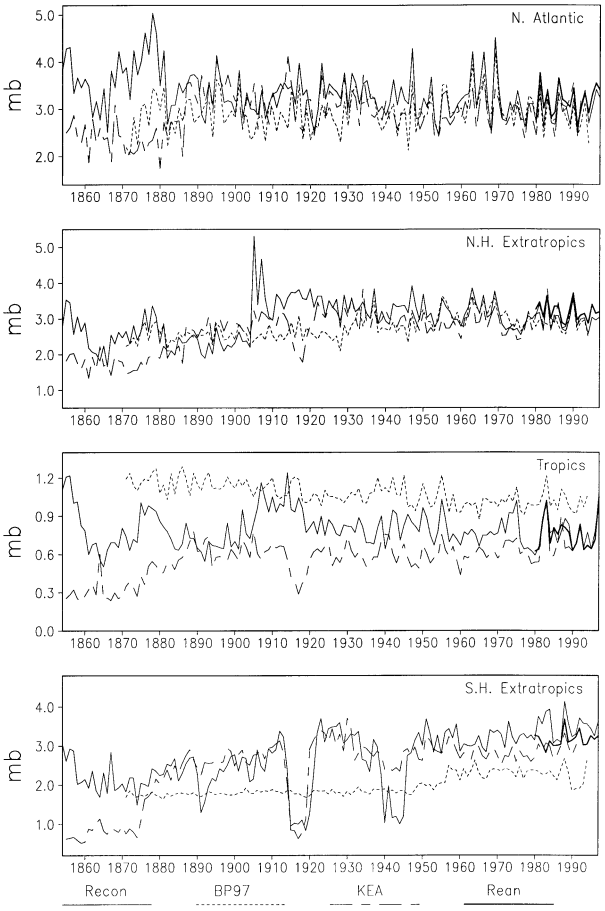

FIG. 7. Spatial std dev, as in Fig. 6, for the North Atlantic (23°–60°N, 70°W–0°), the Northern Hemisphere extratropics (23°–60°N), the Tropics (23°S–23°N), and the Southern Hemisphere extratropics (60°–23°S).

Hemisphere south of 40°S or the southeast South Pacific, while we produce an analysis for the entire region, which is damped when data are sparse.

Over most of the analysis period the sampling is good for the North Atlantic. Before 1880 in the North Atlantic, the Recon standard deviation is less stable, indicating that even in the North Atlantic that may be as far back as a useful analysis may be computed. In the same period KEA and BP97 standard deviations are damped. For the entire Northern Hemisphere extratropics, the Recon has an unrealistic jump in spatial standard deviation shortly before 1910, due to insufficient North Pacific sampling before 1910. The North Pacific error estimates for the 1900–09 period (not shown) indicate larger errors over most of the basin compared to later decades. For all regions the BP97 standard deviation is most consistent and tends to be higher than the others in the Tropics and lower in the extratropics. Sampling appears to be sufficient after 1910 for all regions except parts of the Southern Hemisphere, where sampling is insufficient in the 1910s and the early 1940s.

Spatial correlations between the Recon and the other

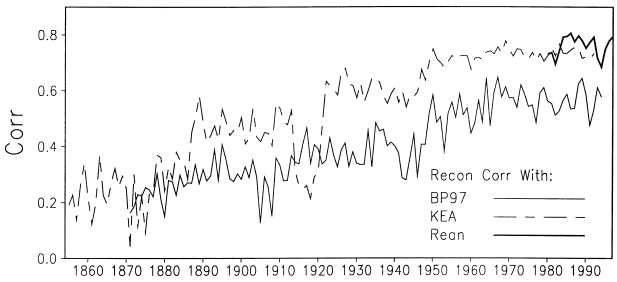

FIG. 8. Spatial correlation (60°S–60°N) between the reconstruction and the comparison analyses: BP97, KEA, and the Kalnay et al. (1996) reanalysis (Rean). Shown are annual averages of monthly values.

analyses indicate how closely the anomaly patterns match (Fig. 8). Correlations with the Rean over our base period are high, indicating a close matching of anomaly patterns for that overlap period. This high correlation is helped by the fact that the Recon is based on Rean modes. However, the KEA analysis uses COADS-based modes, and it also has a high correlation after 1950. The BP97 analysis correlates strongest after 1950 when the sampling is best, although its correlation is lower than with the COADS-based KEA analysis. Although the input data used are mostly the same, the BP97 analysis methods involves less smoothing of the input data and, therefore, it can contain more data noise. That additional noise may reduce the correlations. For both historical analyses, correlations with the Recon decrease before 1950 and are low before 1900.

