Singular Value Decomposition of Wintertime Sea Surface Temperature and 500-mb Height Anomalies

JOHN M. WALLACE, CATHERINE SMITH, AND CHRISTOPHER S. BRETHERTON

Department of Atmospheric Sciences, University of Washington, Seattle, Washington

(Manuscript received 23 July 1990, in final form 28 May 1991)

ABSTRACT

Single field principal component analysis (PCA), direct singular value decomposition (SVD), canonical correlation analysis (CCA), and combined principal component analysis (CPCA) of two fields are applied to a 39-winter dataset consisting of normalized seasonal mean sea surface temperature anomalies over the North Pacific and concurrent 500-mb height anomalies over the same region. The CCA solutions are obtained by linear transformations of the SVD solutions. Spatial patterns and various measures of the variances and covariances explained by the modes derived from the different types of expansions are compared, with emphasis on the relative merits of SVD versus CCA. Results for two different analysis domains (i.e., the Pacific sector versus a full hemispheric domain for the 500-mb height field) are also compared in order to assess the domain dependence of the two techniques. The SVD solution is also compared with the results of 28 Monte Carlo simulations in which the temporal order of the SST grids is randomized and found to be highly significant.

As expected, the leading SVD modes explain substantially more of the squared covariance between the two fields than any of the CCA modes, while the paired expansion coefficients of the leading CCA modes are more strongly correlated than any of the SVD modes. The expansion coefficient for the leading SVD mode is almost identical to the leading principal component of the SST field, regardless of whether the 500-mb height field is hemispheric or restricted to the Pacific sector. SST patterns strongly resembling the second and third EOFs are also recovered among the three leading SVD modes.

The leading CCA mode in the expansion based on the three leading singular value vectors for the Pacific sector resembles the pattern of anomalies observed in association with ENSO. The other modes more closely resemble the patterns derived from PCA of the 500-mb height field than those for the SVD modes on which they are based. The SVD and CPCA solutions for the first three modes proved to be quite similar.

The SVD and CCA solutions based on the hemispheric 500-mb height field are indicative of a coupling between the interannual variability of North Pacific and North Atlantic SST by virtue of their mutual relationship to one of the atmosphere's most prominent planetary wave patterns.

1. Introduction

In the companion paper (Bretherton et al. 1992; hereafter denoted BSW), we compared direct singular value decomposition (SVD) with the more commonly used analysis techniques [i.e., canonical correlation analysis (CCA), principal component analysis (PCA) of the individual fields, and combined principal component analysis of the two fields (CPCA)] for isolating spatial patterns in two fields that tend to occur synchronous with one another. In this paper we will compare the performance of these methods in documenting the relationships between the seasonal mean, wintertime sea surface temperature (SST) and 500-mb height (Z_{500}) fields over the North Pacific. As in many climate applications of these techniques, we will be dealing with fields in which the number of spatial degrees of freedom is far less than the number of grid points used to represent them and with a relatively small number of realizations in the time domain.

For illustrating the applications of these techniques, we have purposely chosen a familiar problem for which substantial literature exists based on a variety of analysis techniques, periods of record, and choices of analysis domain. Relationships between SST anomalies and anomalies in the atmospheric geopotential height field over the North Pacific have been investigated by Namias (1973, 1975, 1978), Kawamura (1984), Lanzante (1984), Iwasaka et al. (1987), Wallace and Jiang (1987), Namias et al. (1988), and Wallace et al. (1990), among others. The dominant correlation patterns in the 500-(or 700-) mb height field (in isolation) have been documented by Wallace and Gutzler (1981), Namias (1981), Horel (1981), Esbensen (1984), Barnston and Livezey (1987), and Kushnir and Wallace (1989), and the structure of the North Pacific SST field has been documented by Namias (1972), Davis (1976), and Weare et al. (1976). We will have occasion to refer to spatial patterns identified in some of these studies at several points in the paper.

Corresponding author address: John M. Wallace, University of Washington, Department of Atmospheric Sciences, AK-40, Seattle, WA 98195.

