JOURNAL OF GEOPHYSICAL RESEARCH, VOL. 103, NO. C9, PAGES 18,567-18,589, AUGUST 15, 1998

Analyses of global sea surface temperature 1856–1991

Alexey Kaplan, Mark A. Cane, Yochanan Kushnir, Amy C. Clement,
M. Benno Blumenthal, and Balaji Rajagopalan
Lamont–Doherty Earth Observatory of Columbia University, Palisades, New York

Abstract

Global analyses of monthly sea surface temperature (SST) anomalies from 1856 to 1991 are produced using three statistically based methods: optimal smoothing (OS), the Kalman filter (KF) and optimal interpolation (OI). Each of these is accompanied by estimates of the error covariance of the analyzed fields. The spatial covariance function these methods require is estimated from the available data; the time-marching model is a first-order autoregressive model again estimated from data. The data input for the analyses are monthly anomalies from the United Kingdom Meteorological Office historical sea surface temperature data set (MOHSST5) [*Parker et al.*, 1994] of the Global Ocean Surface Temperature Atlas (GOSTA) [*Bottomley et al.*, 1990].

These analyses are compared with each other, with GOSTA, and with an analysis generated by projection (P) onto a set of empirical orthogonal functions (as in *Smith et al.* [1996]). In theory, the quality of the analyses should rank in the order OS, KF, OI, P, and GOSTA. It is found that the first four give comparable results in the data-rich periods (1951-1991), but at times when data is sparse the first three differ significantly from P and GOSTA. At these times the latter two often have extreme and fluctuating values, prima facie evidence of error. The statistical schemes are also verified against data not used in any of the analyses (proxy records derived from corals and air temperature records from coastal and island stations). We also present evidence that the analysis error estimates are indeed indicative of the quality of the products. At most times the OS and KF products are close to the OI product, but at times of especially poor coverage their use of information from other times is advantageous.

The methods appear to reconstruct the major features of the global SST field from very sparse data. Comparison with other indications of the El Niño – Southern Oscillation cycle show that the analyses provide usable information on interannual variability as far back as the 1860s.

1. Introduction

The recent surge of interest in decadal to centennial climate variability is prompted by the need to distinguish between natural and anthropogenic changes in the past century and by immediate interest in decadal cycles that impact societies, such as the Sahel drought or the U.S. dust bowl. In our data-driven science the study of longer periods quickly runs into the limitations imposed by data availability. For most climate variables, adequate global coverage is rarely available before 1950, providing only a few realizations of decadal variability, too few to permit any firm conclusions. Thus we must find ways to extend these data sets into the past. Some extant historical data has not yet been incorporated into our standard data sets [*Parker et al.*, 1995], and proxy techniques such as dendrochronology and geochemical analyses of corals and ice cores provide a limited amount of "new" observations of the past; otherwise, there is no alternative but to extract as much information as possible from the scanty instrumental record available to us.

Sea surface temperature (SST) is one of the more important climate variables in the observational database. The record we have is derived almost entirely from observations taken on volunteer observing ships and thus is concentrated in shipping lanes. Coverage is especially poor before 1880 and during the two World Wars of the twentieth century. In this paper we present global analyses of SST from 1856 to 1991. Our analysis methods are specifically designed to recover large-scale features from sparse data. These, we presume, are the more climatically important ones. The methods sacrifice detail which could be captured at times of greater data coverage.

Kaplan et al. [1997] (hereafter referred to as K97) reported on a number of procedures that were used to produce analyses of Atlantic sea surface temperature (SST) fields from 1856-1991. The focus of that paper was methodology (the development, justification, and verification of analysis and error estimation techniques). For that reason, attention was restricted to the relatively data-rich Atlantic north of approximately 30^{o}S. In this paper we extend the analysis to the global SST field over the same period. Of previously reported work, our procedures are most similar to the schemes of *Smith et al.* [1996] and *Shriver and O'Brien* [1995], with differences that become important only at times of sparse coverage. Our methods differ from successive correction methods [e.g., *Pan and Oort*, 1990] and from the National Centers for Environmental Prediction optimal interpolation (NCEP OI) analysis of *Reynolds and Smith* [1994] in utilizing nonlocal covariance information.

Section 2 summarizes our methodology, while section 3 has a brief account of the data source. A detailed description of methods and data are given by K97. The examples of sections 4–7 are offered in hopes of building the reader's confidence in the products these methods produce, as well as a sense of their limitations. The schemes generate theoretical error estimates as an important part of the product, but it is necessary to establish that they are truly indicative of the uncertainty of the analysis. (Formulas for the theoretical error estimates are given by K97.) Most often we verify against the NCEP OI product [*Reynolds and Smith*, 1994], which makes use of remote sensing data and the Tropical Atmosphere/Ocean array of the Tropical Ocean Global Atmosphere program (TOGA TAO) and other moorings in addition to merchant ship observations. This product is available only for the satellite era. Though we take this analysis as a standard of comparison, we are mindful of the fact that it is not free of error, especially during periods when the satellite observations are contaminated by volcanic emissions (not to take anything away from the NCEP OI bias correction of satellite data which makes use of in situ measurements). Section 4 presents some comparisons of different methods and products. Section 5 reports global-scale experiments with data withheld, while section 6 compares with independent data at points (corals, islands, and coastal stations). Section 7 presents some standard indices based on SST such as NINO3 (mean SST for the eastern equatorial Pacific 5^{o}S–5^{o}N, 150^o–90^{o}W). Section 8 discusses the reasons for believing our product to be superior to the presently available ones, together with our understanding of its shortcomings.

2. Methods

We have observations $\mathcal{T}_n^{\rm o}(x_n^{\rm o})$ at times $n = 1, \cdots, N$ available at certain subsets (with elements $x_n^{\rm o}$) of a set of M space grid points x since not all M grid points are observed at each time. We wish to construct the best estimate of a gridded field $\mathcal{T}_n$ for all times $n = 1, \cdots, N$. We construct a linear estimate by solving a least squares minimization problem. In this we follow Gauss and most of our colleagues in meteorology and oceanography, whether they employ optimal interpolation, Kalman filters or smoothers, adjoint methods, four-dimensional variational assimilation, etc. [cf. *Ghil and Malanotte-Rizzoli*, 1991].

The four methods we consider may be viewed as least squares estimation procedures that minimize different versions of the cost function

$$\mathcal{S}[\alpha_1, \alpha_2, \cdots, \alpha_N] = \sum_{n=1}^{N} (\mathcal{H}_n \alpha_n - \mathcal{T}_n^{\circ})^T \mathcal{R}_n^{-1} (\mathcal{H}_n \alpha_n - \mathcal{T}_n^{\circ}) + \sum_{n=1}^{N-1} (\alpha_{n+1} - \mathcal{A}_n \alpha_n)^T \mathcal{Q}_n^{-1} (\alpha_{n+1} - \mathcal{A}_n \alpha_n). \quad (1)$$

