An Optimal Regional Averaging Method with Error Estimates and a Test Using Tropical Pacific SST Data

SAMUEL S. SHEN

Climate Prediction Center, National Weather Service/NOAA, Camp Springs, Maryland, and Department of Mathematical Sciences, University of Alberta, Edmonton, Alberta, Canada

THOMAS M. SMITH, CHESTER F. ROPELEWSKI, AND ROBERT E. LIVEZEY

Climate Prediction Center, National Weather Service/NOAA, Camp Springs, Maryland

(Manuscript received 17 December 1996, in final form 3 November 1997)

ABSTRACT

This paper provides a systematic procedure for computing the regional average of climate data in a subregion of the earth surface using the covariance function written in terms of empirical orthogonal functions (EOFs). The method is optimal in the sense of minimum mean square error (mse) and gives an mse estimate of the averaging results. The random measurement error is also included in the total mse. Since the EOFs can account for spatial inhomogeneities, the method can be more accurate than those that assume a homogeneous covariance matrix. This study shows how to further improve the accuracy of optimal averaging (OA) by improving the accuracy of the eigenvalues of the covariance function through an extrapolation method. The accuracy of the authors' procedure is tested using cross-validation techniques, which simulate past sampling conditions on the recent, well-sampled tropical Pacific SST and use the EOFs independent to the month being tested. The true sampling error of the cross-validated tests is computed with respect to the $1° \times 1°$ data for various sampling conditions. The theoretical sampling error is computed from the authors' derived formula and compared to the true error from the cross-validation tests. The authors' numerical results show that (i) the extrapolation method can sometimes improve the accuracy of the eigenvalues by 10%, (ii) the optimal averaging consistently yields smaller mse than the arithmetic averaging, and (iii) the theoretical formula for evaluating the OA error gives estimates that compare well with the true error.

1. Introduction

The spatial average of a climate field can be a useful index indicating the change of a climate state. The global average annual mean of the surface air temperature in the past 150 years is a well-known example of global climate change index. It is desirable to have an optimal averaging method that yields both accurate average and sampling error since the historical observations were often sparse. Using EOFs to take account of spatial inhomogeneity, Shen et al. (1994) developed such an optimal method for the average on the entire globe. They concluded that about 60 well-distributed stations can yield a global average annual mean surface air temperature with an error less than 10% compared with the natural variability. Because the spatial average of a climate quantity over a region is sometimes a more effective index compared to the global average for some

important climate phenomena, there is a need of an optimal regional averaging method that can also take account of spatial inhomogeneity. For instance the average of the eastern equatorial Pacific SST is significantly correlated with the ENSO index and hence a good indicator of the strength of an El Niño event (Peixoto and Oort 1992, 412–449).

The objective of this paper is to develop an optimal and systematic averaging procedure to spatially average the anomaly of a climate quantity in a region. To increase the accuracy level of the optimal averaging, an extrapolation method is introduced to refine the eigenvalues of the covariance matrix of the climate anomaly field. The major procedure here is an extension of that of Shen et al. (1994) from global averaging to regional averaging. The extension also includes (i) the use of the extrapolated eigenvalues, and (ii) the consideration of the random measurement error when estimating the total sampling error. The new method improves the accuracy of averaging and it is important for various kinds of climate change detection and climate analysis studies. The extrapolation method used to refine the computations of the eigenvalues is remarkably effective and also appears to be new in statistical climatology.

Corresponding author address: Samuel S. Shen, Department of Mathematical Sciences, University of Alberta, Edmonton, AB T6G 2G1, Canada.
E-mail: sam.shen@ualberta.ca

The method is tested using the optimal interpolation (OI) sea surface temperature (SST) for the tropical Pacific region (20°S, 20°N) × (155°E, 105°W) as the input data. The OI SST data are on a 1° × 1° grid (Reynolds and Smith 1994) from January 1982 to December 1995. Because of the inclusion of dense satellite observations, anchored to all available ship and buoy observations, the OI gives the most accurate SST available. Since one can never obtain the absolute truth of the SST data in a region, in this study, the OI SST is used to define the "truth" with quotation marks. The data used for the cross-validation tests are subsets of the full 1° × 1° grid. The data distribution of these subsets, that is, the gapped data, are determined by the historical sampling conditions. The mean square error (mse) of averaging the gapped data in the cross-validated tests is computed using the EOFs and eigenvalues obtained from different subsets of the OI data.

To outline our goals, let $T(\hat{\mathbf{r}})$ be a climate anomaly field over a region Ω, let $\mathcal{N}$ denote an observation network, and $\tilde{T}_j = \tilde{T}(\hat{\mathbf{r}}_j) = T(\hat{\mathbf{r}}_j) + E_j$, $j \in \mathcal{N}$ be samples (i.e., data), where E_j is the random measurement error (instrument error, reading error, and other random artificial influences), and $T(\hat{\mathbf{r}}_j)$ is the error-free value of the climate anomaly field at the point $\hat{\mathbf{r}}_i$. The systematic errors are assumed to have been removed from the raw data, and hence the remaining random error is uncorrelated with the anomaly field and with errors at other locations:

$$\langle E_i T \rangle = 0, \qquad \langle E_i E_j \rangle = 0 \quad \text{when} \quad i \neq j, \quad (1)$$

where $\langle \cdot \rangle$ denotes the ensemble average. Our two goals are (a) to find the best spatial average of the anomaly field T over Ω (section 2), and (b) to illustrate an extrapolation method that can improve the accuracy of eigenvalues (section 3). Computational results are presented in section 4 to demonstrate the usefulness of our OA procedure to climate studies. Discussions and conclusions are in section 5.

2. Optimal regional averaging and its mse

a. Mse sampling error

The average of the field T over a region Ω is

$$\overline{T} = \frac{1}{A} \int_\Omega T(\hat{\mathbf{r}}) \, d\Omega, \qquad (2)$$

where A is the area of the region Ω. Our objective is to use the sampling data $\tilde{T}(\hat{\mathbf{r}}_i)$ to estimate this quantity with the maximum accuracy. The method employed in this paper is referred to as optimal averaging and has been discussed by Kagan (1979), Vinnikov et al. (1990), Gandin (1993), Smith et al. (1994), and Shen et al. (1994). Here the version of OA similar to that of Shen et al. (1994) is followed, in which EOFs defined on the entire globe were computed for 40 yr of data (1950–89) using the series expansion of spherical harmonic functions. The method of Shen et al. (1994) is extended in this study from global averaging to regional averaging where spherical harmonics are no longer an orthonormal basis. Hence in this paper we do not use spherical harmonics but instead take an area factor into account when computing eigenvalues and EOFs. Namely, the EOFs are computed from area-weighted OI data for the period of January 1982–December 1995. The effect of random measurement errors is included in the total sampling error, as is done in Vinnikov et al. (1990) and Gandin (1993).