2. Data and preprocessing

The SST data used in this study comes from the Comprehensive Ocean–Atmosphere Data Set (COADS) (Fletcher et al. 1983) for the interval December 1946–February 1985, which consists of monthly anomalies of SST on a 2° × 2° grid. We have aggregated these data into seasonal averages for 4° lat by 10° long grid boxes. The 500-mb height data are based on NMC operational analyses, archived on the NMC 47 × 51 octagonal grid. For the calculations, a coarse (23 × 23 point) grid was used, but for display purposes, the full resolution was used. The domain of the analysis described in the next four sections and the location of the grid points is shown in Fig. 1. Note that the 157-point SST grid extends equatorward as far as 10°N; not quite far enough to sample the large interannual variability associated with the El Niño–Southern Oscillation phenomenon. The regional 500-mb height grid is comprised of 125 points. For both SST and 500-mb height time series at each grid point, standardized anomalies were computed for each of the 39-winter (December–February) seasons in the record.

In computing the covariance matrices referred to in the remainder of this paper, we performed the calculations with and without weighting SST at each grid point by the square root of the cosine of latitude as suggested by North et al. (1982) and found that it made very little difference. Results proved to be somewhat more sensitive to whether the time series used to compute the temporal covariance matrix were normalized. In the interests of simplicity in keeping track of the fractions of the variances and covariances explained by the various modes, we have elected to show the results based on unweighted, normalized variances and covariances (i.e., the off-diagonal elements of C_{SZ} are correlation coefficients).

The squared covariance between the two fields [cf. BSW, Eq. (5)] is $\|C\|_F^2 = 1245$. Since the covariance matrix is based on normalized, equally weighted time series, this quantity is dimensionless. In order to eliminate the dependence on the size of the matrices, we will use what we will refer to as the *normalized squared covariance,* defined as $\|C\|_F^2/(N_s \cdot N_z) = 1245/(157 \times 125) = 0.063$.[1] In subsequent sections of the paper comparisons will be made of how much of the normalized squared covariance can be accounted for by the various modes in the separate and combined PCA expansions and in the SVD and CCA expansions, based on the methodology developed in section 5 of BSW.

Before the results are shown, a few preliminary comments regarding statistical significance are in order.

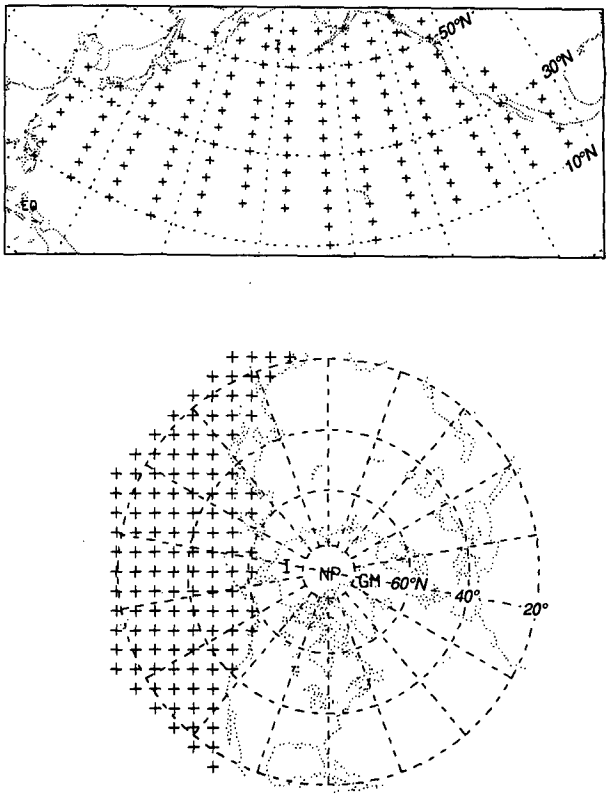


FIG. 1. Domain of the analysis in sections 3–6. The upper panel shows the 157 grid points used to represent the SST field, and the lower panel shows the 125 grid points used to represent the 500-mb height field. All grid points are equally weighted in the analysis.

In the absence of year-to-year correlation, the number of degrees of freedom in the correlation coefficients derived from the dataset used in this study would be 37 (2 less than the number of winters). Most of the extratropical time series exhibit positive 1-year lag correlations; the strongest of which approach values of 0.3. According to the algorithm in Leith (1973), such a level of autocorrelation would reduce the effective number of degrees of freedom from 37 to ~20. For 20 degrees of freedom, the 95% (99%) confidence level based on a two-tailed test corresponds to a correlation coefficient of 0.42 (0.54) in absolute value.