The unknowns α_n , $n = 1, \cdots, N$ here are the vectors of amplitudes $(\alpha_n^1, ..., \alpha_n^L)$ at time n of a set of spatial basis functions e_l, $l = 1, \cdots, L$ such that

$$\mathcal{T}_n = \sum_{l=1}^{L} e_l \alpha_n^l = E\alpha_n \quad (2)$$

(the matrix E comprises e_l, $l = 1, \cdots, L$ as its columns). We will take E to be defined by a set of gridded empirical orthogonal functions (EOFs) and expect $L \ll M$. $\mathcal{H}_n = H_n E$ maps α_n into the observations at time n (H_n does this in the physical space),

$$\mathcal{T}_n = H_n \mathcal{T}_n + \varepsilon_n^{\circ} = \mathcal{H}_n \alpha_n + (H_n \varepsilon_n^r + \varepsilon_n^{\circ}) \stackrel{\text{def}}{=} \mathcal{H}_n \alpha_n + \tilde{\varepsilon}_n^{\circ}, \quad (3)$$

where $\tilde{\varepsilon}_n^{\circ}$ includes not only the "traditional" observational and sampling error ε_n°, but also the influence of truncated modes ε_n^r, i.e., those not included in the L modes comprising E (the equation defining new variables is marked "def" above its equal sign). This "effective observational error" $\tilde{\varepsilon}_n^{\circ}$ is assumed unbiased and white in time with expected covariance $\mathcal{R}_n = \langle \tilde{\varepsilon}_n^{\circ} \tilde{\varepsilon}_n^{o\,T} \rangle = R_n + H_n \langle \varepsilon_n^r \varepsilon_n^{r\,T} \rangle H_n^T$. Observational and sampling error covariance R_n is estimated on the basis of intrabox data variability and the number of measurements per gridbox, while $H_n \langle \varepsilon_n^r \varepsilon_n^{r\,T} \rangle H_n^T$ is computed from the truncated EOF modes. The $\mathcal{A}_n$, $n = 1, \cdots, N-1$ are linear models relating successive times,

$$\alpha_{n+1} = \mathcal{A}_n \alpha_n + \varepsilon_n^{\mathrm{m}}, \quad (4)$$

where the model error (system noise) $\varepsilon_n^{\mathrm{m}}$ is unbiased and white with expected covariance $\mathcal{Q}_n$. In this work we follow K97 by finding $\mathcal{A}_n = \mathcal{A}$ (and the corresponding $\mathcal{Q}$) as a first-order autoregressive (AR(1)) model fitting data in EOF space. Our four methods result from differing versions of (1); in all cases more detail may be found in work by K97.

2.1. Optimal smoother (OS). Solving (1) as is results in the OS solution; our solution procedure is the fixed interval OS algorithm of Rauch–Tung–Striebel [*Rauch et al.*, 1965]. The OS solution is the best estimate of $\mathcal{T}_n$ because it utilizes all available information (observations at all space points and all times and all the estimated error information, as well as the model $\mathcal{A}$). This solution minimizes the misfit between the analysis and observations subject to the constraint (4), the model equation. Thus it is a four-dimensional variational method with the model imposed as a weak constraint [*Sasaki*, 1970]. In this it differs from the usual adjoint method, in which the model is imposed as a strong constraint, effectively treating it as if it were error free [cf. *Ghil and Malanotte-Rizzoli*, 1991; *Miller and Cane*, 1996].

2.2. Kalman filter (KF). In the KF the analysis at time $K \leq N$ is based only on the data at times $n \leq K$; no information is used from times after K. It is equivalent to solving (1) for each K with the N in (1) replaced by K. It thus uses data at all times only at the final time $K = N$ and so is formally suboptimal at all other times. As with the OS procedure, the analysis is best fit to the data subject to the time-dependent relation (4).

2.3. Optimal interpolation (OI). In this procedure, there is no connection between successive times; $\mathcal{A}_n = 0$ for all n, so $\mathcal{Q}_n = \mathcal{C}$. Here $\mathcal{C} = E^T C E$ is the time-independent spatial covariance of α and $C = \langle \mathcal{T} \mathcal{T}^T \rangle$ is the covariance of the signal in the physical space. Hence (1) reduces to

$$\mathcal{S}_n^{\mathrm{OI}}[\alpha_n] = (\mathcal{H}_n \alpha_n - \mathcal{T}_n^{\circ})^T \mathcal{R}_n^{-1} (\mathcal{H}_n \alpha_n - \mathcal{T}_n^{\circ}) + \alpha_n^T \mathcal{C}^{-1} \alpha_n \quad (5)$$

for $n = 1, \cdots, N$. The OI analysis at time n uses no information from other times but does make optimal use of all the observational data available at time n. It fills holes by using the covariance relations to other points and weights observations according to the error estimate $\mathcal{R}_n$. Even when an observation is available at a point x the analysis takes note of its estimated error and seeks to improve on it by using data at covarying points.

2.4. Projection method (P). Only the first sum in (5) is retained, so there is no further constraint in time or space as in sections 2.1-2.3. One finds the amplitudes of the basis functions that give the best projection on the data. This is essentially the procedure of *Shriver and O'Brien* [1995] and *Smith et al.* [1996]. Data voids are filled with the covariance information carried by the EOFs that comprise the basis set. With ample data the error $\mathcal{R}$ will be much smaller than $\mathcal{C}$, the covariance of the variable itself, so the OI solution will be controlled by the first term of (5). In this case the P analysis will be close to the OI analysis. If there are fewer than L data points, then the EOF amplitudes will be undetermined and the P method will fail. With somewhat more data these amplitudes will be formally determined, and the P scheme will produce an analysis, but one likely to be strongly influenced by observational errors.

In principle the quality of the analysis improves as one moves up the list from P to OS; formal error estimates are given by K97. However, the OI method adds the need for the covariance estimates $\mathcal{C}$, while the OS and KF require $\mathcal{Q}$ and $\mathcal{A}$. Each step up relies on more information and will only improve the analysis if this information is close enough to being correct.

The most important information is the stationary spatial covariance of the field, C. It is the source of the EOFs and the key ingredient in all the recipes for filling data voids. Our algorithm for estimating C is arguably the most original feature of our analysis procedure. A complete account of it appears in section 3.2 of K97; a heuristic description will be given here.

We begin by calculating C_{raw}, the covariance of the observational data in the relatively data-rich period 1951-1991. Because of data gaps and observational error, C_{raw} will not be a true covariance, even of the sample. We therefore smooth C_{raw} in each spatial direction. As explained in K97, this is done as if the filter had been applied to the spatial field at each time and would be an exactly equivalent operation if the data had no gaps. The result, C_{f}, preserves the large-scale relations in the original covariance, which we presume to be correctly estimated by the data sample, while eliminating the small-scale variations, which we presume to be dominated by observational error. The smoothing also removes some of the variance of the original data. This is undesirable since we believe that the original observational sample is a good estimate of the true variance once the error variance is removed. That is, we estimate true variance as $V = \text{diag}\,[C_{\text{raw}} - R]$, where R is the estimated observational error (discussed below). This formula for V assumes that the errors are not correlated with the signal. We now rectify the unfortunate effect of the smoothing by inflating the variance in C_{f} back up to the values in V; see K97 for a precise account. The resulting covariance matrix C_{v} has the same correlation structure as the smoothed matrix C_{f} but a better estimate of the variance; from it we calculate the EOFs that are to be used as a basis set. There is one final step, taken to correct for the fact that smoothing the covariance artificially reddens the variance (eigenvalue) spectrum of the EOFs. A percentage β of each eigenvalue is taken away; then the variance removed in this way is added back uniformly to each eigenvalue, whitening the spectrum somewhat. The percentage β is tuned as described by K97, the general idea being that in the OI analysis the variance of the first EOF, the most robust structure, will come close to the estimated first eigenvalue.