As shown in Fig. 8, the Recon and Rean correlate well in their overlap period. Further validation of the Recon and comparison with the other analyses is given in Table 1. Validation is shown for the Rean for the 1980–92 period when all of the analyses are defined. Although the Rean itself may contain errors, it is a high-quality analysis and these comparisons help to show the relative quality of the different analyses. Comparisons are done using spatial statistics over the 60°S–60°N region, and averaging those statistics over the 1980–92 period. Here the data were all averaged to the same 5° grid to ensure that the 5° BP97 analysis is not unfairly disadvantaged.

The statistics considered in Table 1 are the correlation (r), the rms difference (rmsd), and the ratio of the analysis variance to the Rean variance (R_{var}). The variance ratio gives an indication of how damped the analysis is

TABLE 1. Spatial statistics over the 60°S–60°N region computed against the reanalysis, averaged over 1980–92. Statistics are for the reconstruction (Recon), BP97 analysis, and KEA analysis. Shown are the correlation r, rms difference, rmsd, and the ratio of the analysis variance to the reanalysis variance R_{var}. (The rmsd units: mb.)

	r	Rmsd	R_{var}
Recon	0.77	1.62	1.07
BP97	0.62	1.94	0.75
KEA	0.73	1.56	0.88

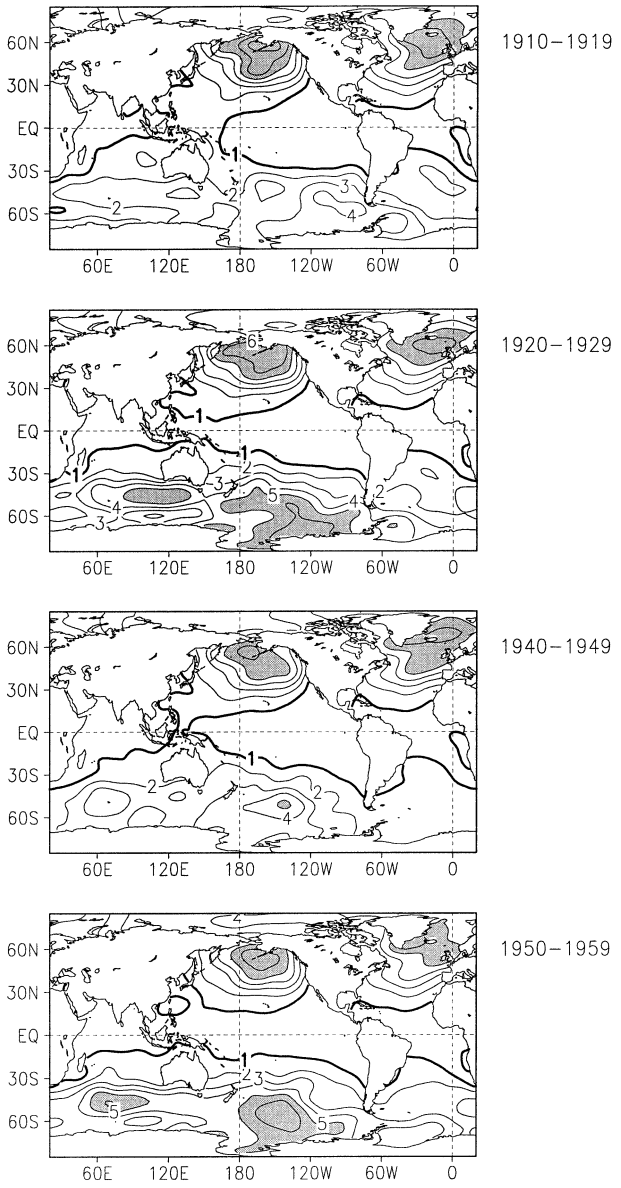

FIG. 9. Temporal std dev (mb) of the reconstruction for 1910–19, 1920–29, 1940–49, and 1950–59. Contour interval is 1 mb, the 1-mb contour is heavy, and values greater than 5 mb are shaded.

