An Improved Method for Analyzing Sparse and Irregularly Distributed SST Data on a Regular Grid: The Tropical Pacific Ocean

THOMAS M. SMITH AND ROBERT E. LIVEZEY

NOAA/NWS/NCEP/CPC, Camp Springs, Maryland

SAMUEL S. SHEN

Department of Mathematical Sciences, University of Alberta, Edmonton, Alberta, Canada

(Manuscript received 21 January 1997, in final form 3 September 1997)

ABSTRACT

An improved method for interpolating sparsely sampled climatological data onto a regular grid is shown. The method uses the spatial and temporal covariance of the field, along with the sparse data, to fill the full grid. This improves on similar methods that have recently been developed by eliminating the development of features that are not sufficiently supported by the data (i.e., overfitting). Statistical tests are used to tune the method to represent as much variability as the spatial–temporal information will support without overfitting. The method is further improved by a data-checking procedure that detects and removes suspect data. The method is developed and evaluated by interpolating tropical Pacific sea surface temperature (SST) monthly anomalies to a regular grid for the 1856–1995 period. Ship data averaged to 5° squares are used as input and are interpolated to a complete 1° grid. Comparing the results to interpolations using other methods shows this method's quantitative improvements where satellite data are available for validation. Comparisons in the presatellite era show sharper and stronger anomaly patterns with this method, compared to another method developed for use with sparse data. Also shown are several periods when data are so sparse that only very weak SST anomalies may be reliably reconstructed in the tropical Pacific (i.e., before 1870 and 1915–25). In future research, the global SST and possibly other climatological fields will be gridded using improved methods.

1. Introduction

Interpolation of historical climatological data onto a regular grid is often necessary in order to use those data for climate studies, and various interpolation techniques have been developed. When data are sufficiently dense, many techniques give almost identical results. Problems arise when data become sparse and irregularly distributed. This problem has been addressed using the data's large-scale covariance structure as defined by a set of empirical orthogonal functions (EOFs) (Rayner et al. 1995; Shriver and O'Brien 1995; Smith et al. 1996; Kaplan et al. 1997a, Kaplan et al. 1997b, manuscripts submitted to *J. Geophys. Res.*, hereafter K97a, K97b). Here we begin with the method of Smith et al. (1996, hereafter SRLS), which was developed to grid sea surface temperature (SST) anomalies for the period beginning 1950. The post-1950 period was chosen because there is moderate to good sampling in that period. With poor sampling, that approach can lead to overfitting errors (i.e., generating large anomalies not supported by the sparse observations). We show how to build onto the method of SRLS to allow SST interpolation with very sparse data. The new method prevents overfitting with sparse data while slightly better representing variations supported by the available data.

The procedure is developed and tested using monthly SST anomalies in the tropical Pacific Ocean (20°S–20°N 155°E–105°W), from January 1856 to October 1995. Here, a limited region is used to develop the method and show its improvements over previously used methods. The selected area is one of the most important for studies of interannual variability, but also one of the most difficult to produce gridded analyses for prior to the satellite era. In future research the study area will be enlarged. Data are supplied by the U.K. Meteorological Office (UKMO) and are an updated version of the atlas of Bottomley et al. (1990). They consist of averages of individual observations on 5° longitude–latitude squares (5° superobservations). The UKMO data are used in this study because they have already been carefully bias corrected for the pre-1942 period, when biased bucket temperatures are most common (Parker et al. 1994; Folland and Parker 1995). Also, in

Corresponding author address: Dr. Thomas M. Smith, Analysis Branch, Climate Prediction Center, National Centers for Environmental Prediction, World Weather Building, 5200 Auth Road, Room 605, W/NP52, Camp Springs, MD 20746.
E-mail: wd52ts@sgi25.wwb.noaa.gov

the extremely data-sparse period before 1950, there is little advantage gained by going from 5° superobservations to higher-resolution superobservations, such as the 2° Comprehensive Ocean–Atmosphere Data Set (COADS) data (Slutz et al. 1985). With either size superobservations, they will be sparsely distributed. For these reasons, we decided to use the corrected 5° superobservations from the UKMO. We form anomalies (departures from norms) of these data using the climatology of Reynolds and Smith (1995), and assign each SST value to the center of its 5° square.

Spatial covariance is computed using the 1° resolution monthly optimal interpolation (OI) of SST (Reynolds and Smith 1994) for the 1982–95 period. The OI is an accurate analysis of that period because of the use of bias-corrected satellite data as well as in situ observations. This relatively brief and recent period may be used because, as discussed below, we separately account for interdecadial variations, and we also use the covariance to interpolate month-to-month SST increments rather than the full anomaly. The full anomaly is then computed by adding the interpolated increment to the previous month's analysis, with the interdecadal component added on later. Results show that this approach is justified. Interpolation of the full anomaly all in one step requires a longer basis period to fully describe the anomaly, which requires the use of lower quality basis data. It also requires more dependence on the statistical structure obtained from that lower quality data, because anomalies are more complicated than monthly increments.

2. The method

A simple way to interpolate SST data that are reasonably dense, such as after 1950 in the tropical Pacific, is to use the method of SRLS,

$$T(n) = \mu(n) + \sum_{m=1}^{M} w_m \psi_m(n), \qquad (1)$$

where $T(n)$ is the SST anomaly analysis at location n, $\mu(n)$ is the low-frequency anomaly, ψ_m the anomaly EOF (for M modes), and w_m is the weight for mode m. Weights are computed for the set of modes to minimize the mean-squared error (MSE) of the interpolated field compared to the available observations, as discussed in detail by SRLS. However, the method of SRLS cannot be used with extremely sparse data, such as is often encountered with SST before 1950. The method described below was developed to overcome the limitations of the SRLS method without taking away from its strengths. There are several parts to the new interpolation method, each described later as follows. Section 2a: Before interpolation, the low-frequency (decadal or longer) SST anomaly is removed from the data, to be added back on later. Removal of the low-frequency anomaly is referred to as recentering the anomaly. Section 2b: Data increments from the previous month's recentered analysis (i.e., from the first guess, G) are interpolated using the increment spatial covariance. This 1-month incremental correction is then added onto G to form the recentered anomaly for the month. Temporal information is included through the first guess. Rotated covariance EOFs of monthly increments are used to define the modes of spatial covariability. Section 2c: Screening regression is used to ensure that only those modes of spatial variability that are supported by the available data are used, to avoid overfitting. Section 2d: Modes of spatial variability that are not supported by the available data are damped to avoid unsupported anomaly persistence. Section 2e: A system for checking data increments is developed to find and discard suspect superobservations. Section 2f: These components are used to obtain the anomaly analysis for each month. In section 3 the statistical tuning of the method is discussed, which is needed to ensure the highest possible accuracy.

The 5° superobservations are interpolated to a 1° regular grid by using 1° resolution EOFs. Although some small-scale variations must be lost by the use of 5° superobservations, the use of 1° EOFs makes it possible to reconstruct some sharp SST anomaly gradients associated with several leading EOFs. This approach assumes that the 5° superobservations can be assigned to the center of the 5° box. Results show that this assumption does not cause large errors.

a. Removal of low-frequency anomalies

Separate interpolation of the very low frequency signal reduces the amount of variance that the statistical interpolation must account for. Since decadal or longer timescale SST anomalies describe large-scale variations that can be estimated more directly from the available SST observations (e.g., Parker et al. 1994), they can be computed separately and removed from the data. With those anomalies accounted for separately, the limited-period but high quality OI SST data may be used to more accurately estimate subdecadal spatial covariance.

As was shown by SRLS, separate interpolation of low-frequency anomalies improves the interpolation accuracy. The improvements in skill should be even greater when interpolating over the much longer period of data used here, because the low-frequency signal is substantial over the >100-yr period. The low-frequency anomaly is therefore computed and removed from the 5° superobservations before computation of statistics and gridding. After interpolation, the low-frequency anomaly is added back on.