In the absence of a physical model connecting SSTs at successive times we also estimate the $\mathcal{A}_n$ from the data; they are multivariate AR(1) models. In principle they could be full matrices that vary monthly, but the tests described by K97 led to the conclusion that the best we could do with the limited data available is to take $\mathcal{A}_n$ to be independent of time and diagonal in form; $\mathcal{A}_n = \mathcal{A} = \text{diag}[a_1, \cdots, a_L]$, where

$$a_l = \sum_{n=1}^{N-1} \alpha_{n+1}^l \alpha_n^l \Big/ \sum_{n=1}^{N-1} \alpha_n^l \alpha_n^l. \tag{6}$$

Here α_n^l are components of the solution obtained from the OI analysis (which does not use the model $\mathcal{A}$) . Values a_l are lag-one correlations for EOF amplitudes; they are almost monotone decreasing from $a_1 = 0.97$ to $a_{80} = 0.23$. $\mathcal{A}$ can be characterized as a damped persistence model in the EOF space because every month it predicts EOF amplitudes of the previous month scaled down by the factors a_l. It follows immediately from (4) that $\mathcal{Q}_n = \mathcal{Q} = \mathcal{C} - \mathcal{A}\mathcal{C}\mathcal{A}^T$, independent of time.

3. Data

The basis of all historical SST analyses are reports from merchant ships. All our analyses take the Global Ocean Surface Temperature Atlas (GOSTA) monthly averages of individual SST observations for 5^o latitude by 5^o longitude bins [*Bottomley et al.*, 1990] as the observational data; we use version MOHSST5, which incorporates the Comprehensive Ocean–Atmosphere Data Set (COADS) compilation [*Woodruff et al.*, 1987] for the boxes where originally there were no GOSTA data. It also corrects for systematic biases in the SST measurements [*Parker et al.*, 1994; *Folland and Parker*, 1995], which is important because while our methods reduce the influence of random measurement errors, they cannot correct for biases. In all four procedures the climatological monthly mean for the period 1951–1980 is subtracted out; the analysis product is the anomaly about this mean.

We take the observational error to be spatially uncorrelated (R_n is diagonal; see K97 for discussion of this assumption) but inhomogeneous, depending on the local variance and on the number of observations in each gridbox that month. The idea is that the error is a combination of a measurement error and a sampling error that arises from having too few measurements to estimate precisely the mean over the month and the $5^o \times 5^o$ area. As in K97, we make use of the greater resolution of the COADS monthly mean data to estimate the intrabox variance σ. Then with $N_{\rm obs}$ observations in box m for month n (these data are part of the GOSTA products) the observational error variance is

$$R_n^{mm} = \sigma^2(m)/N_{\rm obs}(m,n). \tag{7}$$

As a benchmark for our analyses in the period 1982–1991 we use the NCEP OI analysis of *Reynolds and Smith* [1994]. It is produced weekly at $1^o \times 1^o$ grid resolution; we average it to monthly $5^o \times 5^o$ means. This analysis blends all available in situ and bias-corrected satellite data to produce what should be considered as the best estimate of recent SST fields available at present.

4. Comparisons of Methods and Products

4.1. Comparison of different methods. We compare among the four analysis products (OS, KF, OI, and P) and the GOSTA product. In principle they should generally rank in the order OS, KF, OI, P, and GOSTA. With very sparse data coverage, however, the projection product P can exaggerate errors and be worse than the input data GOSTA because the EOF amplitudes are ill determined. In view of the limited skill of the time-dependence model $\mathcal{A}$ we may anticipate that the OS and KF products will not improve much on the OI except at times when data is extremely scarce.

Figure 1 shows that overall the OI product is close to the OS, and the KF is closer still, while the projection P is quite different, especially in the Pacific, which is not as well sampled as the Atlantic or Indian Ocean. Figure 2 shows that the differences are large at the times when data is sparse, especially prior to 1880 and during the World Wars; the 1918 influenza pandemic is evident. Differences are virtually nonexistent in the well-sampled period after 1960. The root-mean-square (rms) difference between OS and P (OI, KF) is 1.14^{o}C (0.21^{o}C, 0.13^{o}C) for the period 1856-1900; it drops to 0.34^{o}C (0.12^{o}C, 0.08^{o}C) for 1901-1950 and is only 0.04^{o}C (0.03^{o}C, 0.02^{o}C) for 1951-1991. Note that P is occasionally further from OS than GOSTA is (Figure 2). This occurs when there are too few data to give a robust projection amplitude for the EOFs, leaving them inordinately sensitive to observational errors.

Hereafter we will generally show only the OS analysis. That OS is superior in fact as well as theory was checked in the experiments with withheld data described below; the comparison is with the NCEP OI analysis. The globally averaged rms differences among the three "optimal" schemes are rarely more than a few hundredths of 1^{o}C, whereas P can differ by as much as 1^{o}C in years like 1918 or 1941, when observations are especially scarce. The detailed comparison among these schemes and its theoretical interpretation is presented in greater detail by K97.

4.2. Sensitivity to the number L of EOFs retained. For simplicity we discuss only the OS procedure; results for the OI and KF are similar. The discussion by K97 and that by *Smith et al.* [1996] provide some justification for choosing L =80 EOFs, which retains 75% of the total estimated variance. But since this choice must remain somewhat arbitrary, we explore the consequences of different values of L. Figure 3 maps the rms difference over the period 1856-1991 between the L =80 solution and several other choices. Little is gained over this period by increasing the number of EOFs; reducing the number appreciably has a noticeable effect. Figure 4 shows the time series of the global rms difference from the L =80 solution. In contrast to the different analysis methods of Figure 2, data density has little relation to the average differences, although there is some tendency for the differences to vary more in periods of poor data coverage. As in the maps of Figure 3, L =60 or L =120 are not very different from L =80. Figure 5 shows the rms differences with the NCEP OI for L =40, 60, 80, and 120; L =15 (not shown) is clearly inadequate. There is little to distinguish between 60, 80, and 120, even in this period (1982-1991) when data is relatively plentiful. The difference map for L =120 resembles the pattern of estimated observational error (not shown) and the pattern of difference between NCEP OI and GOSTA (not shown), suggesting that this analysis is "drawing" too close to the data.

Hereafter we use L =80 for our work on the grounds that little if anything is lost, and it is computationally more convenient. It captures 75% of the global SST variance, so obviously we are not even trying to retrieve much of the small-scale structure in the global SST field. The sparse data coverage of all but the recent three or four decades does not allow it. We do expect to recover the structure of the larger scale, lower frequency features of the global SST more likely to impact or reflect coherent climate variability. The more sophisticated methods hold the promise of doing so even in the face of substantial data gaps.

4.3. Comparison with other products. The global average of the differences between the OS with 80 EOFs and NCEP OI (Figure 5c) is 0.3°C. Figure 6 puts this value in perspective by showing rms differences for 1982-1991 between OS and GOSTA (Figure 6a), OS and *Smith et al.* [1996] (Figure 6b), OS and NCEP OI (Figure 6c), and *Smith et al.* [1996] and NCEP OI (Figure 6d). The mean rms differences on Figures 6b, 6c, and 6d are all about 0.3°C (more precisely 0.31°, 0.32°, and 0.29°C). Thus the difference between our OS and the NCEP OI is about the same as the difference of the latter with the *Smith et al.* [1996] analysis, despite the fact that our procedures are independent of the NCEP OI while *Smith et al.* [1996] use the NCEP OI product as the sample for the covariance structure at the heart of their technique. However, the mean rms difference between OS and GOSTA (Figure 6a) is 0.41°C, so our OS analysis is closer to the NCEP OI than it is to the GOSTA data that provides its raw material.