The linear estimator of the average, denoted by $\hat{\overline{T}}$, is

$$\hat{\overline{T}} = \sum_{j \in \mathcal{N}} w_j \tilde{T}_j, \qquad (3)$$

where $\mathcal{N}$ denotes the observation network on which the gapped data are distributed and the weights w_j satisfy a normalization condition:

$$\sum_{j \in \mathcal{N}} w_j = 1. \qquad (4)$$

This condition is needed because our data contain trends. So to guarantee that

$$\langle \hat{\overline{T}} \rangle = \sum_{j \in \mathcal{N}} w_j \langle \tilde{T}_j \rangle$$

when $\langle \tilde{T}_j \rangle = \langle \hat{\overline{T}} \rangle$, it is necessary to have the normalization condition (4) as discussed by Kagan (1979). This will be discussed further in section 4b.

The sampling error is measured by the mean square error

$$\epsilon^2 = \langle (\overline{T} - \hat{\overline{T}})^2 \rangle, \qquad (5)$$

and we define the covariance function by

$$\rho(\hat{\mathbf{r}}, \hat{\mathbf{r}}') = \langle T(\hat{\mathbf{r}}) T(\hat{\mathbf{r}}') \rangle. \qquad (6)$$

The following notations are adopted:

$$\rho_{ij} = \rho(\hat{\mathbf{r}}_i, \hat{\mathbf{r}}_j) = \langle T_i T_j \rangle \qquad (7)$$

and

$$\overline{\rho}_j = \frac{1}{A} \int_\Omega d\Omega \langle T(\hat{\mathbf{r}}) T_j \rangle. \qquad (8)$$

In terms of data $\tilde{T}_i$ and measurement errors $\langle E_i^2 \rangle$, the covariance matrix ρ_{ij} can be written as

$$\rho_{ij} = \langle T_i T_j \rangle = \langle \tilde{T}_i \tilde{T}_j \rangle + \langle E_i^2 \rangle \delta_{ij}, \qquad (9)$$

where δ_{ij} is the Kronecker delta that is equal to 1 when $i = j$ and 0 otherwise.

The mse can be written as

$$\begin{aligned} \epsilon^2 &= \langle \overline{T} \overline{T}' \rangle - 2 \langle \overline{T} \hat{\overline{T}} \rangle + \langle \hat{\overline{T}} \hat{\overline{T}}' \rangle \\ &= \frac{1}{A^2} \int_\Omega d\Omega \int_\Omega d\Omega' \, \rho(\hat{\mathbf{r}}, \hat{\mathbf{r}}') - 2 \sum_{j \in \mathcal{N}} w_j \overline{\rho}_j \\ &\quad + \sum_{i,j \in \mathcal{N}} w_i w_j \rho_{ij} + \sum_{i \in \mathcal{N}} w_i^2 \langle E_i^2 \rangle. \end{aligned} \qquad (10)$$

To minimize the mse, a Lagrange function is constructed:

$$\mathcal{L}(w_1, \cdots, w_N) = \epsilon^2(w_1, \cdots, w_N) + 2\Lambda\left(\sum_{j \in \mathcal{N}} w_j - 1\right),$$
(11)

where Λ is the Lagrange multiplier and N is the number of stations in the network $\mathcal{N}$. The partial derivatives

$$\frac{\partial \mathcal{L}}{\partial w_j} = 0, \qquad j = 1, 2, 3, \ldots, N,$$

and

$$\frac{\partial \mathcal{L}}{\partial \Lambda} = 0,$$

lead to

$$\sum_{j \in \mathcal{N}} \rho_{ij} w_j + \langle E_i^2 \rangle w_i + \Lambda = \overline{\rho}_i, \qquad i = 1, 2, \ldots, N; \quad (12)$$

$$\sum_{j \in \mathcal{N}} w_j = 1.$$
(13)

The solution of the above set of equations yields the optimal weights $w_1, w_2, \ldots, w_N$ for computing the optimal averaging by (3).

The covariance matrix (ρ_{ij}) can be approximated by

$$\rho_{ij} = \frac{1}{M_\gamma} \sum_{\gamma=1}^{M_\gamma} \tilde{T}_i(\gamma)\tilde{T}_j(\gamma) + \langle E_i^2 \rangle \delta_{ij}.$$
(14)

In this expression it is assumed that the time series $T(\hat{\mathbf{r}}_i, t)$ satisfies an ergodic process [the ensemble average $\langle T(\hat{\mathbf{r}}_i, t)T(\hat{\mathbf{r}}_j, t)\rangle$ is equal to the temporal average, which is approximated by the above summation with respect to the time variable γ]. The value M_γ is the maximal length of the data streams to be processed. It should be pointed out that $\tilde{T}_j(\gamma)$ may be serially correlated, but due to the short length of the data streams of the recent accurate observation, it is still the best approximation to estimate the covariance matrix (ρ_{ij}) by (14) rather than throwing out some data so that the remaining data is serially independent of each other. From (14) one can see that the rank of the computed covariance matrix (ρ_{ij}) is M_γ since M_γ is usually much less than the total number of OI grids. Thus the covariance matrix (ρ_{ij}) is often not a full rank matrix and this can cause some errors in computing the EOFs and variances as discussed in Zwiers and Shen (1997).

To solve (12) and (13), one needs to find $\overline{\rho}_i$, the average of the covariance function around the station $\hat{\mathbf{r}}_i$. Hence, the original problem of averaging T is partially converted into an averaging problem of the covariance function. This conversion, although mathematically straightforward, is important since it provides us a new way of evaluating sampling errors due to the fact that $\overline{\rho}_i$ can be computed from the averages of the EOFs. In this analysis we take advantage of the fact that the leading EOFs (i.e., characteristic climate patterns) are often stable for several decades and therefore can be estimated from more recent and relatively more accurate observations. They can also be estimated from climate models, such as GCMs (Zwiers and Shen 1997) or even energy balance models (Kim and North 1993). The algorithm for computing EOFs from the area-weighted data and that for computing $\overline{\rho}_i$ are described in the next section.

b. Computation of $\overline{\rho}_i$ and mse

Our method of computing $\overline{\rho}_i$ is illustrated using the monthly SST OI data from January 1982 to December 1995. Here these data are regarded as the truth. Anomalies of SST are computed with respect to the climatology of Reynolds and Smith (1995). The OI SST anomalies are gridded on a $1° \times 1°$ grid, denoted by OI, which is a very dense network for the monthly SST field. Let M_{OI} denote the length of the data stream (equal to 168 months). Then (14), with exclusion of the random error $\langle E_i^2 \rangle$, is used to compute the covariance matrix (ρ_{ij}^{OI}) from the OI data. The exact eigenvalue problem is

$$\int_\Omega \rho(\hat{\mathbf{r}}, \hat{\mathbf{r}}')\psi_k(\hat{\mathbf{r}}') \, d\Omega' = \lambda_k \psi_k(\hat{\mathbf{r}}).$$
(15)

Here $\psi_k(\hat{\mathbf{r}})$ is the kth EOF (or mode) and λ_k is the variance (eigenvalue) of $T(\hat{\mathbf{r}})$ on the kth mode ($k = 1, 2, \cdots$). The approximate eigenvalues of the above continuum eigenproblem can be estimated by a discretization procedure given by

$$\sum_{j \in OI} \hat{\rho}_{ij}\hat{v}_j^{(k)} = \hat{\lambda}_k \hat{v}_i^{(k)},$$
(16)

where

$$(\hat{\rho}_{ij}) = \left(\sqrt{A_i}\,\rho_{ij}^{OI}\sqrt{A_j}\right)$$
(17)

is the modified covariance matrix,

$$\hat{v}_j^{(k)} = \psi_k(\hat{\mathbf{r}}_j)\sqrt{A_j}$$
(18)

are the modified eigenvectors satisfying the normalization condition

$$\sum_{j \in OI} (\hat{v}_j^{(k)})^2 = 1,$$
(19)

and A_j is the area associated with the station $\hat{\mathbf{r}}_j$. For uniform latitude–longitude grid boxes, one has

$$A_j = \cos\phi_j \Delta\theta\Delta\phi,$$
(20)

where ϕ_j is the latitude of $\hat{\mathbf{r}}_j$ and the $\Delta\theta$ and $\Delta\phi$ are the zonal and meridional box dimensions, respectively, which are measured in radians. The linear spatial unit (i.e., the length unit) is in the radius of earth: $R = 6376$ km.