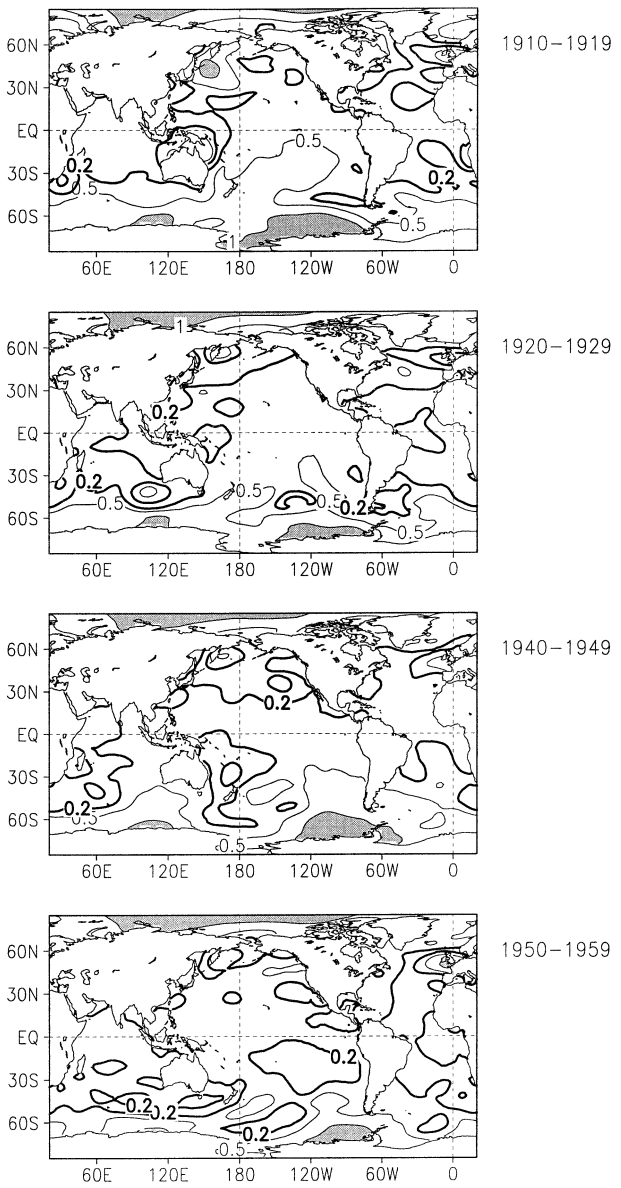

FIG. 10. Normalized error estimates for annual averages of the reconstruction, averaged for the decades 1910–19, 1920–29, 1940–49, and 1950–59. The contour interval is 0.5, and the 0.2 contour is also added as a heavy line. Values greater than one are shaded.

[reflected by α^2 in Eq. (10)]. The Recon has the highest correlation and the R_{var} closest to one, and it has a low rmsd. The KEA analysis has slightly lower rmsd and second best correlation and R_{var}. The BP97 analysis does not compare as well as the others in this period. The Recon and KEA give similar comparisons, with the Recon slightly better over this period.

Temporal standard deviations of the monthly reconstruction anomalies for the 1910s, 1920s, 1940s, and 1950s (Fig. 9) help to show why Southern Hemisphere spatial standard deviations vary in the 1910s and the 1940s. In those two decades, the Southern Hemisphere high-latitude standard deviation is damped because of sparse sampling. For the region north of 30°S, the standard deviation does not change much between these periods. In the 1950s the standard deviation is similar to the Rean standard deviation (Fig. 4), although there are differences.