Low-frequency anomalies are computed from a running mean of monthly UKMO anomalies optimally averaged within 10° longitude–latitude squares. The optimal averaging (OA) method used here is similar to the method described by Smith et al. (1994) and is described in the appendix. Time series of the OA used here are

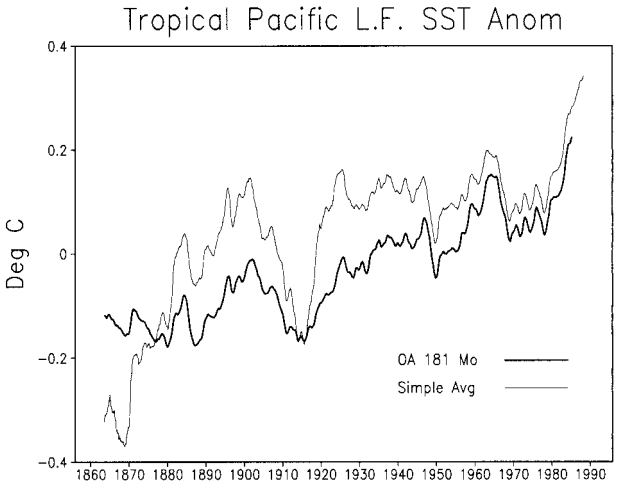

FIG. 1. Area-average low-frequency anomalies in the tropical Pacific (20°S–20°N, 155°E–105°W) from UKMO data, a 181-month average period (°C), using the 10° OA superobservations and the raw data anomalies (simple averaging).

similar to those shown by Smith et al. (1994) for the same regions. A running mean with a length of about 14 yr is appropriate because that is the length of the OI-basis period. We use a 15-yr (181 month) running mean as a compromise that filters out interannual variations and is long enough to span some extremely sparse data periods (e.g., 1913–23).

The OA-based 181-month running mean, averaged over the entire tropical Pacific area (Fig. 1), has a generally upward trend over the period, similar to the running mean from simple averaging (i.e., area-weighted averaging) of the available data. However, the OA-based mean exhibits less artificial low-frequency variability from sampling across extreme sparse-sampling periods before 1880 and around 1915. For the first 90 months of the record, the first available low-frequency anomaly is used, and for the last 90 months of the record, the last available low-frequency anomaly is used. For the remainder of this study, it should be understood that the 5° superobservation anomalies are recentered by removal of these low-frequency anomalies before statistical interpolation, and that the low-frequency anomalies are added back onto the interpolation afterwards.

b. Data increments

Data increments are computed for gridding to further reduce the work that the EOF-based gridding scheme must perform, by using information from the previous month's analysis to help form the recentered anomaly. Because SST anomalies generally have decay times of 3–6 months (e.g., see SRLS), their 1-month increments require less information to interpolate than the full anomalies do. As discussed by Thiébaux (1997), this technique can provide greater accuracy in practical situations when a statistical description of the increment is available. That is the case with SST, due to the availability of the OI analysis. Data increments (or monthly tendencies) are differences between the (recentered) 5° superobservation anomaly, D, and a first guess, G,

$$I(n) = D(n) - G(n). \quad (2)$$

The increments, I, are defined at all data locations. Because the monthly SST anomalies have long timescales, a slightly damped previous month's analysis is used for G. Damping of G is proportional to the 1-month lag autocorrelation of anomalies and is discussed in detail later. By using a first guess and interpolated increments, temporal information is used in addition to the spatial information, which is important when data are sparse.

By running the analysis backward in time, information from the month after the analysis month is included. Since the autocorrelation is equally valid in both directions, a backward run is as valid as a forward run. We perform the analysis in both forward and backward directions, and average the results to include time information in both directions. The forward analysis uses an initial first guess of zero anomaly, and the backward analysis uses an initial first guess of the last forward analysis.

To grid the data increments, a set of increment EOFs are computed and varimax rotated. The increment EOFs are computed from month-to-month SST anomaly differences. Data used are the OI analysis, using all months in the 1982–95 period. Increments are smaller than full anomalies, so computation of increment covariance structure requires accurate data such as the OI to prevent noise from excessively contaminating the EOFs.

Rotated modes are generally more localized and more closely resemble observed structures than unrotated modes (Richman 1986) and, therefore, are preferred when not all modes are always used. If all EOF modes in a set were to be used always, then rotation would be unnecessary. However, with sparse data not all modes will always be sufficiently supported by the available data, and therefore we will generally choose a subset of the total for interpolation. Weights for the set of EOFs supported by the data are computed using least squares best-fit criteria. The computation of the weights is the same as in SRLS except that only those supported modes are used, while SRLS used all modes.

c. Screening regression

Screening regression is necessary in order to avoid overfitting (i.e., using modes that are not supported by the available data). Screening is a tool for choosing a best set of predictors given a set of predictands, and for interpolation it amounts to choosing the rotated EOF modes that are supported by the monthly data. Using modes that are not supported by the data would lead to overfitting, which occurs when there are no data near the strongest loadings of a mode and the weight for the mode is entirely determined by data in weak-loading

regions. Since the weight reflects the least squared best fit only where data exist, a little noise in the data can cause the unsupported mode to have a large weight in order to better fit the noise in the weak-loading region. However, the large weight would cause an unsupported local maxima or minima at the locations where the mode's loadings are strongest (where there are no data). Screening regression eliminates overfitting problems by not using unsupported EOF modes in the fit.

To determine how well a mode is supported, we compute for each mode m and for each set of data the variable

$$f_v(m) = \frac{\sum_{n=1}^{N} \delta_n \psi_m^2(n) \cos\phi_n}{\sum_{n=1}^{N} \psi_m^2(n) \cos\phi_n}. \quad (3)$$

Here, the summations are over all N spatial points, $\cos\phi_n$ is the cosine of latitude ϕ_n, and δ_n is 1 if there is an observation at point n, and 0 otherwise. The variable $f_v(m)$ gives the fraction of the total variance of mode m that is directly sampled. The minimum value of $f_v(m)$ for acceptance of the mode (f_v^*) is found by tuning, described below.

d. Damping of unsupported modes

As noted in section 2b, the first guess, G, is the previous month's recentered anomaly slightly damped. There is no need to damp that part of G that is supported by the data at the analysis month, because that part of the guess will be directly corrected by the data. In order to damp only those parts of G that are not supported by the data, the previous month's (recentered) anomaly analysis is projected onto the full set of increment EOFs. Since the recentered anomalies are built up by a series of increment fits, with an assumption of zero anomaly first guess for the first analysis, projection of the anomalies onto the increment EOFs is justified and completely represents G. The previous month's anomaly analysis is defined at all points so there is no problem with projecting it onto the full set of EOFs.

Thus, the first guess is

$$G(n) = \sum_{m=1}^{M} [\Delta_m + (1 - \Delta_m)c_m] wg_m \psi_m(n), \quad (4)$$

where wg_m is the weight for rotated increment EOF mode ψ_m, and c_m is the 1-month lag autocorrelation of the SST anomaly projected onto the increment EOF mode ψ_m. Here, the variable $\Delta_m = 1$ if mode ψ_m is supported by the data and 0 otherwise, as determined from the data distribution using (3). Note that the weights wg_m are for the least squares best fit of the full recentered anomaly from the previous month's analysis, but the EOFs ψ_m are the same increment EOFs used to grid the increments.

e. Data checking

Ship observations of SST tend to be extremely noisy (Reynolds and Smith 1994). Since errors in the ship observations will be represented in the interpolated analysis, it is highly desirable to remove as many questionable 5° superobservations as possible. However, with sparse data, care must be taken to avoid discarding any superobservations that may represent real variations.

As an example of the sort of errors possible in the superobservations, consider March 1985. In that month the 5° superobservations of ship data show anomalies between +1°C and +4°C in the eastern equatorial Pacific. The size of the anomalies are physically realistic, but the high quality OI analysis shows negative anomalies in the same region. Buoy data also show negative anomalies in the region. The ship-based 5° superobservation anomalies are positive only for a few months, and then rapidly switch to negative anomalies consistent with the OI and the buoy data. Clearly, this is a case of erroneous ship reports (either instrument or ship-position errors). The problem is that for most of the ship-data record there is no higher quality analysis with which to compare the ship-based superobservations.