Since the eigenfunctions $\psi_n(\hat{\mathbf{r}})$ form an orthonormal functional basis, the covariance function can be expanded into an EOF form

$$\rho(\hat{\mathbf{r}}, \hat{\mathbf{r}}') = \sum_{n=1}^{\infty} \lambda_n \psi_n(\hat{\mathbf{r}})\psi_n(\hat{\mathbf{r}}').$$
(21)

The EOF representations of $\rho(\hat{\mathbf{r}}, \hat{\mathbf{r}}_i)$ and $\overline{\rho}_i$ are, respectively,

$$\rho(\hat{\mathbf{r}}, \hat{\mathbf{r}}_i) = \sum_{n=1}^{\infty} \lambda_n \psi_n(\hat{\mathbf{r}}) \psi_n(\hat{\mathbf{r}}_i) \qquad (22)$$

and

$$\overline{\rho}_i = \sum_{n=1}^{\infty} \lambda_n \psi_n(\hat{\mathbf{r}}_i) \overline{\psi}_n, \qquad (23)$$

where $\overline{\psi}_n$ is defined as

$$\overline{\psi}_n = \frac{1}{A} \int_{\Omega} \psi_n(\hat{\mathbf{r}}) \, d\Omega, \qquad (24)$$

which is the average of the eigenfunction $\psi_n(\hat{\mathbf{r}})$. In practice one has to compute an approximate value of this by numerical integration:

$$\overline{\psi}_n \approx \frac{1}{A} \sum_{j \in OI} \psi_n(\hat{\mathbf{r}}_j) A_j \approx \sum_{j \in OI} \hat{v}_j^{(n)} \frac{\sqrt{A_j}}{A}. \qquad (25)$$

In summary, to compute $\overline{\rho}_i$ we

1) compute the covariance matrix $[\rho_{ij}^{OI}]$ according to (14) (excluding the term $\langle E_i^2 \rangle \delta_{ij}$) and the modified covariance matrix $\hat{\rho}_{ij}$ according to (17),
2) solve the eigenvalue problem for the modified covariance matrix $\hat{\rho}_{ij}$ to obtain eigenvalues $\hat{\lambda}_k$ and normalized eigenvectors $\hat{v}_i^{(k)}$, and
3) use (25) to compute $\overline{\psi}_n$ and (18) to compute $\psi_n(\hat{\mathbf{r}}_i)$, and finally compute $\overline{\rho}_i$ by

$$\overline{\rho}_i \approx \sum_{n=1}^{M_c} \hat{\lambda}_n \psi_n(\hat{\mathbf{r}}_i) \overline{\psi}_n. \qquad (26)$$

The quantities $\hat{\rho}_{ij}$ and $\overline{\rho}_i$ will be used in (12), which, together with (13), determines the optimal weights w_1, ..., w_N for averaging. The eigenvalues $\hat{\lambda}_n$, eigenvectors $\psi_n(\hat{\mathbf{r}}_i)$, and their averages $\overline{\psi}_n$ will be used to calculate the total sampling error given by (27) below. The sum in (26) above and (27) below for n in practice runs through a relatively small number of modes M_c, say 20, since the higher modes are contaminated by noise and the inclusion of these modes may increase error. As discussed by Kagan (1979) it is important to avoid adding more detail to the covariance function than can be justified by the amount of data available to compute them. However, this problem is lessened by using EOFs and the mse formula [(27)] below since each mode is scaled by its eigenvalue. This forces the first few, most important modes, to dominate. In practical computations, one may choose the cutoff mode number M_c according to the criterion of 80%–95% variances explained (by the first M_c modes).

By (10) and (21), the final expression of mse is obtained in terms of M_c EOF modes:

$$\epsilon^2 \approx \sum_{n=1}^{M_c} \hat{\lambda}_n \left[\overline{\psi}_n - \sum_{j \in \mathcal{N}} w_j \psi_n(\hat{\mathbf{r}}_j) \right]^2 + \sum_{i \in \mathcal{N}} w_i^2 \langle E_i^2 \rangle. \quad (27)$$

Since this formula includes the EOF patterns, if the observations are along the node lines of an EOF (where the EOF is equal to zero) or in the fine spatial structure area of an EOF, the sampling error is large for the corresponding mode. Thus this formula is also useful for future observation network design.

Second, the sampling error formula (27) implies that the mse is linearly proportional to the eigenvalues and is in a square relationship with the numerical integration errors of the eigenfunctions. This is the mathematical basis of many researchers' opinion that to estimate the mse of an OA it is crucial to obtain highly accurate eigenvalues, and the exact shapes of the eigenfunctions do not matter as much. Therefore it is desirable to compute the eigenvalues as accurately as possible.

3. Extrapolation method for eigenvalue refinement

The purpose of the extrapolation method presented here is to increase the accuracy of the eigenvalue estimations. For instance, in the test described in section 4, using this method one can use a $5° \times 5°$ network to get eigenvalues of about the same accuracy as those obtained from a $2° \times 2°$ network. This is an important consequence since for many climatological datasets, it is only possible to interpolate the data onto a coarse network.

We use h to denote the grid box size (in radians). It is assumed that if one uses a network of size h to estimate the eigenvalues, the accuracy is of second order. Namely,

$$\lambda - \hat{\lambda} = Ch^2 + O(h^4), \qquad (28)$$

where λ is the true eigenvalue, $\hat{\lambda}$ is the one estimated from the network, and C is a constant of order one. This assumption, though not proved mathematically, is true in many practical cases. It is numerically verified for the OI SST data in a rectangular region (20°S, 20°N) $\times$ (155°E, 105°W).

In typical practice, grid sizes are usually smaller than tens of degrees; in terms of radians, the grid size $h = $ several degrees $\times \pi/180$ is a small number (less than 0.3). The square of it is less than 0.1, which is the order estimation (second-order accuracy) of the error for computing the eigenvalues. Of course one would desire to have higher-order accuracy, say, $O(h^4)$ where $h^4 < 0.01$. The purpose of the extrapolation method is to raise the order of accuracy. The idea is to optimally and linearly combine the eigenvalues computed from different networks. Suppose that h_1 and h_2 are the sizes of two different networks. The assumption of second-order accuracy is

$$\lambda - \hat{\lambda}^{(1)} = Ch_1^2 + O(h_1^4), \qquad (29)$$

$$\lambda - \hat{\lambda}^{(2)} = Ch_2^2 + O(h_2^4), \qquad (30)$$

where the superscripts (1) and (2) signify the networks.