When multiple regression is applied to a random dataset in which the number of independent samples (in the time domain) is rather small, the expected value of the multiple correlation coefficient increases rapidly with the number of predictors included in the regression equation. In a similar manner, the expected value of the correlation between the expansion coefficients $a_k(t)$ and $b_k(t)$ increases with the number of spatial degrees of freedom in the fields: the larger the number of spatial degrees of freedom that are available in the fields, the easier it is to construct linear combinations that are highly correlated in the time domain. Hence, the existence of impressively high correlation coeffi-

[1] This is equivalent to the square of the temporal correlation coefficient between SST and Z_{500}, averaged over all pairs of grid points in the two fields. It would attain a value of unity if both field variables, at every one of their grid points, exhibited a perfect linear correlation with the other field variable at every one of its grid points.

cients between the respective $a_k(t)$ and $b_k(t)$ derived from SVD and CCA does not, in and of itself, guarantee statistical significance.

A well-established procedure exists for testing the significance of canonical correlations (Bartlett 1941, 1947). It is not directly applicable to the present study because it requires a knowledge of the number of spatial degrees of freedom in the two fields. We are not aware of the existence of any formal significance tests for the results of direct SVD. In this paper, we will rely on a simple Monte Carlo approach for assessing statistical significance.

3. Principal component analysis

As a basis of comparison for our subsequent calculations, we first performed conventional PCA on the individual covariance matrices of the normalized SST and 500-mb height fields for the 39 winters.

Figure 2 shows the first three EOFs of Pacific SST, scaled such that the value at each grid point is the correlation coefficient between the time series of the principal components (i.e., the expansion coefficients of the EOFs) and the SST anomalies at that point. Note that in many of the subsequent figures in this section the spatial domain encompasses an area larger than was used in the PCA. The fractions of variance of the normalized fields explained by the respective modes are indicated in the upper right-hand corner of each map. In these and all subsequent figures, the reference time series used in generating the correlation map is denoted by the label in the upper-left corner of the map. The label consists of a letter denoting the type of analysis (in this case E for EOF), followed by the mode number, followed in parentheses by the field name (SST or Z_{500}).

The spatial pattern associated with E_1(SST), which accounts for 20.8% of the total variance of Pacific SST, exhibits an out-of-phase relationship between temperature anomalies in the central Pacific (~40°N, 155°W) and those in the tropical Pacific and along the west coast of North America. This pattern resembles the dominant mode in previous EOF analyses of North Pacific SST based on various SST grids and periods of record (e.g., Davis 1976; Weare et al. 1976; Iwasaka et al. 1987). The eigenvalues for the next two modes are not well separated according to the criterion of North et al. (1982), but nevertheless, one of them will prove to be of interest. The EOF mode, E_2(SST), which accounts for 10.3% of the variance, exhibits maximum amplitude in the subtropics, and E_3(SST), which accounts for 9%, exhibits a banded structure with maxima along 50°N and 15°N separated by a band of the opposite polarity along 25°N in the western Pacific.

The first two EOFs of normalized wintertime 500-mb height field shown in Fig. 3 are based on data for the Pacific sector, whereas most of those in the literature are based on the full hemispheric domain. The $E_1(Z_{500})$