We conclude that the limits of data coverage and of "state-of-the-art" analysis techniques leave an uncertainty in present day analyses of the order of 0.3°C. Figure 6e is the OS error estimate (for the period 1982-1991) at the scales of the analysis (i.e., smoothed to 80 EOFs) and so would give the data error in the analysis; its global mean value for this period is about 0.1°C. Figure 6f shows the estimated error due to smoothing for our OS scheme, i.e., the estimated variance in the truncated modes. The global mean is above 0.3°C. All analysis procedures (ours, *Smith et al.* [1996], and NCEP OI) do some smoothing (implicitly if not explicitly), but no two smooth in exactly the same way. The different effective smoothings may account for much of the differences, but it is difficult to make a quantitative estimate. A priori we expected the OS to be closer to *Smith et al.* [1996] than to NCEP OI since the smoothings of the first two are somewhat similar. We expected differences between OS and NCEP OI to be greater than those between *Smith et al.* [1996] and NCEP OI because *Smith et al.*'s [1996] EOFs are derived from the NCEP OI fields. In fact, the differences among these three products turned out to be comparable. It seems likely that the different covariance estimates used by each are responsible for the major part of the differences. A deeper understanding of why this is so will require a cooperative effort among the groups producing these products and is beyond the scope of this paper.

5. Verification on Global Fields: Experiments With Withheld Data

We repeated an experiment described by K97 (Figures 7 and 8 of K97) by withholding data in a large region of the well-sampled North Atlantic (see Figure 7). Generally the schemes are able to fill the hole with fair accuracy, with the error decreasing as the procedures become more sophisticated and computationally demanding: for the P, OI, KF, and OS schemes the maximum differences between the analyses with and without the data are 0.193°, 0.165°, 0.162°, and 0.161°C, respectively. The greatest surprise in these experiments is illustrated in Figure 7, which compares the NCEP OI to the global analyses of the present paper and the Atlantic analyses of K97. With all data included, the global and Atlantic only results are quite comparable, but when data is withheld, the global analysis does a better job of filling the hole. It seems that data from other oceans add useful information, even though the Atlantic is well observed outside the withheld region.

We now turn to some withheld data experiments directly relevant to the 136 year global analysis. In these experiments we apply the data coverage mask from a sparsely sampled month in the past to a well sampled month in the recent period; below we refer to each of these as a "simulated analysis." Such a reconstruction with few observations may be compared to the analysis with the full set of ship observations as well as to the NCEP OI product. The first of these (Plate 1) uses the sparse data coverage of December 1868 for December 1991. The OS analysis begins in January 1982 and runs 10 years through December 1991; it uses the corresponding data masks from 1859 to 1868. In order to account for the lesser number of observations at the earlier time, the 1990s data is contaminated by Gaussian noise, white in time and space, with variance equal to the difference between the estimated observational error variances at the simulated and actual times (compare (7)).

The OS analysis for December 1868, shown in Plate 1b, indicates this to be an El Niño period. Note from Plate 1a that apart from a few observations near the South American coast (which support the El Niño pattern for this month), there are no observations in the equatorial Pacific. The analysis procedure has reconstructed the characteristic tropical Pacific El Niño signature largely on the basis of the highly correlated and relatively well observed Indian Ocean. The error analysis (Plate 1g) indicates that the reconstruction of the eastern equatorial Pacific is well above the error bars and should be believed. However, the patterns in the North Pacific are generally weaker than the estimated error. Other evidence that 1868-1869 was an El Niño – Southern Oscillation (ENSO) warm event includes the deficit in all-India rainfall in 1868 (Plate 6) and the historical analysis of *Quinn* [1992].

Additional corroboration of the 1868 analysis comes from applying the 1868 data mask to 1991. The result, shown in Plate 1c, may be compared with the "full" OS analysis with all data shown in Plate 1d. In view of the difference in data coverages the similarity between the two analyses is surprising. Though the two are not identical (e.g., the equatorial warm anomaly extends farther west in the simulated analysis, while the North Pacific warm spots are stronger in the west and weaker in the east), all the major features of the full analysis appear in the simulated one. This point is reinforced by the comparison of the two to the NCEP OI analysis (Plates 1e and 1f); the differences between each of our analyses and the NCEP OI are far larger than the differences due to data coverage. The largest differences are at high latitudes, especially in the southern hemisphere at the limits of our data coverage. Only a few regions of the Pacific distinguish between our two analyses. A way in which the two differ strongly is the estimated error (Plates 1g and 1h). While the analyses are close, the added data of the recent period strongly increases one's confidence in the full analysis.

Plates 2 to 5 are similar in format to Plate 1; they pair the years 1877 with 1986, 1918 with 1990, 1941 with 1982, and 1942 with 1988. The OS analysis uses the corresponding data masks in each case; for example, for the 1877/1986 pair the data masks for 1873 to 1882 are applied to data for 1982 to 1991. The general conclusions suggested by the 1868/1991 case of Plate 1 are supported by the following cases: the major features of the full OS analysis are captured surprisingly well in the data sparse simulations; confidence is higher in the full analysis (as measured by the expected errors), and the differences between the two OS analyses are typically smaller than the difference of either with the NCEP OI analysis. In all five cases the OI and KF are quite similar to the OS. While the full projection (P) analyses are quite close to the others in this data-rich period, the simulated P analyses are drastically different, with large regions in error by several degrees Celsius; the global mean rms difference with the NCEP OI for the simulated P analyses corresponding to Plates 1e-5e are 2.0^o, 1.0^o, 1.2^o, 1.0^o, and 0.6^oC, respectively.

We comment further on a few particular features of these figures. In 1877 there was a devastating famine in India (Plate 6). It is well known [e.g., *Quinn*, 1992] that there was a very strong El Niño in that year; this is apparent in Plate 2b, where the warm anomaly reaches a maximum of 3.6^oC. As in 1868, the Indian Ocean is well sampled, but this time there is some data in the equatorial Pacific cold tongue region (near 130^oW). The simulated analysis (Plate 2c) reconstructs the moderate 1986 El Niño; differences with the NCEP OI (Plate 2e) are generally small in amplitude and areal extent. The coverage in 1918 is among the very poorest in the record. The 1918 analysis (Plate 3b) indicates a strong event [cf. *Quinn*, 1992] but is rather featureless away from the eastern equatorial Pacific, which did have some data (see Plate 3a). It is hard to believe this unusual global pattern, but the simulation (Plate 3c) does do quite a credible job of reconstructing the interesting warm patterns of the non-El Niño year of 1990 (compare Plates 3d and 3e). The 1941 analysis (Plate 4b) depicts the well-known strong event of that poorly covered war year. It suggests that the warm anomaly was unusually broad and the western North Pacific was unusually cold. Both OS analyses of the mammoth 1982 El Niño (Plates 4c and 4d) extend the warm region farther to the west than the NCEP OI analysis (Plates 4e and 4f).

Plate 5 pairs the poorly observed 1942 ENSO cold event [*Kiladis and Diaz*, 1989] with the relatively well observed cold event of 1988. According to the OS analysis (Plate 5b) the 1942 event was comparable in strength to the recent one. Again, the simulated OS analysis is quite close to the full OS (Plates 5c and 5d). The OS analyses are similar to the NCEP OI with one important exception: their version of the equatorial cold tongue is markedly warmer in the grid boxes between 5^oN and 5^oS, 105^oW and 135^oW (Plates 5e and 5f).

This discrepancy is much like that for May 1988 considered by *Smith et al.* [1996]; see their Figures 5 and 7. The covariance structure in our analysis, which is derived entirely from ship reports, does not have enough information in the vicinity of the equator and 125^{o}W to appreciate that the temperatures there should be related to those to the east and west not the north and south; it has insufficient knowledge of the structure of the equatorial cold tongue. The *Smith et al.* [1996] covariance, being built from the NCEP OI, has been educated by the TOGA TAO moorings at 125^{o}W and 110^{o}W to know the strength and narrow meridional scale of the cold tongue. Note that the rms difference of our analyses from the NCEP OI (Figure 5) has a steady rise in the eastern equatorial Pacific when the number of EOFs used increases. Thus the analysis believes more and more details of our estimated covariance structure, which unfortunately is deficient in this region.