TABLE 1. Eigenvalues ($\times$ 1000) of the tropical Pacific SST directly computed from networks (first 10 eigenvalues and four different network resolutions.

Grid reso-lution	Mode									
	1	2	3	4	5	6	7	8	9	10
1 $\times$ 1	270.2	56.3	20.5	15.3	12.9	11.3	8.8	7.1	6.8	4.8
2 $\times$ 2	268.5	56.0	20.1	15.1	12.7	11.2	8.7	7.0	6.7	4.7
5 $\times$ 5	255.7	54.3	17.9	14.0	11.5	10.5	8.0	6.6	6.1	4.3
10 $\times$ 10	217.9	49.4	14.7	11.2	9.5	8.7	5.8	5.7	5.1	3.6

The linear combination of the two estimated eigenvalues is

$$\hat{\lambda}^{(1,2)} = w_{1,2}\hat{\lambda}^{(1)} + w_{2,1}\hat{\lambda}^{(2)}, \tag{31}$$

where

$$w_{1,2} + w_{2,1} = 1. \tag{32}$$

Multiplying (29) by $w_{1,2}$ and (30) by $w_{2,1}$ and adding the two resulting equations together, we get

$$\lambda - \hat{\lambda}^{(1,2)} = C(w_{1,2}h_1^2 + w_{2,1}h_2^2) + O(h_2^4).$$

Without loss of generality it has been regarded that $h_1 \leq h_2$ in the above expression. To force the fourth-order accuracy, one must have

$$w_{1,2}h_1^2 + w_{2,1}h_2^2 = 0. \tag{33}$$

This equation and the normalization equation (32) lead to the optimal weights:

$$w_{1,2} = \frac{h_2^2}{h_2^2 - h_1^2} \quad \text{and} \quad w_{2,1} = \frac{h_1^2}{h_1^2 - h_2^2}. \tag{34}$$

From these equations one can see that one of the weights must be negative. This is the reason why the method is called extrapolation in contrast to interpolation (for which both weights are positive). If $h_2 = 2h_1$, then

$$w_{1,2} = \frac{4}{3}, \qquad w_{2,1} = -\frac{1}{3}. \tag{35}$$

In general the more dense network renders more accurate eigenvalues. Hence if $h_2 \geq h_1$, $\hat{\lambda}^{(2)}$ helps $\hat{\lambda}^{(1)}$ with a correction equal to

$$\lambda^{CC} = w_{2,1}(\hat{\lambda}^{(2)} - \hat{\lambda}^{(1)}), \tag{36}$$

and $w_{2,1}$ is negative. If the network *consistently* underestimates the eigenvalues, then $\hat{\lambda}^{(2)} \leq \hat{\lambda}^{(1)}$ and the correction quantity λ^{CC} is positive. Otherwise, if the network *consistently* overestimates the eigenvalues, then $\hat{\lambda}^{(2)} \geq \hat{\lambda}^{(1)}$ and the correction quantity λ^{CC} is negative. For the SST field considered in this paper, networks consistently underestimate the true eigenvalues and $\hat{\lambda}^{CC} > 0$. See Table 1.

In the above, the word consistently is emphasized because only under this condition can our extrapolation method be applied. Table 1 shows that the coarser grid consistently yields lower eigenvalues and hence our extrapolation method can be safely applied. If one network

underestimates the eigenvalues and the other makes overestimates, then one would expect a more complex extrapolation method than the one presented here. A nonsystematic situation may be a consequence of a highly inhomogeneous field with a coarse observation network, since some important finer structures cannot be captured by a coarse network, and hence there is inconsistency of over- and underestimations.

Another point is how to choose h_1 and h_2. When h_1 and h_2 are too close to each other, $|w_{1,2}|$ and $|w_{2,1}|$ are very large and lead to computational instability. Therefore the correction is sometimes highly unreliable. If $h_1 \ll h_2$, then $w_{2,1}/w_{1,2}$ is too small and hence the correction amount λ^{CC} is too small to be effective. Our experiments suggest a range of

$$h_2/h_1 = 2 \pm 0.5. \tag{37}$$

Similarly one can derive the sixth-order accuracy combination when combining the results of the fourth-order accuracy and requiring at least three networks. The higher-order extrapolation makes a finer tuning of the eigenvalues. The correction amount λ^{CC} is usually small for this finer tuning, and sometimes it is too small to be noticeable. Our experiments on the SST data in the region (20°S, 20°N) $\times$ (155°E, 105°W) show that the fourth-order extrapolation renders no significant improvement to the eigenvalues, while the second-order extrapolation does reduce the errors of the eigenvalues up to 10% for some modes.

4. Computational results

This section describes our computational results on eigenvalues, eigenfunctions, optimal averaging, and its mse. First, we present the results of the eigenvalues and eigenfunctions, which give a rough estimation of the variance of the spatially averaged SST anomaly.

a. Total variances and eigenvalues

The area-weighted total variance $\langle \int_\Omega T^2(\hat{\mathbf{r}}) \, d\Omega \rangle$ can be used as another verification for the eigenvalue computation and a reference quantity to determine the number of modes used in sampling error calculations. This variance $\langle \int_\Omega T^2(\hat{\mathbf{r}}) \, d\Omega \rangle$ can be computed directly from numerical integration:

$$\left\langle \int_\Omega T^2(\hat{\mathbf{r}}) \, d\Omega \right\rangle \approx \frac{1}{168} \sum_{\gamma=1}^{168} \sum_{j \epsilon OI} T^2(\hat{\mathbf{r}}_j, \gamma) \cos\phi_j (\pi/180)^2$$

$$= 0.473[°C]^2 R^2, \tag{38}$$

where γ indicates the time variable in the unit of month, ϕ_j is the latitude of the grid point $\hat{\mathbf{r}}_j$, and $\cos\phi_j(\pi/180)^2$ is the area of the jth 1° $\times$ 1° grid box.

Eigenvalues were computed for the SST anomalies [with respect to the Reynolds and Smith (1995) climatology] over the tropical Pacific region (20°S, 20°N) $\times$ (155°E, 105°W) using monthly data from January

TABLE 2. Extrapolated eigenvalues ($\times$ 1000) for the tropical Pacific SST where the grid labeled $r° \times r°$ is the combination of the $r° \times r°$ and $2r° \times 2r°$ grids.