One way to detect physically unrealistic reports is to examine their month-to-month variability. In regions where the reported SST anomaly changes more rapidly than is physically reasonable, it can be assumed that there is an error. The undamped anomaly increments are therefore checked for all data, and data that are found to change too rapidly are not used in the statistical interpolation.

The OI anomaly 1-month increment standard deviation (σ_I) is computed at all spatial points. The range of σ_I in the tropical Pacific is approximately 0.25°–0.75°C, with largest values in the eastern equatorial Pacific where changes in upwelling can force rapid SST variations. Absolute increments from the undamped previous month's analysis, $T_{t-1}(n)$, are therefore computed,

$$I_A(n) = |D_t(n) - T_{t-1}(n)|, \quad (5)$$

and compared to $A\sigma_I$. Here A is an amplitude to be preset (e.g., 3 or 4). If $I_A(n) > A\sigma_I(n)$, then $D_t(n)$ is rejected. Tuning, described below, is used to find the optimal amplitude A.

Although the 5° superobservations have undergone rough quality control in their development (Bottomley et al. 1990), the example cited above shows the need for further data checking. The data-checking method outlined here removes the most suspect superobservations through a spatial–temporal consistency check. Since SST anomalies normally do not change rapidly from month to month, this approach is justified. However, only consistency with analyses based on other superobservations can be checked, and therefore situations involving systematic errors over a region or in two or more months are very difficult to find.

f. The complete interpolation method

After the data are checked for outliers, screening regression is applied to find which modes are supported by the remaining data (to define Δ) and then the first-guess weights (wg) and the damped first guess (G) are computed. The increments from G are computed and the weights for the increment gridding (w) are determined for the supported modes. Adding the damped first guess to the interpolated increment yields the full recentered interpolated SST anomaly analysis,

$$Tr_t(n) = \sum_{m=1}^{M} \psi_m(n)[\Delta_m(w_m + wg_m) + (1 - \Delta_m)c_m wg_m]. \quad (6)$$

Adding the low-frequency anomaly back onto (6) gives the full interpolated anomaly,

$$T_t(n) = \mu(n) + Tr_t(n). \quad (7)$$

Note that when all rotated EOFs are supported by the data, $\Delta_m \equiv 1$ and (7) is equivalent to (1), the method of SRLS, but with a slightly different basis. For the next month's analysis the process is repeated, with the guess G_{t+1} computed using weights for the projection of Tr_t onto the full set of modes ψ_m.

As discussed in section 2b, to include temporal information in both directions, we run (7) both forward and backward, and then average the results. The difference between the forward and backward runs gives an indication of the stability of the method in extremely data-sparse periods, as discussed later. Stability can develop for two reasons: either enough data to sample from both directions or no data for an extended period. Thus, stability alone cannot be used as a measure of accuracy.

3. Tuning

Tuning of the interpolation method is done to estimate the parameters needed. Estimates are found to optimize the skill of the fit throughout the period of record. These parameters include the number of rotated anomaly increment EOF modes to choose from (M), the cutoff criterion for choosing modes (f_v^*), the 1-month lag correlations for each mode (c_m), and the amplitude for rejection of suspect data (A). Similar experiments also helped in the development of the method by showing an advantage in gridding anomaly increments and a first guess compared to directly gridding anomalies, and also an advantage in removing low-frequency variance anomalies before gridding and then adding it back in afterward.

Cross-validation tests over a part of the OI period are used for tuning. Briefly, the cross-validation tests are conducted as follows. The 1985–89 60-month period is chosen for analysis, because it contains both warm and cool episodes in the tropical Pacific and because it is long enough to yield stable cross-validation statistics. For each month in that period, a separate set of rotated increment EOFs is computed. These cross-validation EOFs are computed using all data over the OI period except for a 17-month window centered on the month. That window is removed to produce EOFs that are approximately independent of the month. Note that after tuning, interpolation for the full record uses just one set of EOFs using all available OI data. Data used for interpolation in the cross-validation period are the UKMO superobservations for the period, except that they are subsampled to simulate historical sampling conditions. If there is a superobservation at a location in the historical month as well as in the cross-validation month, then the superobservation is used for interpolation. Results from the test are evaluated by computing mean square error (mse) with respect to the OI data, for the given historical sampling conditions and the given choice of parameters. By varying the parameters in these tests, their optimal values can be determined for the given historical sampling conditions. The historical grids tested are for the periods 1885–89, 1905–09, 1915–19, 1935–39, and 1955–59. More details of cross validation are given by SRLS.

Testing for the number of rotated anomaly increment EOF modes was done setting $A = 4.0$ and $f_v^* = 0.005$, and computing c_m from the UKMO data. As shown below, these parameter settings are reasonable estimates. Truncations at 8 and 14 modes were selected for testing using the method of O'Lenic and Livezey (1988). There was almost identical skill for all periods for either 8 or 14 modes. Since additional EOFs do not add skill, we chose eight to avoid possibly introducing noise from the higher modes. It is reassuring that relatively few EOFs are needed, since the higher-order modes are more likely to contain variations that may not have been well sampled in the past. The f_v^* value is relatively small because we are using 1° resolution EOFs with 5° resolution data (assumed to fill the 1° cell at the center of the 5° area with other 1° cells in the 5° area set to missing), so even a completely filled 5° grid would fill only 4% of the 1° grid.

The optimum f_v^* was found by testing a range of values (0.01–0.001) with eight rotated EOFs. For every period, the minimum cross-validation mse occurs with f_v^* at or close to 0.005. We performed this test using two sets of c_m, one computed by projection of OI anomalies onto the EOFs and the other by projection of the 5° ship anomalies onto the EOFs. Since the ship data are more noisy than the OI data, their c_m's are slightly lower than those from the OI anomalies. Results were slightly better when the c_m's were computed from ship anomalies. Thus we fix f_v^* at 0.005 and use c_m's from projection of ship anomalies. Table 1 contains the cross-validation mse for various f_v^* using this choice for c_m's. Also shown is the average number of 5° superobservations for each period within our test region.

Tests to determine the optimal data cutoff amplitude, A, were done by fixing all other parameters as determined above and varying A between 3.0 and 4.5. Results

TABLE 1. Cross-validation experiment results using (6) with data checked with $A = 4.0$, and eight rotated increment EOFs and c_m's from the UKMO data. The mean-square errors (°C²) for different cross-validation grids for choices of the f_v* are tabulated. The constant zero-anomaly fit has mse = 0.49°C². Also tabulated are the average number of monthly superobservations in each period (maximum number = 160).

1885–89	1905–09	1915–19	1935–39	1955–59	f_v*
0.23	0.33	0.47	0.30	0.15	0.010
0.20	0.29	0.45	0.24	0.15	0.007
0.22	0.24	0.39	0.23	0.15	0.005
0.23	0.25	0.38	0.29	0.15	0.003
0.38	0.37	0.48	0.33	0.15	0.001
37	38	7	59	124	N (avg)

are best (i.e., the cross-validation mse is lowest) with $A = 4.0$. Table 2 shows the total number of observations per decade and the number of observations discarded with $A = 4.0$. The amount discarded is always less than 10% of the total, and often 5% or less. This is many more than would be expected if there were no errors and increments were normally distributed, but is not surprising considering that ship observations tend to be noisy (e.g., see Reynolds and Smith 1994). All results discussed below use these parameter choices.

To summarize, tuning experiments show that the optimal parameter settings are $M = 8$, $f_v^* = 0.005$, $A = 4.0$, and c_m computed from the ship data.