Figure 8 shows the averaged SST anomaly over the $10^o \times 10^o$ box centered at 0^o, 125^{o}W from NCEP OI, *Smith et al.* [1996], our OS, and GOSTA. For OS and *Smith et al.* [1996] the correlation for 1950 to 1991 is 0.91. The correlation between OS and the NCEP OI is 0.92, while the correlation between Smith et al. and NCEP OI is 0.96. These are all quite close, but the OS is late with the rapid SST rise in 1982 and late with the rapid decline in 1988.

To check our understanding of the problem, we redo our analysis replacing the ship observations from 1982-1991 with the NCEP OI analysis in the construction of the covariance matrix. (We differ from *Smith et al.* [1996] in continuing to use data from 1951 to 1981. Though the NCEP OI is of exceptionally high quality, we regard the period after 1981 to be too unusual to use as the sole basis for a long-term analysis. It does not provide a set of patterns and a distribution of variance representative of long-term variability.) The result for this "new" optimal smoother (OSn) analysis is shown in Figure 8 as the dashed-dotted curve. Agreement is improved for the 1982 and 1988 periods. Figure 9 shows the NCEP OI, Smith et al., original OS, and OSn analyses for May 1988 [cf. *Smith et al.*, 1996, Figure 5]. The OSn analysis produces a sharper cold tongue than the original OS, bringing it closer to the NCEP OI. Like the *Smith et al.* [1996] field, it is not as intense as the NCEP OI.

While producing the OSn analysis was very useful for our interpretation of the actual analysis error in the vicinity of 0^o, 125^{o}W, it posed the additional problem of estimating EOFs from an inhomogenous sample. Since we are yet to address this problem adequately, we do not feel comfortable enough with the OSn analysis to use it instead of the OS as our standard product. This problem of obtaining reliable pattern sets from nonuniform samples will be addressed in our future work.

The withheld data experiments in Plates 1 to 5 are compromised by the fact that data from the verification period was also used in estimating the covariance matrix in the initial stage of the analysis. This difficulty is not easily overcome; the data from 1982-1991 is needed in the covariance estimation to connect the tropics and midlatitudes. It is also needed for verification because in the presatellite era, there are no estimates of global SST comparable in quality to the NCEP OI fields. So it is difficult to make an unequivocal conclusion from verifications before 1980.

To test the analysis for a period not used in estimating the covariance structures, we carried out additional experiments as follows: the NCEP OI SST fields from January 1992 to July 1996 were chosen as the "true" solution. These "true" data were resampled and corrupted by noise according to the data availability and our estimates of observational error for a certain period in the past. Then the results of the analysis were compared with the NCEP OI solution. In the first experiment we used the data mask of 8 years earlier (so that the data of 1994 was resampled according to data availability in 1986), and in the second experiment it was taken 76 years earlier (so that 1994 was resampled as if it was 1918).

With the 8 year shift, the rms of the simulated observational error is 0.35^{o}C, and the rms difference of the OS and the NCEP OI analyses is 0.37^{o}C. Most of this is the portion of the NCEP OI solution which does not project onto the 80 EOF patterns used in our analysis; the difference of the OS solution and the projection of the NCEP OI solution on the set of EOFs is 0.11^{o}C. In good agreement with this number, our theoretical estimate of the large-scale error in OS is 0.08^{o}C. For the second experiment (1916-1920 sampling) the rms error for available observations is 0.74^{o}C, and there are many locations where the SST is not observed at all. The rms deviation of the OS from the NCEP OI is now 0.48^{o}C, and its deviation from the NCEP OI projected onto 80 EOFs is 0.31^{o}C. The last number is consistent with the theoretical error estimate, 0.28^{o}C. For comparison, a similar experiment for the GOSTA data from the period 1984–1988, resampled and corrupted to simulate 1916–1920, gave rms differences of 0.45^o and 0.33^{o}C for the full- and large-scale only NCEP OI fields. (The theoretical error depends only on the sampling and so is the same as in the 1990s case.) Overlap with the period used in the covariance estimate made little difference in the result.

The years 1992-1996 are outside the period used in constructing the covariance estimate and are marked by strikingly different behavior. Thus these experiments demonstrate that even with limited data the reduced space OS is able to reconstruct the global SST in a period when the covariance structure is different from the one used by the analysis procedure.

6. Verification Against Independent Point Observations

Figure 10 plots the OS analysis at Tarawa, together with 3σ error bars; also shown is the GOSTA product and the proxy series from corals [*Cole et al.*, 1993]. The OS is an obvious improvement on GOSTA, which clearly can have large errors at times when the sampling is sparse. Values are often very far from the coral proxy values; moreover, at some of these times the GOSTA values are outliers (values beyond the limits of any of the well-sampled times). This alone makes it unlikely that they are real, but in addition, it is difficult to imagine a plausible physical scenario that would permit such values; what possible source could there be for such cold waters at Tarawa? We did comparisons with other coral time series as well, but they raised enough interesting issues about the interpretation of the proxy records to warrant a separate report.

In Figure 11 we compare the GOSTA and OS SST products to air temperature measurements at nearby land boxes from the compilation of *Jones* [1994]. The land measurements are not SST and may have problems of their own [e.g., *Parker*, 1994], so the correlation need not be high everywhere even if the SST products were perfect. All curves are distributions of time correlation coefficients over the set of points used in the comparisons. The dashed curve is for the GOSTA product, and the solid curve is for the OS analysis but only over the set of grid points and times where GOSTA data is available. The difference between the two shows that the OS improves on the GOSTA even when observations are available. The dashed-dotted curve is for the OS analysis at points where GOSTA data is not available, generally in the nineteenth century. It is not surprising that it is not as good as for points where there are observations, and it is encouraging that it is only marginally below the GOSTA values.

7. SST Indices

Figure 12 shows time series of some standard SST indices based on the OS analysis. The GOSTA version, shown for comparison, is again questionable when the number of observations is very small. The error bars on the OS values are generally quite tight, except for the period before about 1870 and during the World Wars. The nineteenth century Pacific indices have large uncertainties.

Evidence as to the quality of these series is offered in Figure 13, which gives the rms differences of the indices of Figure 12 between GOSTA and NCEP OI (dots) and between the OS analyses and the NCEP OI (open circles) for 1982-1991, together with the estimated 2σ error bar. Also shown is the difference between *Smith et al.* [1996] and the NCEP OI (star). In addition to the full GOSTA and OS using all available ship data for 1982-1991, we show the five simulated GOSTA and OS analyses of Plates 1-5 in which the data is sampled to match the data availability in those five poorly observed periods. The results indicate that even in poorly sampled times the variability in these indices is captured in our OS analysis. Only the 1860s case is markedly worse than the full coverage case. In contrast, the GOSTA values are typically quite poor in all but the well-sampled modern period.

The indices are averages over large areas, which serve to reduce sampling error. With this in mind, it is somewhat disconcerting that the disagreements (notably including *Smith et al.* [1996] versus NCEP OI) are as large as they are. Although our error bars almost always cross zero, indicating that the differences are within expected 95% error limits, we believe that these differences are due less to observational and sampling error than to differences in the covariance structures assumed in the different analyses. This issue was discussed in section 4; Figure 13 shows that its influence on area averages is not trivial.