Grid reso- lution	Mode									
	1	2	3	4	5	6	7	8	9	10
1 $\times$ 1	270.7	56.3	20.6	15.3	13.0	11.4	8.9	7.1	6.8	4.8
2 $\times$ 2	270.9	56.3	20.6	15.3	13.0	11.4	8.8	7.1	6.8	4.8
5 $\times$ 5	268.3	56.0	18.9	14.9	12.2	11.1	8.7	6.9	6.5	4.5

1982 to December 1995. Table 1 shows the eigenvalues computed from different grid sizes and Table 2 shows the extrapolated eigenvalues. Table 1 clearly demonstrates that the more sparse grid consistently yields lower estimation of the eigenvalues and hence the extrapolation method described in section 3 can be safely applied. The combination of the $1° \times 1°$ and $2° \times 2°$ grids yields the most accurate eigenvalues one can obtain and that are thus regarded as the truth. The summation of the first 50 of these eigenvalues is

$$\sum_{n=1}^{50} \hat{\lambda}_n = 0.464.$$

Because of

$$\langle T^2(\hat{\mathbf{r}}) \rangle = \sum_{n=1}^{\infty} \lambda_n \psi_n^2(\hat{\mathbf{r}}),$$

and the normalization condition for $\psi_n^2(\hat{\mathbf{r}})$, one has

$$A\langle T_{(m)}^2 \rangle = \left\langle \int_\Omega T^2(\hat{\mathbf{r}})\, d\Omega \right\rangle = \sum_{n=1}^{\infty} \lambda_n \approx \sum_{n=1}^{50} \lambda_n$$

$$= 0.464\ [°C]^2 R^2,$$

where R is the radius of earth and $\langle T_{(m)}^2 \rangle$ is the mean value of the SST variance. The total variance $0.464[°C]^2 R^2$ computed from the summation of the eigenvalues is in a good agreement with $0.473[°C]^2 R^2$ computed from the direct numerical integration shown in (38), and this is an independent verification for the eigenvalue computation. The area of Ω is $A = 1.194\ R^2$. Hence

$$\langle T_{(m)}^2 \rangle \approx \frac{0.464}{1.194} = 0.389[°C]^2.$$

The standard deviation is

$$\sqrt{\langle T_{(m)}^2 \rangle} = 0.624[°C]. \tag{39}$$

The first 50 extrapolated eigenvalues from the $1° \times 1°$ grid explains 98% (=0.464/0.473) of the total variance, while the first 20 eigenvalues already explains 93% (=0.442/0.473). Considering that the high modes contains much noise, it was determined to take the first 20 modes in our sampling error estimation, that is, $M_c = 20$ in (26), (27), and the pertinent equations in the rest of this paper.
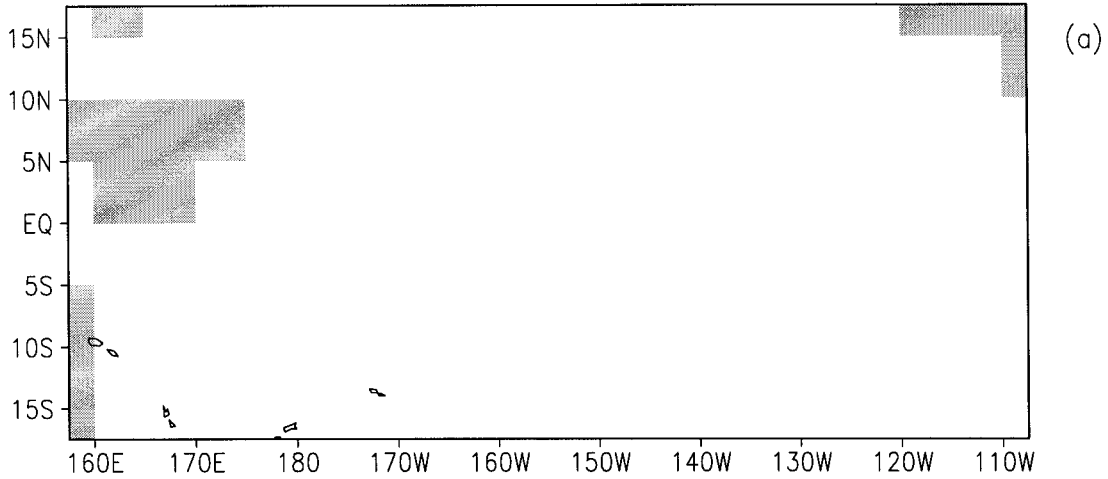
b. Average temperature

The strong correlation between the SST and the evaporation–precipitation in the tropical Pacific requires accurate assessment of various types of SST characteristics, one of which is the spatial average. A possible strong correlation between the El Niñõ and the tropical Pacific average SST motivates us in the present careful study of optimal averaging and its error estimation (Peixoto and Oort 1992, 412–449). The optimal averaging method developed here makes use of the most accurate estimation of variances by an extrapolation and the EOFs derived from the recently available OI data of Smith et al. (1996). Our results include not only the spatial average of the historical data from 1856 to 1995 but also the statistical sampling error. The sampling error is computed from both the theoretical formula (27) and cross validation for various 14-yr periods.

The historical sampling conditions, that is, the number and locations of the sampling points, are considered for six 14-yr periods in the tropical Pacific region (20°S, 20°N) $\times$ (155°E, 105°W). The sampling locations are at the center of each 5° $\times$ 5° box with the following four boundary sampling points: 17.5°S, 157.5°E; 17.5°N, 157.5°E; 17.5°S, 107.5°W; and 17.5°N, 107.5°W. Thus the maximal number of sampling points is 20 $\times$ 8 = 160. See Figs. 1 and 2 for illustrations. The sampled data are the 14-yr OI data between January 1982 and December 1995, for the SST characteristics of the OI data are known and hence can be used to examine the goodness of the the historical sampling conditions by estimating the sampling errors. This test of using the historical sampling condition on the more recent data is in the category of cross validation. The six historical periods tested are 1882–95, 1912–25, 1932–45, 1952–65, 1972–85, and 1982–95. Data used to define the locations of the OI data for each cross-validation test network are from an updated version of the atlas of Bottomley et al. (1990), provided by the U.K. Meteorological Office. These data are on a 5° $\times$ 5° grid with special values to mark where there are no data for a certain month. The sampling conditions for October 1938 and January 1979 are shown in Figs. 1a and 1b, respectively. The total number of sampling points as a function of time from 1865 to 1995 is shown in Fig. 2.

Cross validation employs a separate set of EOFs for each month averaged, as well as a sparse grid from historical sampling. The EOFs for each month are computed using all monthly data except for the averaging month and eight months on either side of it. This procedure gives a set of EOFs for each month that are approximately independent of the month, to better simulate the accuracy that one may expect in historical periods. The cross validation is explained in more detail, in a different context, in Smith et al. (1996).