4. Results

Results are discussed in two sections: the dependent period (which is within the period from which the EOFs were computed, 1982–95) and the full period. In addition, comparisons to the SRLS method are made. The SRLS method uses (1) to reconstruct using the same UKMO data in the same region, and using the same base period to compute the EOFs. That method uses the same number of EOF modes that SRLS found was optimal for tropical Pacific reconstruction after 1950. As noted by SRLS, this method is not reliable before 1950 when data are too sparse.

a. The dependent period

The root-mean square error (rmse) with respect to the OI shows that even in the relatively data-dense dependent period, the new method generally outperforms the SRLS method (Fig. 2). The improvement is largest in the eastern equatorial Pacific. In the western equatorial Pacific near 168°E, the dependent period's data are noisy, leading to the increased error there. The analysis methods both smooth out much of the noise in the raw data, but data errors always contaminate the analysis. The data-checking procedure in the new method significantly reduces errors from excessively noisy data. Besides having a lower rmse, a larger portion of the signal strength is also captured by the new method (Fig. 3). Incrementally adjusting a first guess, rather than gridding a full (recentered) anomaly is one reason for the new method's improvement. Another is the use of screening regression, which does not allow unrepresentative modes to take away from the fitted signal. The high quality OI standard deviation is the best available estimate of the true SST anomaly standard deviation over the period, which is near 1.6°C at its maximum.

TABLE 2. Total number of UKMO 5° superobservations per decade and the number and percent discarded each decade using $A = 4.0$ (maximum total number = 19 200). For the 1860s the months are Jan 1860–Dec 1869, etc.

Decade	Total	Discarded	% discarded
1860s	481	22	5
1870s	2252	133	6
1880s	4230	141	3
1890s	3892	107	3
1900s	5253	159	3
1910s	2201	87	4
1920s	5183	371	7
1930s	7861	467	6
1940s	8385	644	8
1950s	13 879	890	6
1960s	18 301	814	4
1970s	18 487	1062	6
1980s	18 397	1233	7

Maps of the SST anomalies from the new analysis, the SRLS method, and the OI from several months within the dependent period help to illustrate the performance of the EOF-based methods. For January 1983, near the peak of a very strong warm episode, all three have similar patterns (Fig. 4). A more difficult test is March 1985 (Fig. 5), when the sampling is more sparse

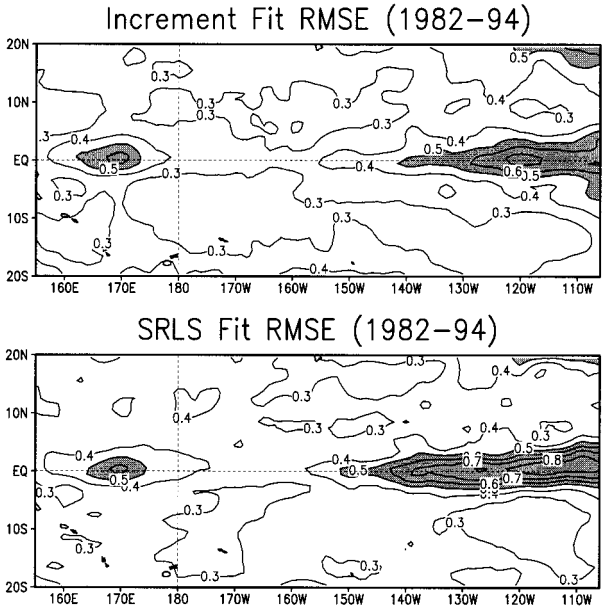

FIG. 2. Rmse with respect to the OI over the 1982–94 period for the improved analysis method (top), and the SRLS method (bottom), in units of °C.

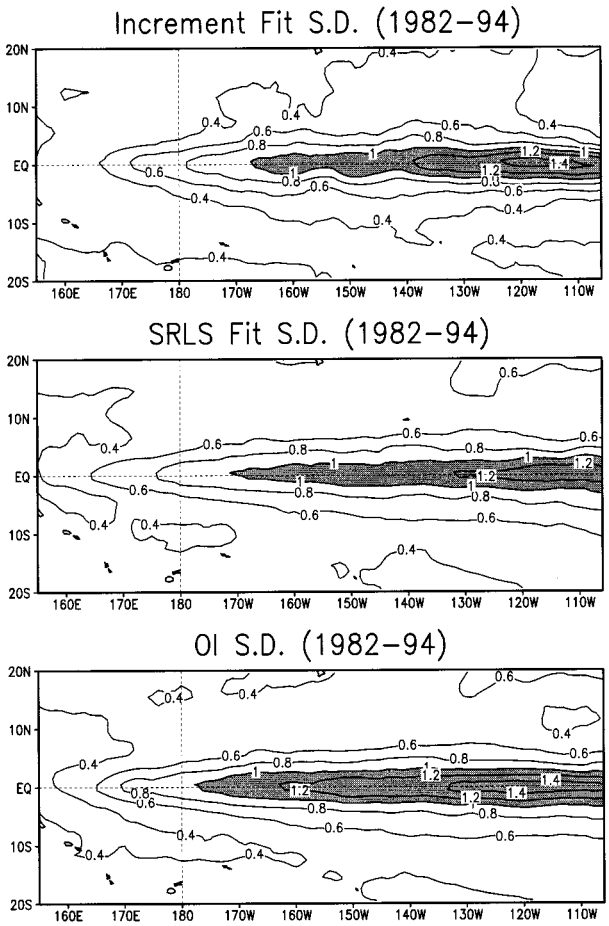

FIG. 3. Standard deviation over the 1982–94 period of the improved analysis (top), the SRLS method (middle), and the OI (bottom), in units of °C.

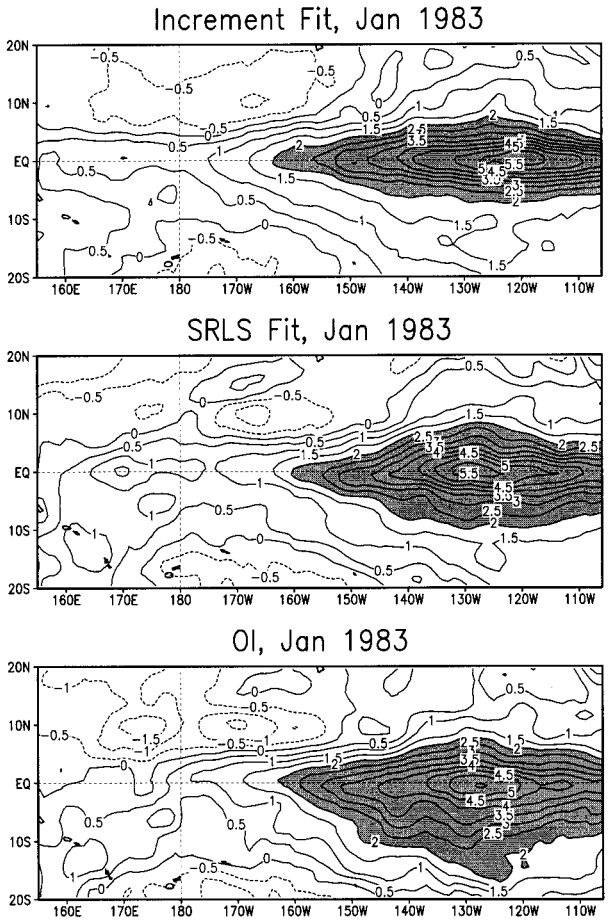

FIG. 4. Interpolated analyses (°C) from the improved method (top) and the SRLS method (middle), and the OI analysis (bottom), for January 1983.

and there are several questionable observations in the eastern equatorial Pacific (discussed in section 2e). For that month, the OI shows weak to moderate cool episode conditions. However, several strong, erroneous positive superobservations in the eastern Pacific badly bias the SRLS method. Mostly because of our data checking, the new method avoids much of the problem. There is not enough accurate ship data to fit the cool episode properly, so the new analysis simply produces weak anomalies in the eastern Pacific.