There is no independent data that allows a straightforward validation of these series. However, they were chosen because they are thought to have some importance in larger patterns of climate variability, which provides some possibilities for indirect verification. Plate 6 shows a histogram of all-India rainfall index [*Sontakke et al.*, 1993], with the bars color coded according to NINO3. The well known, strong but imperfect connection between the two is obvious: El Niño years (warm NINO3) tend to be poor rainfall years. The relation breaks down in the 1850s and early 1860s, perhaps because of data problems, but recovers in the still data sparse 1870s. In particular, the analysis captures the very strong El Niño event of 1877, which was associated with drought and terrible famine in India.

The OS NINO3 was checked against the *Quinn* [1992] list of El Niño events, which is based on a variety of land-based, historical factors known to be associated with El Niño. Our NINO3 identifies 28 of the 37 events listed by Quinn as occurring after 1856, and the magnitude of the events are in good agreement. All nine mismatches are events rated moderate by Quinn, with five of the nine occurring before 1875; the most recent is in 1943. The relation is strong but not perfect; since the two are not different measures of exactly the same thing, it is hard to say how tight the relation should be [cf. *Smith et al.*, 1994]. *Kiladis and Diaz* [1989] list both cold and warm events; again, qualitative agreement is good.

We also compared our OS analysis NINO3 values with another long-lived ENSO index, sea level pressure at Darwin. For annual values their correlation is 0.89 for the 40 years 1951–1990, and for 1911–1950 (1888–1911) it is 0.80 (0.84). Since the 95% confidence limit for 40 years of data falls at 0.80, we conclude that there is no significant change resulting from either changes in data quality or genuine nonstationarity in the climate system.

Figure 14 shows the global average SST, together with global land surface temperature. The curves (smoothed by a five-year running mean filter) track closely through the twentieth century, which is reassuring. While there is no reason for them to be identical, it is difficult to envision a plausible physical scenario that would let them be strikingly different for very long. Thus, flaws in the data are the likely cause for the disparity between them in the nineteenth century. When this difference has been noted in the past, it was often attributed in part to sampling problems in the SST observations [*Houghton et al.*, 1995, and references therein]. In the twentieth century there is no 20 year period where the difference between our analysis and *Hansen et al.* [1996] data exceeds the two standard deviation level of the monthly errors. For the period from 1880-1899 the difference between the two data sets is 0.24°C which is more than seven standard deviations. The difference between our analysis and *Jones* [1994] data are even larger for 1880-1899 and keeps exceeding the two standard deviation level for 1900-1919. There are some important differences between the two land data sets [*Jones et al.*, 1991], but both indicate that the oceans were warmer than the land before 1900 by an amount unlikely to be due to sampling errors.

8. Discussion

Parker et al. [1995, p. 597] state ". . . a desirable goal would be to combine the optimum interpolation method as currently practiced [*Reynolds and Smith*, 1994] with the eigenvector interpolation technique." The work reported here meets this goal. It might be said to exceed it in that our optimal techniques are more sophisticated than current practice. We use eigenvectors (EOFs) as the basis functions for statistical estimation procedures where space time covariance information is derived from data rather than ad hoc formulations.

Global analyses of monthly sea surface temperature (SST) anomalies from 1856 to 1991 are produced using three statistically based methods: optimal smoothing (OS), the Kalman filter (KF) and optimal interpolation (OI). The development of these methods and their application to the Atlantic were described in some detail in a previous paper by *Kaplan et al.* [1997] (K97). In theory all these methods are "optimal" linear estimates, giving the best estimate using observational information at the present time (OI), or at present and all past times (KF), or at all times, past, present, and future (OS). Each method is accompanied by estimates of the error covariance of the analyzed fields. We have taken pains to verify that these error estimates do reflect the actual analysis errors.

In order to generate an analysis that is better than the input data all these methods add structural information about the fields and the observations. OI requires the spatial covariances of the SST field itself and of the observational errors. The OS and KF procedures are usually built on a time dependent physical model; here we replace it with a statistical model which requires the SST covariances at a 1 month time lag. The required spatial and lag-one covariance functions are estimated from the available data; our procedures are described briefly here and in detail by K97. The limited length of the time series and their serial correlations guarantees that the global covariance matrix estimated from data cannot be of full rank. The situation is still worse, since much of the time-space grid is observed inaccurately or not at all. Our procedure concedes this point and aims only to capture the covariance of the major features of the global fields in a matrix of low rank (relative to the dimension of the full spatial grid). Presumably these are the features of greatest interest for climate studies.

The methods would be truly "optimal" only if the true covariances were perfectly known. Still, since the covariances estimates are reasonable and likely to be decent approximations at least for the major features, one may reasonably expect substantial improvement over the raw GOSTA data that is our input. We offered considerable evidence to this effect, including comparisons to withheld data and to independent records from coastal and island stations and from coral proxies. We also showed a number of instances when the GOSTA products give implausible values even for averages over large areas.

Unrealistic values can also arise in analyses derived from projection of the data onto EOFs (such as *Smith et al.* [1996] and *Shriver and O'Brien* [1995]; see K97 for a discussion of the small differences between their versions and ours). This happens only at times when the data is too sparse to determine the full set of EOF amplitudes, and they are to some extent set by the vagaries of the noise in the observations. When data coverage is adequate (which includes most of the period over which the cited authors studied), the projection method results are comparable to OI.

These impacts of sparse data particularly affect the climatically important tropical Pacific, which is poorly observed at almost all times prior to about 1955. For example, the trend in NINO3.4 (mean SST for the eastern Pacific area 5^oS–5^oN, 170^o–120^oW) from 1900 to 1991 calculated from our OS analysis is actually negative (-0.3^{o}C/100 years) and is significantly different at the 95% level from the global warming in SST of 0.4^{o}C/100 years. This difference is consistent with a previously proposed theory [*Clement et al.*, 1996; *Sun and Liu*, 1996; *Neelin and Dykstra*, 1995] and has potential importance in the global warming debate [*Cane et al.*, 1997]. However, the trend derived from the GOSTA product, which is 0.3^{o}C/100 years and has greater uncertainty due to the higher noise level in the data, is not significantly different at the 95% level from either the mean global warming or the OS NINO3.4 cooling. All these trends are calculated by a least squares fit. In view of the high noise level in the GOSTA values we recalculated them with the far more robust Sen slope test [e.g., *Gilbert*, 1987] in which one computes the slope between all pairs of points in the data sample and estimates the overall trend to be the median value. The OS NINO3.4 trend barely changed (by 0.02, though the error bars tightened), but the GOSTA trend is now only 0.04^{o}C/100 years, which is significantly different from the mean global warming in this period.

We also calculated the trend by sampling the OS-analyzed fields according to the GOSTA coverage. The median trends came out to be -0.08^{o}C/100 years, and we conclude that the discrepancy between the OS and GOSTA values are primarily due to the incomplete sampling of the latter.

Plates 1-5 are striking examples from the nineteenth century and the periods of the World Wars (and the 1918 influenza pandemic) of the remarkable ability of these methods to reconstruct the major features of the global SST field from very sparse data. In addition to verifying the analyses by applying these data distributions to a well-sampled modern period, we offered some indirect evidence, such as ENSO effects, that these reconstructions are probably correct. However, the rather large error bars at those times sound a properly cautionary note.

These verifications do not answer all possible criticisms. The test with modern data is not proof against nonstationarity in the covariance structure. The methods glean information about structure from the recent data record and then fill gaps by presuming that missing information bears the same structural relation to what data there is. If the structures were truly different in the nineteenth century, then it fails and the error estimates fail. For example, we justified early values of NINO3 by their apparent agreement with all-India rainfall (Plate 6), but in those years the schemes fill the eastern Pacific in good measure on the basis of the relatively well observed Indian Ocean (namely Plate 1). This fails if the Pacific-Indian connection was structurally different in the nineteenth century. Note that the method assumes that the dominant structures were dominant for all times but not that their amplitudes are unchanging. This is a modeling assumption in these methods, and we regard it as the most plausible way to fill holes and certainly superior to not filling them at all. The latter falls further short of optimal than does our use of somewhat incorrect covariance structures and so is more likely to mislead. At the same time, it is essential to pay attention to error estimates and to keep in mind that with very little data, very little can be known with certainty.