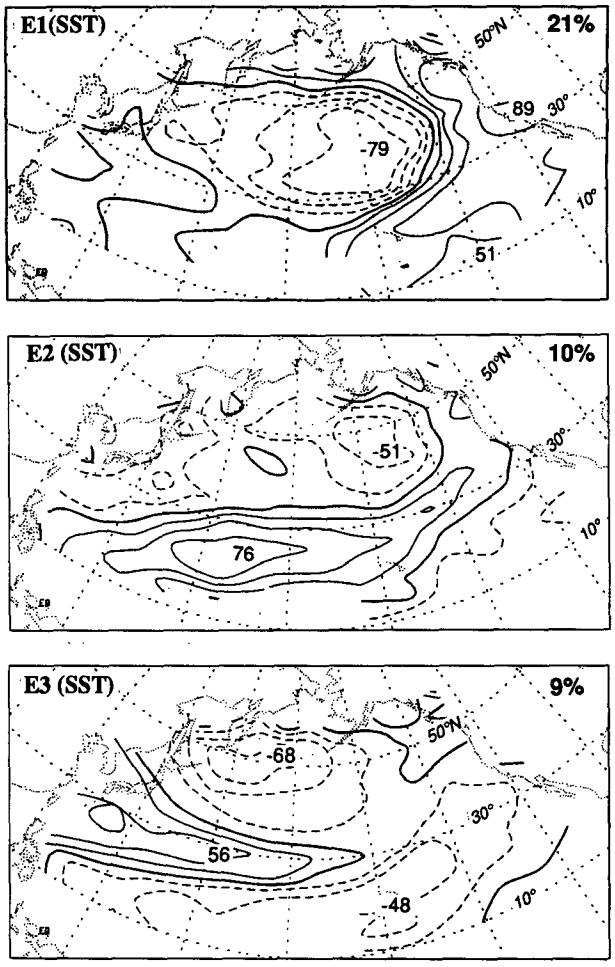


FIG. 2. The first three EOFs of normalized Pacific SST based on data for 39 winter seasons (December–February) 1946–85. The value at each grid point is the correlation coefficient between the time series of the principal component and the SST anomalies at that point. In many of the subsequent figures in this section, the spatial domain encompasses an area larger than was used in the analysis. The fractions of variance of the normalized data field explained by the respective modes are indicated in the upper right-hand corner of each map.

mode, which accounts for 32% of the total variance of 500-mb height over the Pacific sector, consists of a north–south dipole with the zero line along 40°N. It bears some resemblance to the western Pacific (WP) pattern identified by Wallace and Gutzler (1981, hereafter denoted as WG): The temporal correlation coefficient between its principal component and the WP index defined in WG is −0.71. The $E_2(Z_{500})$ mode, which accounts for 23% of the variance, resembles the Pacific/North American (PNA) pattern defined in WG (the corresponding correlation coefficient is 0.81).

In studies of relationships between two different field variables, the expansion coefficients of one field can be correlated with time series of the other field at each grid point to isolate the spatial patterns associated with that mode. [For an example germane to the present

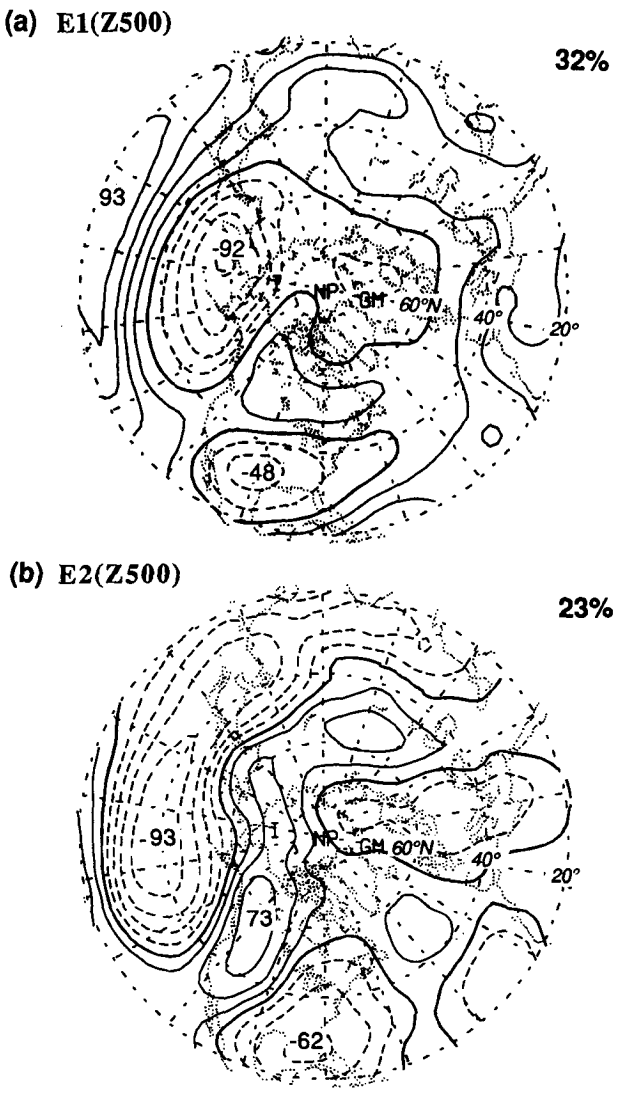


FIG. 3. As in Fig. 2 but for the first two EOFs of normalized wintertime 500-mb height.