In July 1987 (Fig. 6), the new method better represents the strength and shape of the warm episode anomalies across the Pacific, compared to the SRLS method which underestimates the magnitude along the equator. In SRLS May 1988 was examined because of the existence of positive anomalies slightly off the equator, which were sampled by ships, and strong negative anomalies on the equator, which were not well sampled. The new method better represents the phase and extent of the cool anomalies, compared to the SRLS method, although both methods underestimate the magnitude (Fig. 7). Because cool episodes develop from strong upwelling and are often confined to within 5° of the equator, we may in general expect them to be underrepresented when gridding from sparse 5° resolution data.

The analysis of K97b also uses UKMO 5° superobservations to grid data, but they account for much less variance because of their use of 5° EOFs. Their method grids the entire anomaly statistically, and so requires many more EOFs from a much longer base period. Thus, they need to use a coarse-grid analysis to develop their basis EOFs. Increments can be much more accurately described at finer resolution using fewer EOFs based on the shorter period, higher quality OI analysis. Use of OI-based EOFs allows interpolation to a 1° resolution, and use of fewer EOFs filters out noise associated with higher modes. Another difference between their analysis and the increment-based analysis is that they weight EOFs by their eigenvalue, and thus the leading modes are given much more weight. This may not always be desirable because modes that dominate one period may be less dominant in another. Weighting by eigenvalue also forces their method to strongly fit to the leading

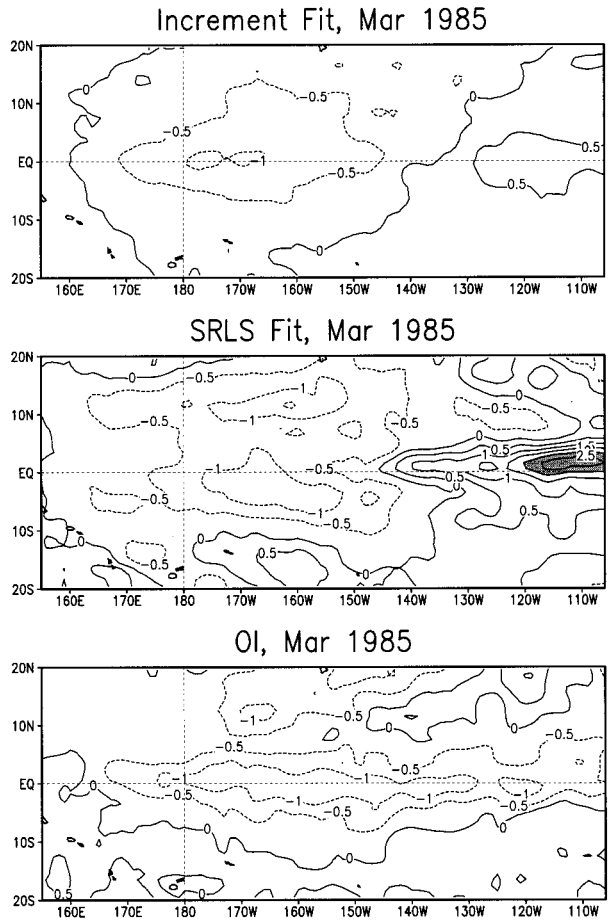

FIG. 5. Interpolated analyses (°C) from the improved method (top) and the SRLS method (middle), and the OI analysis (bottom), for March 1985.

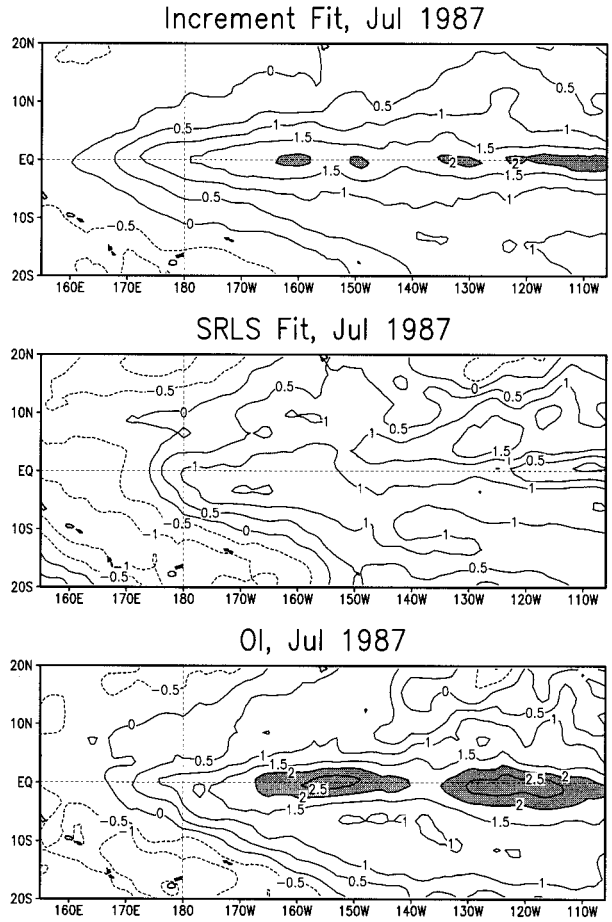

FIG. 6. Interpolated analyses (°C) from the improved method (top) and the SRLS method (middle), and the OI analysis (bottom), for July 1987.

modes, which can cause overfitting in situations where there are errors in extremely sparse data. However, even with all of the differences, their results are consistent with the present study in all but the most extremely data-sparse situations. Although the method of K97b is theoretically optimal, its assumptions are only approximately satisfied and therefore it cannot be assumed to be superior based on theory.

b. The full period

It is reassuring that the new method improves analyses in the recent, well-sampled period. However, the more important test is for earlier periods when sampling is more sparse. In particular for the tropical Pacific SST, sampling is too sparse in the pre-1950 period to reliably use the SRLS method. Time series of the *spatial* standard deviation (σ_s) for both the new method and the SRLS analyses (see SRLS for a definition of σ_s) indicate much greater stability for the improved method when data are sparse (Fig. 8). Variations in σ_s are consistent throughout the record for the new method except in extreme data-sparse periods when the analysis is damped. However, with the SRLS method they increase enormously in the pre-1950 period. The number of observations within the analysis region show that in this case the SRLS method becomes unstable with fewer than about 80 observations (Fig. 9).

As noted before, the new method analysis is the average of a forward running analysis and a backward running analysis. This is done to use information in time both prior to and after the analysis time. In periods with poor or erratic sampling, there can be significant differences between the forward and backward runs. During periods with good sampling there are, as expected, relatively small rms differences between the forward and backward runs (Fig. 10). Largest differences occur when there is some, but sparse sampling. When there are very few data the anomaly field is somewhat flat from either direction.

Time series of the Niño 3.4 area (5°S–5°N by 170°–120°W) SST anomalies from the two runs when differences are large help to illustrate why these difference can occur (Fig. 11). With adequate sampling the two

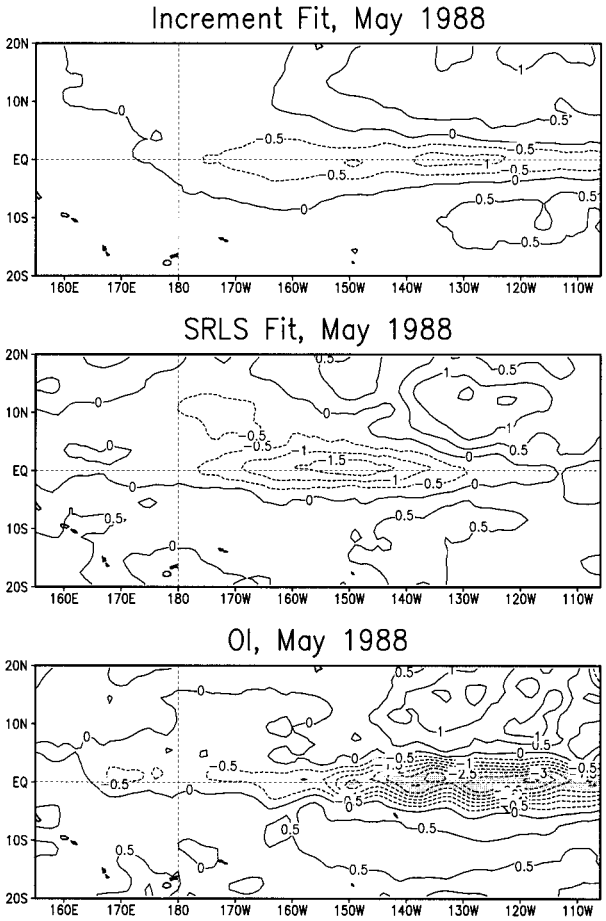

FIG. 7. Interpolated analyses (°C) from the improved method (top) and the SRLS method (middle), and the OI analysis (bottom), for May 1988.