The analysis of the 1988 cold event is a telling example of another problem due to limited data coverage, one stemming from the inability to define an adequate climatology of variability. The TOGA TAO array allows the cold tongue to be defined as a narrow structure, whereas the sparse coverage from ship reports provides only a fuzzy view. Never having seen such a thing, our procedures are powerless to reproduce it. We fixed the problem by making sure the procedure learned about this structure, but other structures (small scale but vigorous coastal phenomena as a probable example) may be missed because they were never observed. This episode points to the wisdom of *Smith et al.* [1996] in extracting structural information from the very recent past when the ship data is augmented by remote sensing and by research measurements such as TOGA TAO. However, while the projection methods ("eigenvector interpolation" in the language of *Parker et al.* [1995]) use only structural information, the optimal schemes also need to know the relative likelihood of the different structures. We felt it would be dangerous to take the highly anomalous 1980s and 1990s as representative of long term statistics and so blended the recent era with the presumably more representative 1950-1980 period. This issue clearly requires further investigation.

A serious shortcoming of these methods is their assumption that the data is unbiased. We know that the data is subject to biases over time as observational methods changed, and despite ingenious and tenacious attempts to remove them [*Parker et al.*, 1995], some doubts remain. While we discern no problems in the twentieth century record, comparison of the long-term trend in the global mean SST with land-based surface temperature suggests that values before 1900 have a small positive bias. This has been noted by others [cf. *Houghton et al.*, 1995] and is usually attributed to sampling problems. Our error estimates indicate that this is not likely to be large enough to account for the difference with land values, leaving measurement bias as the likely explanation. We suspect the SSTs are at fault, but the land-based observations are not problem free [*Parker*, 1994]. It is also possible that nonstationarities in the climate system such as changes in atmospheric circulation are responsible [*Houghton et al.*, 1995].

Finally, we note that even for the relatively well sampled 1982-1991 period the differences among state-of-the-art analysis products [*Smith et al.*, 1996; *Reynolds and Smith*, 1994; this study] are about 0.3°C on average. A cooperative effort among all analysis groups is needed to understand the reasons for such large differences.

Acknowledgments. We thank Michael Jackson and Robert Hackett of the United Kingdom Hadley Centre for making the MOHSST5 data available to us and for their assistance. A multitude of corrections, comments, and advice given in their reviews of this manuscript by David Parker, Dick Reynolds, Bob Livezey, and an anonymous reviewer are appreciated tremendously. Manipulation of large volumes of climate data was facilitated by the Compressing Unlimited Format (CUF) developed by Senya Basin. The help of Virginia DiBlasi-Morris in preparing the manuscript is also appreciated. This study has been supported by NOAA grants NA36GP0074, NA56GP0161, and UCSIO-10075411D/NA47GPO-188. This is Lamont-Doherty Earth Observatory contribution 5806.

References

Bottomley, M., C.K. Folland, J. Hsiung, R.E. Newell, and D.E. Parker, *Global Ocean Surface Temperature Atlas*, Her Majesty's Stn. Off., Norwich, England, 1990.

Cane, M.A., A.C. Clement, A. Kaplan, Y. Kushnir, R. Murtugudde, D. Pozdnyakov, R. Seager, and S.E. Zebiak., 20th century sea surface temperature trends, *Science*, *275*, 957–960, 1997.

Clement, A.C., R. Seager, M.A. Cane, and S.E. Zebiak, An ocean dynamical thermostat, *J. Clim.*, *9*, 2190–2196, 1996.

Cole, J.E., R.G. Fairbanks, and G.T. Shen, Recent variability in the southern oscillation: Isotopic results from a Tarawa atoll coral, *Science*, *260*, 1790-1793, 1993.

Folland, C.K., and D.E. Parker, Correction of instrumental biases in historical sea surface temperature data, *Q. J. R. Meteorol. Soc.*, *121*, 319-367, 1995.

Ghil, M. and P. Malanotte-Rizzoli, Data assimilation in meteorology and oceanography, *Adv. Geophys.*, *33*, 141-266, 1991.

Gilbert, R.O., *Statistical Methods for Environmental Pollution Monitoring*, Van Nostrand Reinhold, New York, 1987

Hansen, J., R. Ruedy, M. Sato, and R. Reynolds, Global surface air temperature in 1995: Return to pre-Pinatubo level, *Geophys. Res. Lett.*, *23*, 1665 -1668, 1996.

Houghton, J.T., L.G. Meira Filho, B.A. Callander, N. Harris, A. Kattenberg, and K. Maskell (Eds.), *Climate Change 1995 - The Science of Climate Change*, 572 pp., Cambridge Univ. Press, New York, 1995.

Jones, P.D., Hemispheric surface air temperature variations: A reanalysis and an update to 1993, *J. Clim.*, *7*, 1794–1802, 1994.

Jones, P.D., T.M.L. Wigley, and G. Farmer, Marine and land temperature data sets: A comparison and a look at recent trends, in *Greenhouse-Gas-Induced Climatic Change: A Critical Appraisal of Simulations and Observations*, edited by M.E. Schlezinger, pp. 153–172, Elsevier, New York, 1991.

Kaplan, A., Y. Kushnir, M.A. Cane, and M.B. Blumenthal, Reduced space optimal analysis for historical data sets: 136 years of Atlantic sea surface temperatures, *J. Geophys. Res.*, *102*, 27,835–27,860, 1997.

Kiladis, G.N., and H.F. Diaz, Global climatic anomalies associated with extremes of the Southern Oscillation, *J. Clim.*, *2*, 1069–1090, 1989.

Miller, R.N. and M.A. Cane, Tropical Data Assimilation: Theoretical Aspects, in *Modern Approaches to Data Assimilation in Ocean Modeling*, edited by P. Malanotte-Rizzoli, 207-233, Elsevier, Amsterdam, 1996.

Neelin, J.D., and H.A. Dykstra, Ocean-atmosphere interaction and the tropical climatology, I, The dangers of flux correction, *J. Clim.*, *8*, 1325-1342, 1995.

Pan, Y.H., and A.H. Oort, Correlation analyses between sea surface temperature anomalies in the eastern equatorial Pacific and the world ocean, *Clim. Dyn.*, *4*, 191-205, 1990.

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

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

Parker, D.E., C.K. Folland, and M. Jackson, Marine surface temperature: observed variations and data requirements, *Clim. Change*, *31*, 559-600, 1995.

Quinn, W.H., A study of Southern Oscillation–related climatic activity for A.D. 622–1900 incorporating Nile River flood data, in *El Niño Historical and Paleoclimatic Aspects of the Southern Oscillation*, edited by H.F. Diaz and V. Markgraf, pp. 119-149, Cambridge Univ. Press, 119-149, New York, 1992.

Rauch, H.E., F. Tung, and C.T. Striebel, Maximum likelihood estimates of linear dynamic systems, *AIAA J.*, *3*, 1445-1450, 1965.

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

Sasaki, Y., Some basic formalism in numerical variational analysis, *Mon. Weather Rev.*, *98*, 875-883, 1970.

Shriver, J.F., and J.J. O'Brien, Low-frequency variability of the equatorial Pacific ocean using a new pseudostress dataset: 1930-1989, *J. Clim.*, *8*, 2762-2786, 1995.