The normalized standard error also helps to show regions and times when the analysis is more reliable. Standard error [from the square root of Eq. (11)] is normalized with the Rean standard deviation. Here we show the error of annual averages, averaged for decades to indicate typical normalized errors for those decades (Fig. 10). In regions where the normalized error is close

TABLE 2. Averages of rmsd from station data (in mb). Analyses compared are the reconstruction (Recon), BP97, KEA, and the reanalysis (Rean, for 1980–97 only).

	1870–99	1900–29	1930–59	1960–97
Recon	1.93	1.62	1.51	1.33
BP97	1.64	1.71	1.82	1.55
KEA	3.55	2.48	2.03	1.59
Rean	—	—	—	0.62

to one, the error is as large as the signal, and the analysis is of little use. Such regions occur at high latitudes, especially in the Southern Hemisphere. In the 1910s and the 1940s, the error in these regions increases slightly. In the 1910s most of the South Pacific has normalized error above 0.5, while for the 1940s error is largest in the southeast South Pacific. The North Pacific shows some higher normalized error values for the 1910s, but they are lower in later decades.

From 1980 to 1997 when the Recon overlaps the Rean, the global and period average normalized error estimate is 0.25. For the same period, the average normalized unbiased rmsd of the Recon, compared to the Rean, is 0.58. Although some of this larger unbiased rmsd is due to errors in the Rean, the comparison suggest that the error estimates for the Recon are numerically too low. However, we believe that the variation of errors in time and space may be used for qualitative guidance about the changing reliability of the Recon.

In most of the Atlantic and Indian basins the normalized error is low over these decades. An exception is in the North Atlantic near Britain, where the normalized error tends to be larger. In that well-sampled region, the unanalyzed COADS values are consistent with our reconstruction, and both show larger variance in the 1950s than in the 1980–97 period. As shown in Fig. 9, this is a region with a strong gradient in the standard deviation, so slight shifts in the mean atmospheric weather patterns can cause large changes in the local SLP variance. Since our assumption of stationary variance is violated in that region, the local error is overestimated.

To better quantify errors in the Recon and the comparison analyses, the rmsd between station data and the analyses is computed using monthly data for several approximately 30-yr periods. Stations selected are the 12 stations with the longest records shown in Fig. 3. These station data are incorporated in the Recon and the other analyses, except for KEA, which uses only COADS data. So for most analyses they are not independent data. However, they help to validate the various analyses.

In Table 2 the rmsd is shown for the analyses for the periods 1870–99, 1900–29, 1930–59, and 1960–97. The Rean is only compared for the 1980–97 period. For each, the rmsd is computed against stations using the ocean grid values closest to the station locations. The average of the 12 values is shown. In the late nineteenth century, BP97 has the lowest rmsd. As indicated in Fig. 6, that is a time when sampling is sparse and the analyses are heavily weighted toward climatology. For the three later periods the Recon has the lowest rmsd, except that the Rean rmsd is lowest in the period. When it is available, the Rean assimilates many different types of data from in situ observations and satellites, giving it an advantage over the historical analyses. Aside from the Rean, the BP97 analysis has the second lowest rmsd for the most recent three periods. The KEA analysis, which is independent of the station data, has the highest rmsd in all periods. These comparisons indicate that the statistical analysis techniques used for the Recon improve the analysis, and inclusion of the station data substantially improves the analysis. Since the Recon analyzes SLP using large-scale modes, information from these stations can influence large regions around the stations.

5. Summary

We present a new global SLP analysis based on COADS data. The input COADS data are first screened using adaptive quality control to remove outliers. Anomalies of the screened data are formed using the 1980–99 reanalysis climatology of Kalnay et al. (1996). Those anomalies are averaged monthly and to 2° squares before statistical analysis.

The statistical analysis separately analyzes low- and high-frequency variations. Low-frequency variations are analyzed by large-scale and temporal averaging. High-frequency variations are analyzed by fitting the COADS anomalies to a screened set of covariance modes. The global set of modes is computed from the 1980–99 Kalnay et al. (1996) data. Modes are screened to exclude those not adequately sampled by the data for a given time. The analysis is computed monthly from 1854 to 1997.