analyses are close, but they damp in different directions through bad sampling periods (e.g., see the 1860–66 period). Since the analysis is developed by adding increments onto the guess anomaly, it may take several months to build up a large anomaly when sampling is sparse. Removal of excessively large increments in the data-checking step can further complicate the situation. Note that the forward and backward runs rarely have significant anomalies of different signs, and then only when data are extremely sparse. Most differences are caused by one giving a strong anomaly while the other gives a weaker anomaly of the same sign. Differences between the forward and backward runs could be greatly reduced by increasing f_v^* and forcing these data-sparse periods to reject all or nearly all modes. However, that would weaken all analyzed anomalies. Since the cross-validation tests show that the best balance is obtained with $f_v^* = 0.005$, we keep that setting.

Comparison to the OI in the most recent period shows that the average of the forward and backward runs is more accurate than either individual run. However, large differences between the forward and backward runs indicate times when we should have less confidence in the analysis due to insufficient sampling. The average of the two runs for the Niño 3.4 region shows reasonable interannual variations after about 1870, indicating at least adequate sampling for most of that period (Fig. 12). The difference between the forward and backward runs suggests that we should use these analyses carefully before 1880, and in the 1910–28 and 1938–43 periods. Comparison of the Niño 3.4 average to the optimal average of seasonal SST anomalies in the Niño 3 region beginning in 1870 (Smith et al. 1994) shows consistent results from both analyses.

To illustrate the type of anomaly patterns obtained from the reconstruction, we show warm and cool January anomaly pairs, for four widely separated times with different levels of sampling. For 1910 and 1912 (Fig. 13), a cool episode is indicated in 1910, with the strongest anomaly near 140°W, while the 1912 warm episode shows a more typical anomaly structure. Anomalies for 1940 show weak, diffuse warm-episode conditions, while 1943 shows a more typical cool episode pattern (Fig. 14). For both of these warm–cool pairs (the 1910–12 pair and the 1940–43 pair) the sampling is relatively sparse, and we should be skeptical about the precise shape and magnitude of the anomaly patterns. However, the Niño 3.4 averages from the forward and backward runs both show consistent anomaly pairs associated with these periods, so the sampling is sufficient to tell the general state of the tropical Pacific.

Anomalies for January 1956 and 1958 (Fig. 15) indicate a weak cool episode followed by a strong warm episode, respectively, while for 1973 and 1974 they indicate a warm episode followed by a strong cool episode (Fig. 16). For both of these periods sampling is fairly good and we can have much more confidence in the details than with the first two shown. As comparisons during the dependent period show, even during relatively well-sampled periods the anomalies patterns still have errors. The rmse map for the dependent period indicates the magnitude of the errors that may by typical after 1960, when sampling is comparable to the dependent period (Fig. 9).

Comparison of these independent-period analyses to those of K97b indicate that the higher resolution of this analysis produces stronger anomalies. In addition, the increment-gridding analysis shows less spatial variability associated with sampling patterns. However, the overall patterns in the two analyses are consistent.

5. Conclusions

We have developed a method for gridding SST anomalies within the tropical Pacific for an extended period beginning in the nineteenth century. There is generally not enough data to reliably interpolate to a grid before 1880. There is also not enough data during and just after the First World War, and also a lack of data early in the Second World War. The decline in sampling associated

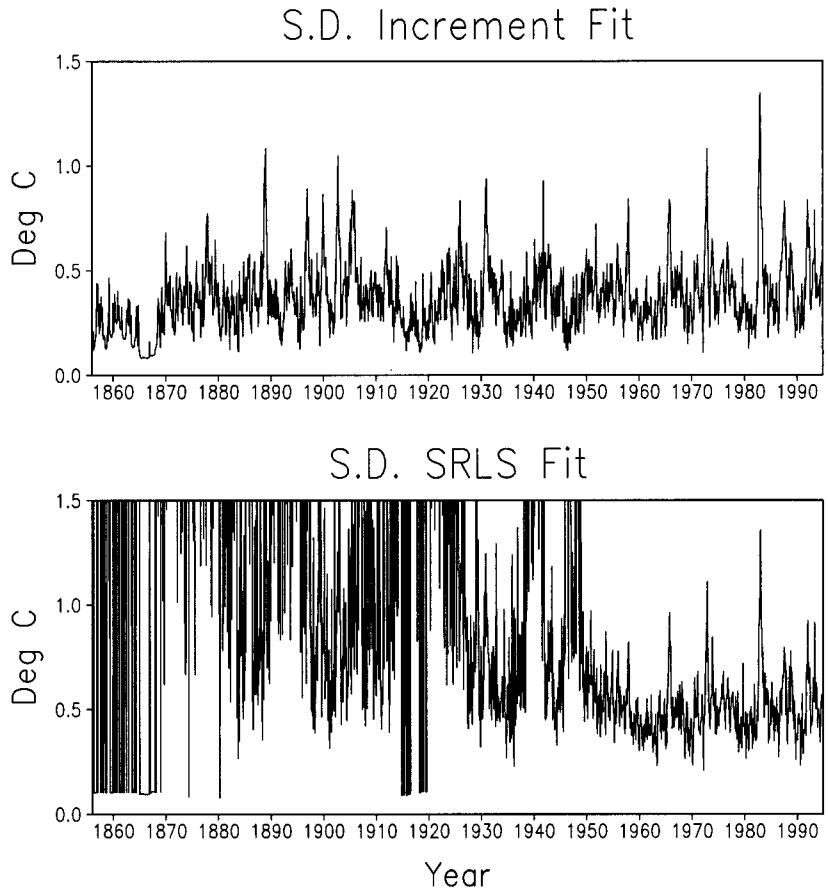

FIG. 8. Spatial standard deviation, monthly from 1856 to 1995, from the improved analysis (top) and the SRLS method analysis (bottom), in °C.

with the First World War is more severe than at any other time after 1880. The gridded SST anomalies for the continuous period beginning in the early 1920s incorporate enough data to at least roughly reflect interannual variations throughout the period, as shown by the time series of Niño 3.4 SST anomalies (Fig. 14).

Anomalies are generally better represented by this analysis method than by the analysis method of Smith

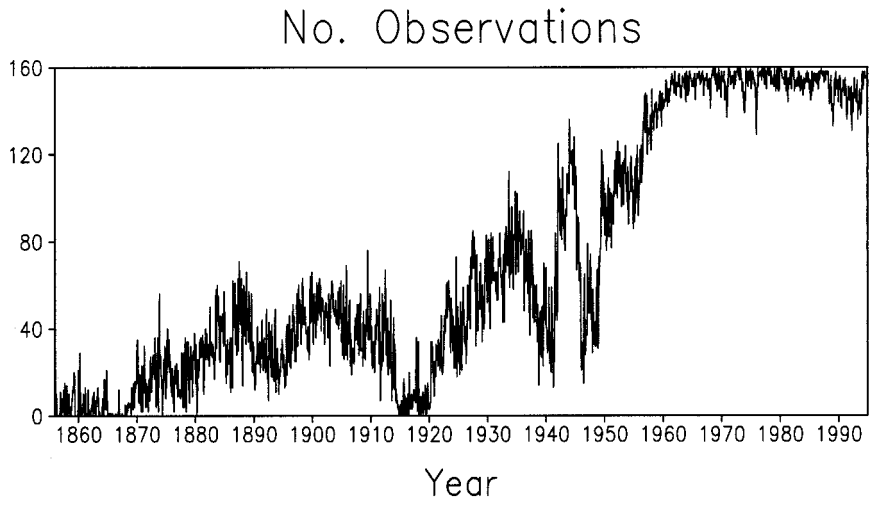

FIG. 9. Number of total observations each month, 1856–1995, in °C.