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

Smith, T.M., R.W. Reynolds, R.E. Livezey, and D.C. Stokes, Reconstruction of historical sea surface temperatures using empirical orthogonal functions, *J. Clim.*, *9*, 1403-1420, 1996.

Sontakke, N.A., G.B. Pant, and N. Singh, Construction of all-India summer monsoon rainfall series for the period 1844–1991, *J. Clim.*, *6*, 1807-1811, 1993.

Sun, D.-Z., and Z. Liu, Dynamic ocean-atmosphere coupling: A thermostat for the tropics, *Science*, *272*, 1148-1150, 1996.

Woodruff, S.D., R.J. Slutz, R.L. Jenne, and P.M. Steurer, A comprehensive ocean–atmosphere data set, *Bull. Am. Meteorol. Soc.*, *68*, 521-527, 1987.

M.B. Blumenthal, M.A. Cane, A.C. Clement, A. Kaplan, Y. Kushnir, and B. Rajagopalan, Lamont-Doherty Earth Observatory, Columbia University, P.O. Box 1000,Palisades, NY 10964-8000.
(e-mail: benno@ldeo.columbia.edu; mcane@ldeo.columbia.edu; clement@ldeo.columbia.edu; alexeyk@ldeo.columbia.edu; kushnir@ldeo.columbia.edu; rbala@ldeo.columbia.edu)

October 3, 1996; revised May 30, 1997; accepted June 13, 1997.

This preprint was prepared with AGU's LaTeX macros v4.
File crcpp formatted November 21, 1999.

Figure 1. Comparison of other analysis products with the optimal smoother (OS) analysis. (a) Projection (P), (b) optimal interpolation (OI), and (c) Kalman filter (KF). Each panel shows the root-mean-square (rms) difference (°C) with the OS analysis averaged over the entire analysis period, 1856-1991.

Figure 2. Time series of global rms differences with the OS analysis. P (solid curve), OI (dashed curve), KF (dashed-dotted curve), and G, the Global Ocean Surface Temperature Atlas (GOSTA) compilation (dotted curve). The GOSTA is averaged only over the points where there is data.

Figure 3. Differing numbers of empirical orthogonal functions (EOFs). Temporally averaged (1856-1991) rms difference (°C) between the $L=80$ EOF (75% of the variance) OS product and the OS products with (a) $L=15$ (44%), (b) $L=40$ (62%), (c) $L=60$ (70%), and (d) $L=120$ (82%).

Figure 4. Differing numbers of EOFs. Global rms difference between the $L=80$ EOF OS product and the OS products with (a) $L=15$ (solid curve), (b) $L=40$ (dashed curve), (c) $L=60$ (dashed-dotted curve), and (d) $L=120$ (dotted curve).

Figure 5. Temporally averaged (1982-1991) rms difference (°C) between the NCEP OI product and the OS products, (a) 40 EOFs, (b) 60 EOFs, (c) 80 EOFs, and (d) 120 EOFs.

Figure 6. The rms differences (°C) for 1982-1991 between (a) OS and GOSTA, (b) OS and *Smith et al.* [1996], (c) OS and NCEP OI, and (d) *Smith et al.* [1996] and NCEP OI. (e) The estimated OS error for the period 1982-1991 at the scale of the analysis, i.e., smoothed to 80 EOFs. (f) The estimated OS error due to the truncation; it is the variance in the EOFs beyond $L=80$.

Figure 7. The rms differences (°C) with the NCEP OI analysis for the period 1982-1991 in the Atlantic. (a) Atlantic OS analysis as by *Kaplan et al.* [1997] (K97), (b) Atlantic OS analysis as by K97 with data withheld in the shaded region, (c) global OS analysis, and (d) global OS analysis with data withheld in the shaded region.

Figure 8. The sea surface temperature (SST) anomaly (°C) averaged over the $10^o \times 10^o$ box centered at 0^o, 125^oW from NCEP OI (thick solid curve), *Smith et al.* [1996] (dashed curve), OS (thin solid curve), and GOSTA (circles). Also shown is the "new" OS (OSn) analysis (dashed-dotted curve) based on a covariance that includes the NCEP OI for the period 1982-1991 (see text).

Plate 1. Analysis for December 1868 and its verification through the experiment with 1991 data. Simulated OS analysis for December 1991 using the data distribution of 1868 versus the standard OS analysis for December 1991 with all GOSTA data. (a) GOSTA observations for December 1868; (b) the OS analysis for December 1868; (c) simulated OS analysis for December 1991 using data distribution of 1868 (see text for details); (d) OS analysis for December 1991; (e) simulated OS minus NCEP OI analysis, December 1991; (f) full OS minus NCEP OI analysis, December 1991; (g) estimated error in simulated OS analysis, December 1991; and (h) estimated error in full OS analysis, December 1991. Units are degrees Celsius.

Plate 2. Same as Plate 1, but for December 1877 verified by December 1986.

Plate 3. Same as Plate 1, but for December 1918 verified by December 1990.

Plate 4. Same as Plate 1, but for December 1941 verified by December 1982.

Plate 5. Same as Plate 1, but for December 1942 verified by December 1988.

Figure 9. Analyses of SST (°C) for the cold (La Niña) month of May 1988. (a) NCEP OI, (b) *Smith et al.* [1996], (c) OS, and (d) OSn.

Figure 10. Comparison of ship-derived products with the coral-based proxy record from Tarawa atoll [*Cole et al.*, 1993]. The dashed curve is the proxy record, the solid curve is the OS analysis, and the dotted curves are 3σ error bars (99.7% confidence limits). The circles are the GOSTA values.

Figure 11. Comparison of SSTs with adjacent land values (1856-1991). All curves are distributions of correlation coefficients in time over the set of points used in the comparisons. The dashed curve is for the GOSTA product, and the solid curve is for the OS analysis, but only over the set of grid points and times where GOSTA data is available. The dashed-dotted curve is for the OS analysis at points where GOSTA data is not available, generally in the nineteenth century.

Figure 12. Time series of some SST indices derived from the OS product (dark curve), with 3σ error bars (light curves); the GOSTA product is also shown (dots). The indices are area-averaged SST anomalies: NA, North Atlantic (35°-60°N, 60°W-0°); NP, North Pacific (30°-50°N, 150°E-120°W); NINO3 (5°S-5°N, 150°-90°W); NINO3.4 (5°S-5°N, 170°-120°W); GEq, global equator, (5°S-5°N); and GL, global, (30°S-60°N).

Figure 13. Differences from the NCEP OI values of SST indices derived from OS products (open circles), GOSTA (dots), and *Smith et al.* [1996] (star). From left to right are the full OS and GOSTA products (all data from 1982-1991), the *Smith et al.* [1996] analysis, and five simulated OS analyses of Plates 1-5: data masks of 1859-1868, 1873-1882, 1910-1919, 1941-1950, and 1936-1945. The vertical line connected to each OS circle indicates the 2σ error bar. The indices are area-averaged SST anomalies as in Figure 12.

Figure 14. Time series of various global temperature estimates (five-year running means). SST from our OS analysis (thick solid curve), GOSTA SST (curve with dots), COADS SST (curve with circles), and land temperatures from *Hansen et al.* [1996] (dashed curve) and *Jones* [1994] (dashed-dotted curve). The differences in the nineteenth century contrast with the consistency in the twentieth century. The corrections in GOSTA largely account for its differences with COADS.

Plate 6. All-India rainfall 1856–1991 [*Sontakke et al.*, 1993]. The bars are color coded to indicate the strength of the annual OS SST anomaly (°C) in the NINO3 area.