Comparisons to other analyses and an error estimate of our analysis help to show the analysis reliability. Because of sparse data, the analysis is not reliable before 1910 except in the North Atlantic, where it is reliable back to about 1880. In the Southern Hemisphere extratropics it is unreliable in the 1910s and the 1940s. After 1950, when sampling is most dense, the analyses compared here are most similar. Comparisons to the Kalnay et al. (1996) SLP for 1980–97, and to station SLP over the historical period, indicate that the new reconstruction is generally more reliable than the comparison analyses.

The sparse sampling in the nineteenth and early twentieth-century limits the value of SLP reconstructions in those periods. However, there are some historical data that have not yet been incorporated. As these new data become available COADS is augmented, and future additions may allow improved future analyses. For example, 1850s to 1860s data from ship's log books are currently being digitized for COADS. Light-ship data from near the U.S. coast are also being digitized. Ad-

ditional data may particularly help future SLP analyses of the nineteenth century, when data are most sparse.

Acknowledgments. We thank S. Woodruff and S. Worley for help with the COADS SLP data and comments on the reconstruction. D. Wuertz made the land-station SLP data available. The reanalysis SLP data were obtained from the JISAO Web site maintained by T. Mitchell. The Kaplan et al. data were obtained from the LDEO data Web site. We also thank W. Esbisuzaki for information and advice about the reanalysis SLP. Useful comments were also provided by D. Easterling, A. Kaplan, J. Lawrimore, and two anonymous reviewers. Funding to support this work was provided through the Office of Global Programs.

REFERENCES

Allan, R. J., N. Nicholls, P. D. Jones, and I. J. Butterworth, 1991: A further extension of the Tahiti–Darwin SOI, early SOI results and Darwin pressure. *J. Climate,* **4,** 743–749.

Basnett, T. A., and D. E. Parker, 1997: Development of the global mean sea level pressure data set GMSLP2. Hadley Centre Rep. CRTN 79, 16 pp.

Davis, R. E., 1976: Predictability of sea surface temperature and sea level pressure anomalies over the North Pacific Ocean. *J. Phys. Oceanogr.,* **6,** 249–266.

Hurrell, J. W., 1995: Decadal trends in the North Atlantic Oscillation and relationships to regional temperature and precipitation. *Science,* **269,** 676–679.

Jones, P. D., T. Jónsson, and D. Wheeler, 1997: Extension to the North Atlantic Oscillation using early instrumental pressure observations from Gibraltar and south-west Iceland. *Int. J. Climatol.,* **17,** 1433–1450.

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

Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteor. Soc.,* **77,** 437–471.

Kaplan, A., Y. Kushnir, and M. A. Cane, 2000: Reduced space optimal interpolation of historical marine sea level pressure: 1854–1992. *J. Climate,* **13,** 2987–3002.

Können, G. P., P. D. Jones, M. H. Kaltofen, and R. J. Allan, 1998: Pre-1866 extensions of the Southern Oscillation index using early Indonesian and Tahitian meteorological readings. *J. Climate,* **11,** 2325–2339.

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

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

Smith, T. M., and R. W. Reynolds, 2002: Extended reconstruction of global sea surface temperatures based on COADS data (1854–1997). *J. Climate,* **16,** 1495–1510.

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

Trenberth, K. E., and D. A. Paolino, 1980: The Northern Hemisphere sea-level pressure data set: Trends, errors and discontinuities. *Mon. Wea. Rev.,* **108,** 855–872.

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

Vose, R. S., R. L. Schmoyer, P. M. Steurer, T. C. Peterson, R. Heim, T. R. Karl, and J. Eischeid, 1992: The Global Historical Climatology Network: Long-term monthly temperature, precipitation, sea level pressure, and station pressure data. Rep. ORNL/CDIAC-53, NDP-041, 325 pp. [Available from Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, Oak Ridge, TN 37831.]

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