The error associated with each cross-validation period is estimated. Since the OI anomalies are regarded as the

Oct 1938 Sampling

(a)

Jan 1979 Sampling

(b)

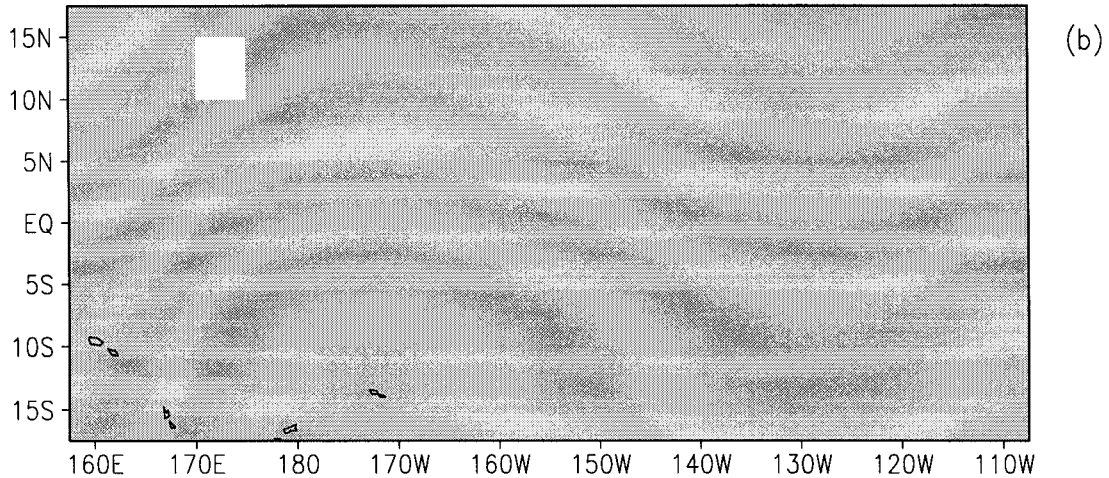FIG. 1. Observation data: (a) October 1938 and (b) January 1979. Gray shadings indicate grids with data.

truth, one can use the $1° \times 1°$ OI to directly compute the true average and the true error of each cross-validation estimate. The cross-validation process here consists of (a) using a set of EOFs computed excluding the month in question and 8 months on either side of the month, for a total of excluded 17 months, and (b) reducing the grid of the OI data to the observational networks of several historical periods and averaging the anomaly field using the reduced grid and cross-validated EOFs for each month.

The averaged OA root-mean-square error for cross validation tests is defined by

$$\text{rmse}_{\text{OA}} = \left\{ \frac{1}{M} \sum_{m=1}^{M} [\text{OA}(m) - \overline{T}_{\text{OI}}(m)]^2 \right\}^{1/2}, \quad (40)$$

where $\text{OA}(m)$ is the optimal average of the SST anomaly computed by the scheme described in section 2 using the OI data at the historical sampling points, $\overline{T}_{\text{OI}}(m)$ is the area-weighted average computed from the $1° \times 1°$ OI, and m is the month index, from 1 to $M = 152$ (i.e., from September 1982 to April 1995) because we have to exclude the eight months from the beginning and the end of the data stream to fit our cross-validation process mentioned above. The quantity, rmse_{OA} is called the true

Number of 5-D Obs
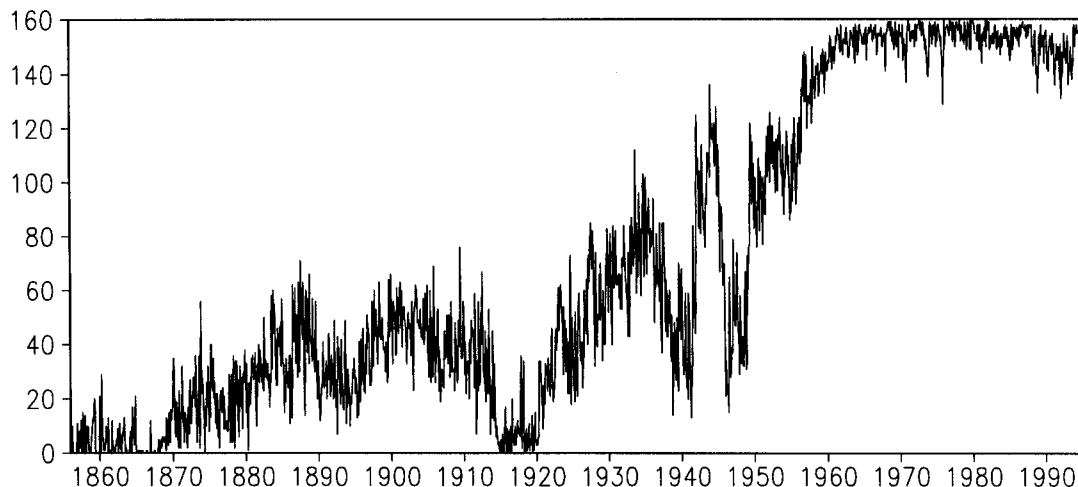

FIG. 2. The total number of observations as a function of time t (month) for our $5° \times 5°$ network of Pacific SST anomalies (20°S–20°N, 155°E–105°W).

OA error since it comes from the comparison with the OI SST field, which is regarded as the truth as discussed in section 1. For every cross-validation period of 14 yr, one can use (40) to compute $rmse_{OA}$, the "True Error: OA," and put the value in the upper-right corner of Figs. 3a–d. The error $rmse_{OA}$ may also be estimated as a function of time by fixing the historical observation network, say that of January 1919, throughout the period. This would yield the $rmse_{OA}$ for the observation network of January 1919. However, we still choose to compute (40) since the observation network generally do not change rapidly and our purpose in computing $rmse_{OA}$ is simply to show that the theoretical error is comparable to the true error.

The true $rmse_{AA}$ for the arithmetic averaging (AA) is computed in a similar way. One only needs to replace $OA(m)$ in (40) by the arithmetic average $AA(m)$ of the SST anomaly. It is similar to the OA case that for every cross-validation period of 14 yr, one can use (40) to compute $rmse_{AA}$, the "True Error: AA," and put the value in the upper-right corner of Figs. 3a–d.

Theoretical OA rsme is computed from the square root of (27), in which $\sqrt{\langle E_i^2 \rangle} = 0.3°C$ (Reynolds and Smith 1994; Parker et al. 1994). An estimate of the theoretical error from AA rmse is obtained by replacing the optimal weights w_i by the uniform weights $1/N$ in the OA error estimation (27). This theoretical rmse is computed for every month in the cross-validation periods indicated on the frame titles of Fig. 3. The numerical results for the rmse curves are shown in Fig. 3. It can be seen that OA rmse is systematically smaller than AA rmse, as one would expect. The difference becomes smaller as the sampling becomes more dense. For the full $5° \times 5°$ grid, the two are the same.

Comparing the true OA $rmse_{OA}$ with the theoretical

OA rmse and AA rmse shows the advantage of using the optimal weights (Fig. 3). The number of sampling points as a function of time is shown below the rmse curves. From Fig. 3 one clearly sees that, compared to the uniform weights, the optimal weights effectively reduce the sampling error when the sampling is poor. When the sampling is dense, that is, Fig. 3d, it makes little difference whether one uses the optimal average or arithmetic average. One can also see that when the sampling is sparse, that is, Fig. 3b, the theoretical OA rmse is about the same as the true OA $rmse_{OA}$, while the theoretical OA rmse slightly overestimates the rmse when the sampling is dense. In the latter case the sampling error is small. Therefore, we conclude that our theoretical sampling formula (27) gives a reasonable assessment of sampling errors.

The anomaly as a function of time is shown in Fig. 4. The thick solid line is the area-weighted average from the $1° \times 1°$ OI. The thin solid line is the optimal average using the historical sampling conditions between 1952 and 1965: $\hat{\tilde{T}}(t) = \Sigma_{j\in\mathcal{N}} w_j \tilde{T}_j(t)$. There is little difference between the two lines. The theoretical error bounds computed from (27) are shown by short dashed lines: $\hat{\tilde{T}}(t) \pm 3 \times rmse$. The spatially averaged anomaly by OA from 1856 to 1995 is shown in Fig. 5, with theoretical error bounds $3 \times rmse$.