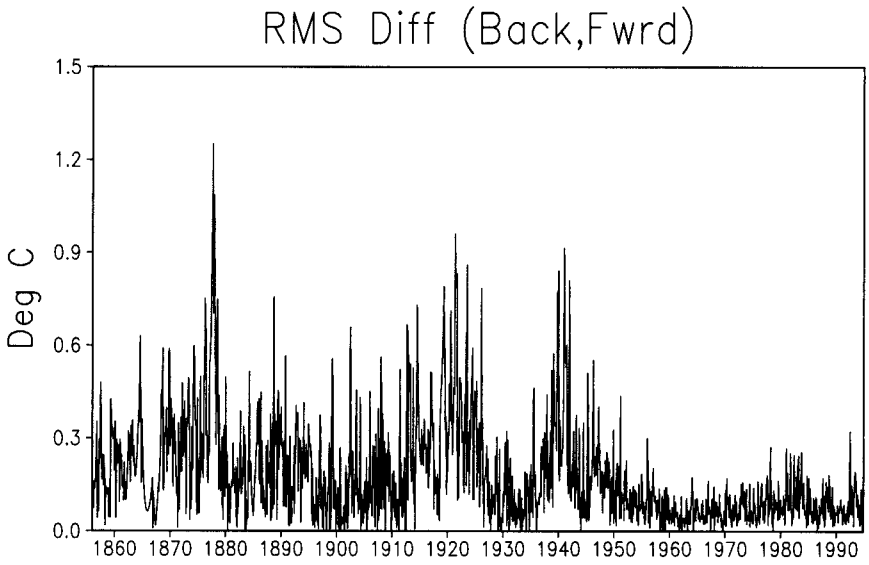

FIG. 10. Spatial rms difference (for the entire averaging area) between the forward run and the backward, in °C.

et al. (1996) for the 1950–92 period. For both techniques cool-episode anomalies, associated with fairly narrow equatorial upwelling, may be weakened because of the coarse resolution of the superobservations used here. In

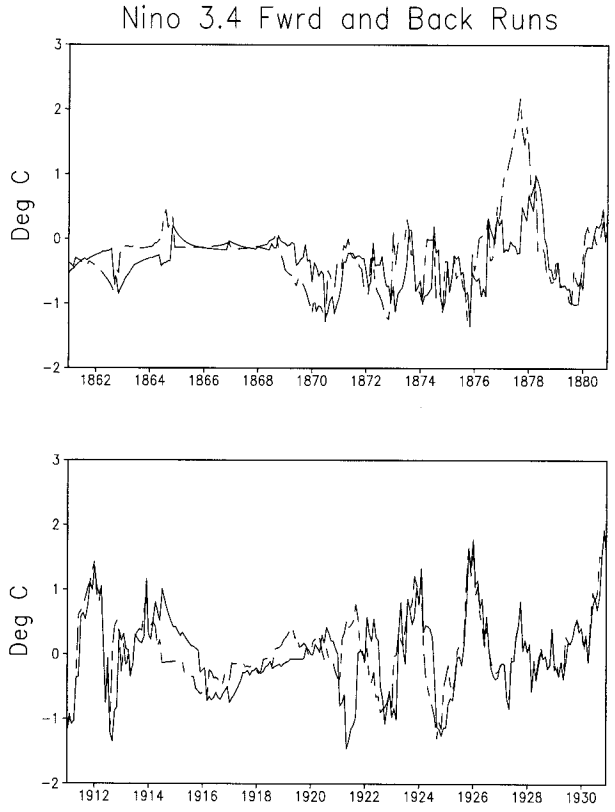

FIG. 11. Niño 3.4 area-averaged SST anomalies from the forward run and from the backward run, for 1862–80 (top) and 1912–30 (bottom), in °C.

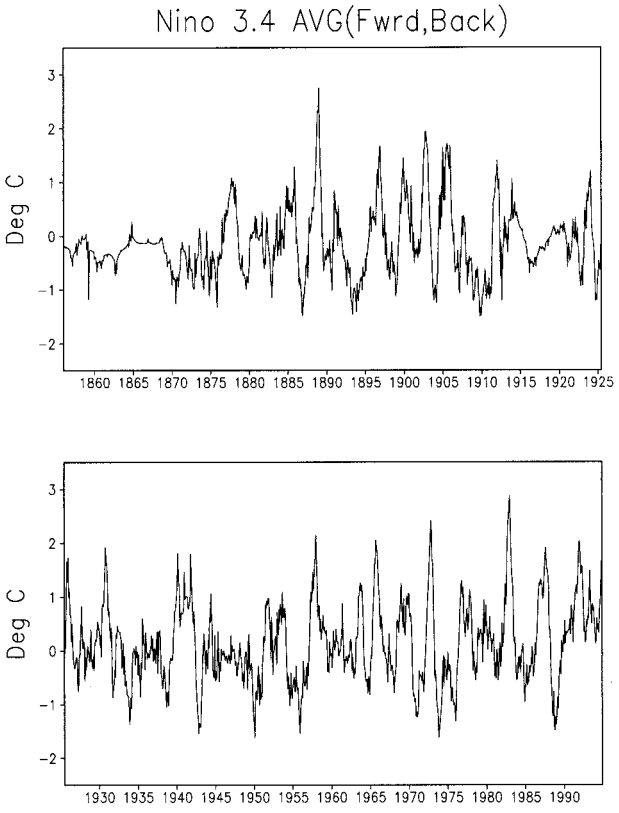

FIG. 12. Niño 3.4 area-averaged SST anomalies from the average of the forward and backward runs, for 1856–1995, in °C.

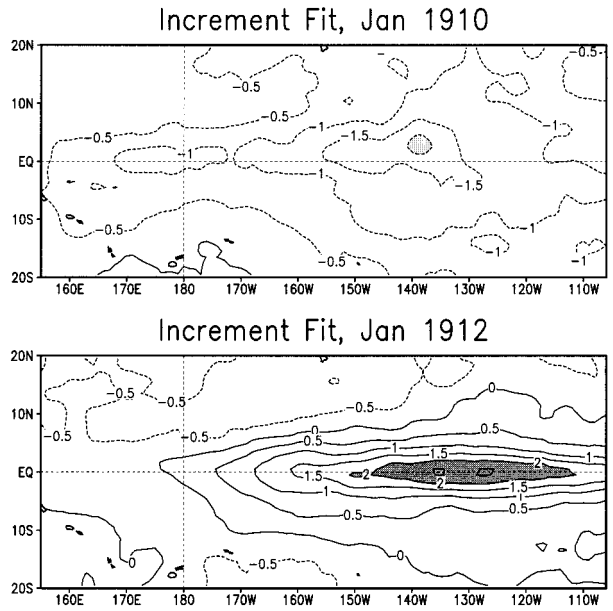

FIG. 13. Interpolated analyses (°C) for January 1910 (top) and January 1912 (bottom).

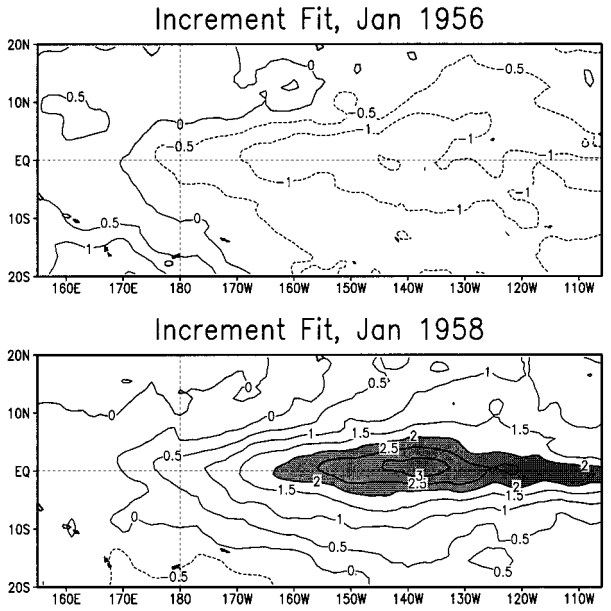

FIG. 15. Interpolated analyses (°C) for January 1956 (top) and January 1958 (bottom).

the dependent period, the new analysis method is only slightly better than the method of SRLS. In the pre-1950 period, when data are often much more sparse, the method of SRLS cannot be used but this method can because it avoids overfitting and incorporates temporal information and a gross data check.

K97a and K97b (hereafter K97) applied an elegant analysis method to Atlantic and global SST, respectively, based on minimization of a penalty function. Their approach also uses an EOF basis for spatial interpolation and the lag-one autocorrelation to tie together succeeding maps, but additionally accounts for some sampling error (in contrast to our gross error check) and errors associated with truncation and order of modes. Beyond the latter, there are two other very important differences between their approach and that described here. First, to avoid the necessity to represent interdecadal and longer variability in our EOFs, we fit the low-frequency

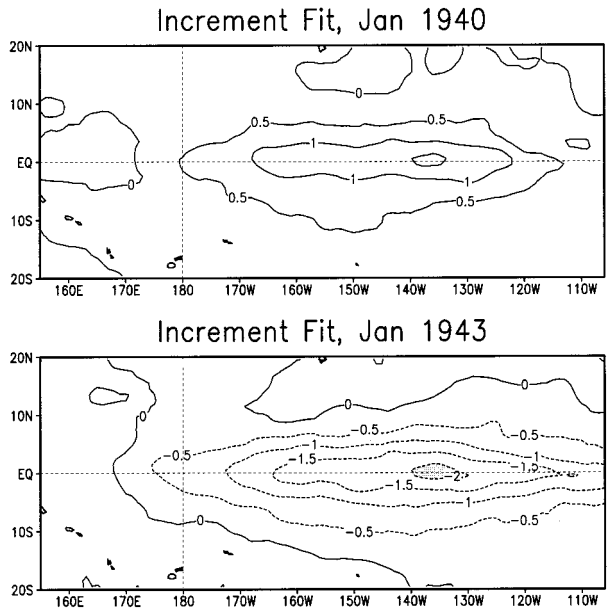

FIG. 14. Interpolated analyses (°C) for January 1940 (top) and January 1943 (bottom).

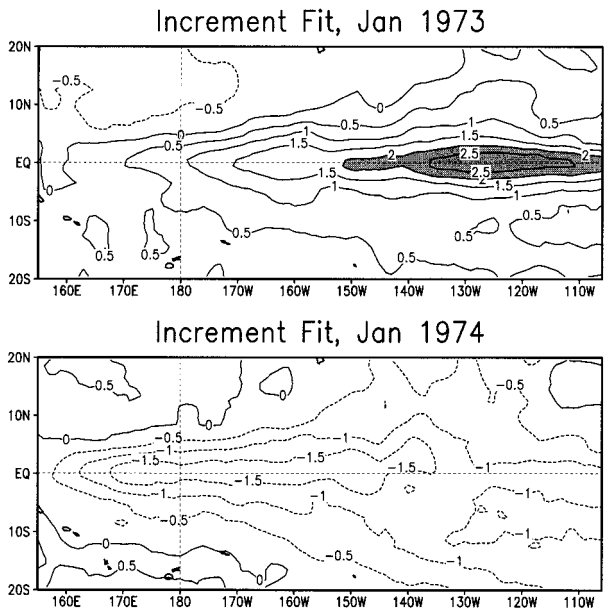

FIG. 16. Interpolated analyses (°C) for January 1973 (top) and January 1974 (bottom).

variations separately and add them back onto the analysis after the EOF fitting. We also fit anomaly tendencies rather than anomalies and add the tendencies onto an analysis first guess. Thus, modes can be developed from the relatively short period (in interdecadal terms) using the more accurate OI data to ensure the proper representation of the structural details of the EOFs. K97 instead analyze anomalies directly with EOFs estimated from an almost 50-yr record of the more coarsely resolved UKMO ship data. The other important difference in the methods is that K97 fit to a large fixed number of EOFs, in contrast to our screened subsets of a modest number of rotated EOFs. It is unclear to what extent their approach is vulnerable to overfitting in very sparse data situations.

The leading EOFs computed are only estimates of the actual leading modes, the variance associated with those leading modes is only approximately constant in time, and data errors only approximately obey the error assumptions incorporated into the method of K97. Therefore, the K97 method is only approximately optimal, and in particular it may have overfitting problems when data are excessively sparse. Those problems are greatly reduced with the addition of more data. It may be possible to completely eliminate overfitting in the K97 method with the use of more accurate EOFs and additional data-error checking. Comparisons shown here indicate that our method may better represent the shape and strength of anomalies, but a more systematic set of diagnostic comparisons of the different approaches may be needed to identify their relative strengths and weaknesses. A collaborative effort to do this is desirable.

The methods of gridding data developed here may be applicable to the global SST as well as to other climatological fields, such as sea level pressure anomalies. Longer reliable series of gridded fields of these variables should help us to better understand seasonal to interannual climatic variations.

Acknowledgments. We especially thank R. W. Reynolds for his many suggestions about ways to improve the analysis method and his review of the results. We also thank K.-C. Mo and C. F. Ropelewski for constructive discussions, and A. Kaplan for providing his analysis for comparisons.

APPENDIX

OA

Optimal averaging (OA) is used to form 10° superobservations for computation of the low-frequency SST anomaly. The OA of a region is a weighted average of data in or near the region. Weights are chosen to minimize the mean-squared error that should occur, provided that the covariance between points and the random data error can be approximated. It is assumed that there are no systematic errors in the data.

The system of equations that must be solved to obtain the weights is the same as in Smith et al. (1994), except that here we normalize the weights as was recommended by Kagan (1979) when averaging data that contain trends. Other differences between this method and that of Smith et al. (1994) are that here the spatial correlation scales are slightly larger (since we are averaging anomalies rather than increments), and data from outside the averaging area and time may be used to form the OA. As in Smith et al. (1994), Gaussian functions are used to estimate the covariance between points as a function of distance. By using simple Gaussian functions in the OA step, the low-frequency anomaly may assume any form supported by the data.

REFERENCES

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

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

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

O'Lenic, E. A., and R. E. Livezey, 1988: Practical considerations in the use of rotated principal component analysis (RCPA) in the diagnostic studies of upper-air height fields. *Mon. Wea. Rev.*, **116,** 1682–1689.

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

Rayner, N. A., M. N. Ward, D. E. Parker, and C. K. Folland, 1995: Using EOF analysis to create a new GISST data set for forcing GCMs. *Workshop on Simulations of the Climate of the Twentieth Century Using GISST,* Res. Tech. Note CRTN56, 11 pp. [Available from Hadley Centre, Meteorological Office, London Road, Bracknell, Berkshire RG12 2SK, United Kingdom.]

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

——, and ——, 1995: A high resolution global sea surface temperature climatology. *J. Climate,* **8,** 1571–1583.

Richman, M. B., 1986: Rotation of principal components. *J. Climatol.,* **6,** 293–335.

Shriver, J. F., and J. J. O'Brien, 1995: Low frequency variability of the equatorial Pacific Ocean using a new pseudo-stress data set: 1930–1989. *J. Climate,* **8,** 2762–2786.

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. Environmental Research Laboratories, 262 pp. [Available from Climate Research Program, Environmental Research Laboratories, 325 Broadway, Boulder, CO 80303.]

Smith, T. M., R. W. Reynolds, and C. F. Ropelewski, 1994: Optimal averaging of seasonal sea surface temperatures and associated confidence intervals (1860–1989). *J. Climate,* **7,** 949–964.

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

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