Another point concerning the OA is the necessity of normalizing weights by (13). To justify this, we computed the true $rsme_{OA}$ by (40) with and without the normalization of weights and compared the results. The comparison shows that the $rmse_{OA}$ without normalization of weights is sometimes larger. For example, for the historical networks in 1932–45, the true $rmse_{OA}$ with normalized weights is 0.0775 and the true $rmse_{OA}$ without the normalization of weights is 0.1009. This implies
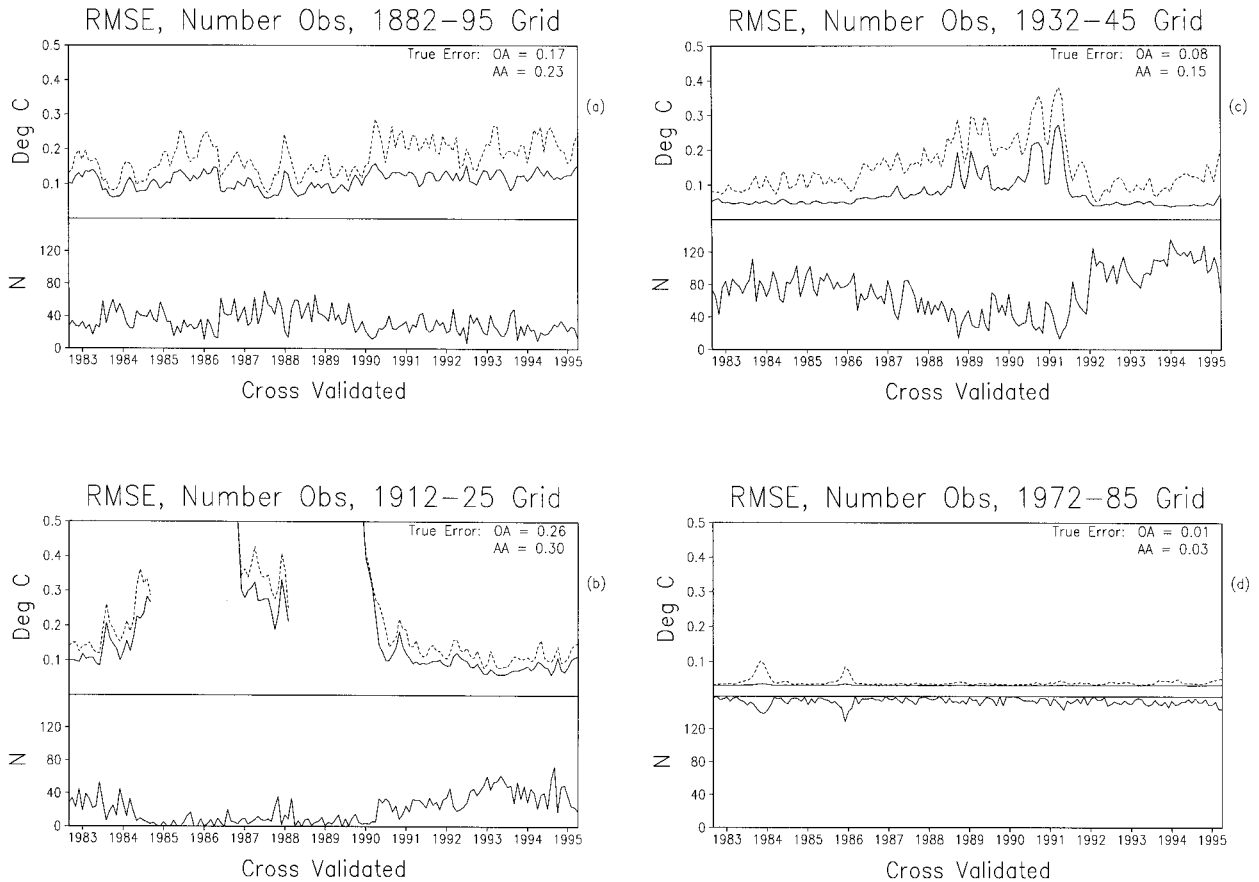
FIG. 3. The theoretical rmse of OA (solid line) and theoretical rmse of arithmetic average (AA, dashed line) for samplings in different periods of time: (a) 1882–95, (b) 1912–25, (c) 1932–45, and (d) 1972–85. Also shown is the number of sampling points N for each reduced grid (max = 160 observations). The ''True Error: OA'' and ''True Error: AA'' were computed from (40).

that the optimal averaging without the normalization of weights has significantly distorted the trend signal of the true average.

5. Conclusions and discussions

We have described an optimal averaging method with error estimates, eigenvalue extrapolation, and a test using tropical Pacific SST data. The optimal averaging method is an improvement over that of Shen et al. (1994), using the EOFs to treat the spatial inhomogeneities. The improvements include (i) a more accurate computation of eigenvalues of the covariance matrix by an extrapolation method, (ii) an extension of the OA on the entire globe to the OA over a part of the earth's surface, and (iii) the inclusion of random observation errors.

The tropical OI monthly SST data were used to compute the EOFs and test the reliability of the improved methods of averaging and gridding. For the OA, when there are more than 30 observations, the OA results agree with the true average with an error less than about 0.15°C. The mse computed from (27) gives a good es-timate of the sampling error although it overestimates the error slightly when sampling is dense.

It is concluded that the regional averaging procedure developed in this paper is reliable and accurate, and takes into account the spatial inhomogeneity. However, our method cannot deal with nonstationarity even though the normalization of the weights in the OA procedure minimizes the distortion of the trends in the data. Further development of the optimal methods may consider the nonstationarity of data (Kim et al. 1996) and application of these methods to other fields.

Finally we wish to discuss the applicability of the method. First, the extrapolation method to refine the eigenvalues can be used if a given network *consistently* yields an underestimate or overestimate. This can be easily tested by comparing the eigenvalues computed from the networks of different observation densities like Table 1. Second, the averaging method can be applied to different climate quantities other than SST anomalies, such as precipitation and 500-hPa height. The crucial part in applying the method to various climate quantities is the EOF properties including both variances and EOF patterns. For instance, for precipitation the variances are
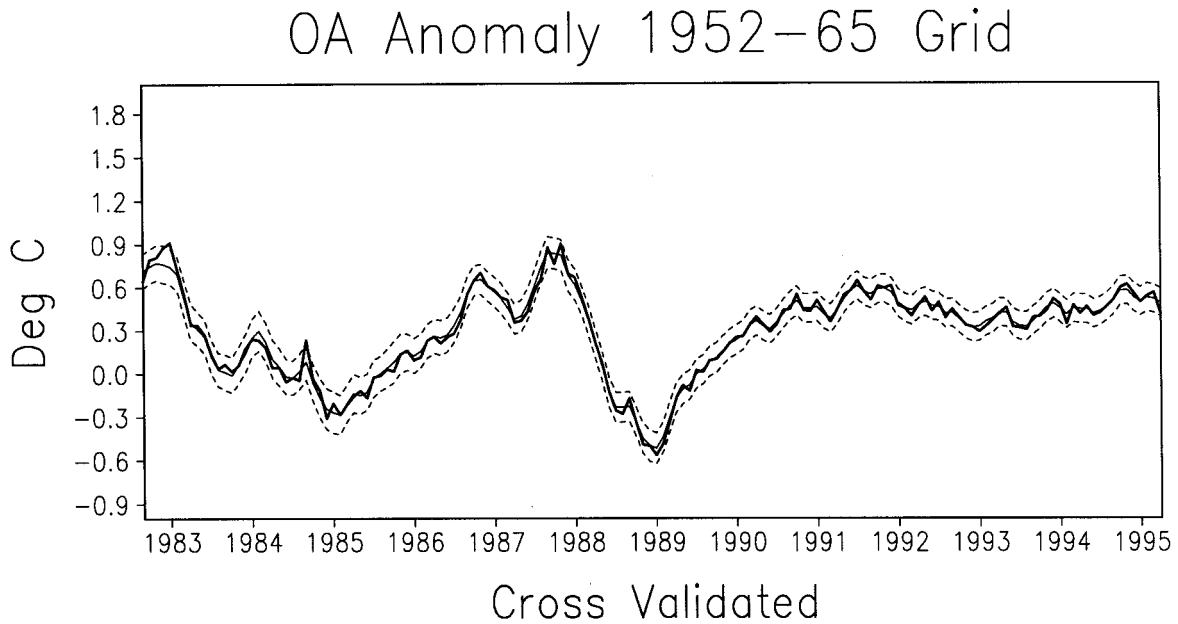
OA Anomaly 1952−65 Grid

FIG. 4. The spatially averaged tropical Pacific monthly anomaly from September 1982 to April 1995 and the error estimates. The thick solid curve is the area-weighted average of the full grid OI data. The thin solid curve is the cross-validated OA result, which was computed from the data on the historical observation networks from September 1952 to April 1965. The dotted lines are the 3 σ error bounds centered around the thin solid line.
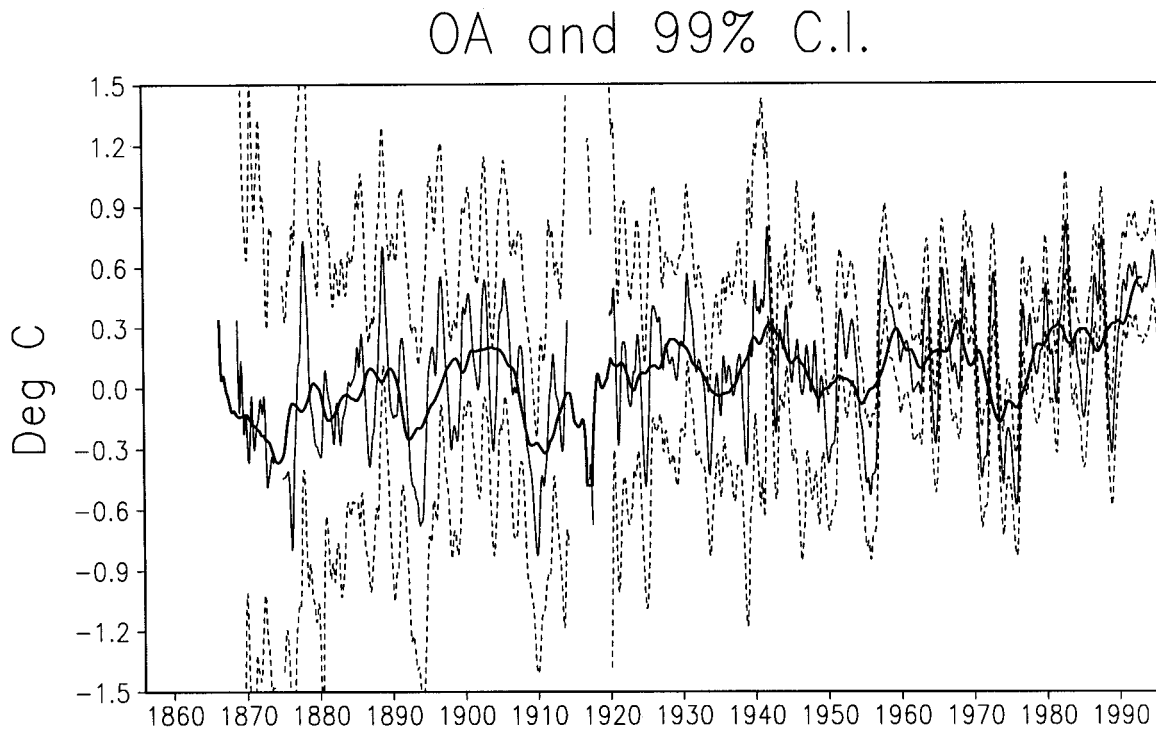
OA and 99% C.I.

FIG. 5. Optimally averaged tropical Pacific monthly SST anomaly from 1856 to 1995 (solid) and its 3 σ error bound (dashed). A 15-point binomial smoother has been applied to the curves. Also shown is the 5-yr running average of the OA (thick curve).

usually very large and the EOF patterns are relatively more complex with respect to SST, and hence it reflects the facts of shorter length scales. Consequently a more dense observation network is needed. Third, the method can also be applied to different regions of the globe. In certain regions, a climate field might be highly inhomogeneous and the condition for extrapolation is not satisfied, but one can still use the optimal averaging formulas (3), (12), (13), and the minimum error formula (27) without refining the eigenvalues, for the refinement of eigenvalues further increases the accuracy of the averaging without increment of observation data, but it is not a necessary step in the optimal averaging procedure.

REFERENCES

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

Gandin, L. S., 1993: Optimal averaging of meteorological fields. National Meteorological Center Office Note 397, 67 pp. [Available from National Center for Environmental Modeling/NWS/NOAA, Washington, DC 20233.]

Kagan, R. L., 1979: *Averaging Meteorological Fields* (in Russian). Gidrometeoizdat, 212 pp. [English translation available from National Center for Environmental Modeling, National Weather Service, NOAA, Washington, DC 20233.]

Kim, K.-Y., and G. R. North, 1993: EOF analysis of surface temperature field in a stochastic climate mode. *J. Climate,* **6,** 1681–1690.

——, ——, and J. Huang, 1996: EOFs of one-dimensional cyclostationary time series: Computations, examples, and stochastic modeling. *J. Atmos. Sci.,* **53,** 1007–1017.

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.

Peixoto, J., and A. H. Oort, 1992: *Physics of Climate.* American Institute of Physics, 520 pp.

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.

Shen, S. S., G. R. North, and K.-Y. Kim, 1994: Spectral approach to optimal estimation of the global average temperature. *J. Climate,* **7,** 1999–2007.

Smith, T. M., R. W. Reynolds, and C. F. Ropelewski, 1994: Optimal averaging of seasonal sea surface temperatures and associated confidence interval (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.

Vinnikov, K. Ya., P. Ya. Groisman, and K. M. Lugina, 1990: Empirical data on contemporary global climate changes (temperature and precipitation). *J. Climate,* **3,** 662–677.

Zwiers, F. W., and S. S. Shen, 1997: Errors in estimating spherical harmonic coefficients from partially sampled GCM output. *Climate Dyn.,* **13,** 703–716.