study, see Iwasaka et al. (1987).] Figure 4 shows the correlation pattern between the principal component of E_1(SST) and the hemispheric 500-mb height field. The atmospheric circulation pattern associated with this mode more closely resembles the PNA pattern as defined in WG and the rotated principal component analyses of Horel (1981), Barnston and Livezey (1987), and Kushnir and Wallace (1989) than E_1 or E_2 of the North Pacific 500-mb height field shown in the previous figure. A similar correlation pattern was obtained by Iwasaka et al. (1987) based on a much shorter dataset. Our pattern accounts for 13% of the variance of the 500-mb height field over the North Pacific (as calculated by taking the spatial average of the square of the correlation coefficients in Fig. 4) for the gridpoints in Fig. 1. The fraction of the squared covariance between the two fields that can be accounted for by virtue of their common link to E_1(SST), as calculated from the algorithms in BSW amounts to 45%.

Figure 5 shows the correlation pattern in the SST field when the time series of the expansion coefficients of $E_1(Z_{500})$ is used as the reference time series. It bears a rather strong resemblance to E_3(SST) (Fig. 2). However, the correlations are not as strong as those between the expansion coefficients of E_1(SST) and the 500-mb height field shown in the previous figure, and the squared covariance fraction is only 39%. Table 1 shows the percentage of the variances of the individual fields and the squared covariance fraction (SCF) between the two fields accounted for by the time series of the various principal components, where the correlation coefficients are derived from maps such as Figs. 2–5. These results will be compared with corresponding results based on direct SVD in the next section.

In the foregoing example, the patterns in the leading mode turned out to be stronger and more consistent with those obtained in previous studies when the PCA is based on the SST field than when it is based on the 500-mb height field. There is no guarantee that either set of patterns is really optimal, in any objective sense, for defining the relationships between the two fields. In the next section, we will demonstrate that direct SVD of the same dataset yields patterns that are more strongly related than any of those shown in section 3.

4. Direct SVD analysis

As input to the SVD subroutine in the *Linpack* statistical program package, we computed the temporal

FIG. 4. Correlation coefficient between the time series of the expansion coefficient of E_1(SST) and the hemispheric 500-mb height field.

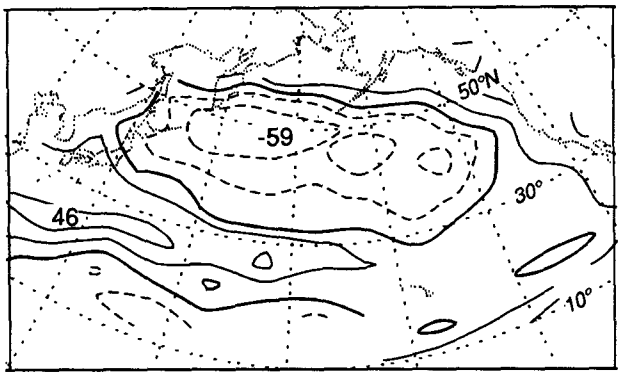


FIG. 5. Correlation coefficient between the time series of the expansion coefficient of $E_1(Z_{500})$ and the SST field over the North Pacific.

covariance matrix between normalized SST at each of the 157 grid points in Fig. 1 (upper panel) and 500-mb height at each of the 125 grid points in Fig. 1 (lower panel). The output consists of the two matrices, each comprised of 125 singular vectors, the "left" matrix for SST and the "right" for 500-mb height, together with the same number of singular values.

As noted in BSW, the singular values provide a measure of the squared covariance fraction SCF accounted for by the various singular value vectors, just as the eigenvalues of a hermitian covariance matrix provide a measure of the relative amounts of variance accounted for by the various principal components. The fraction of the total squared covariance between the two fields accounted for by any mode is proportional to the square of its singular value.

It has become common practice in the literature to scale eigenvectors and rotated principal components such that the value at each grid point is equivalent to the temporal correlation coefficient between the expansion coefficient of that mode and the field variable at that same grid point (see the Appendix of Hsu and Wallace 1985). For temporally normalized fields, the singular vectors can be scaled in an analogous manner by multiplying them by the corresponding singular values and dividing by the temporal standard deviation of the expansion coefficient. Following BSW, these will be referred to as *heterogeneous correlation maps* $r[b_k(t), s_i(t)]$, and $r[a_k(t), z_j(t)]$, since they involve correlations between the expansion coefficients of one field and the gridpoint values of the other field. These maps can also be generated by projecting the singular value vectors onto the corresponding anomaly maps for each of the 39 seasons to obtain expansion coefficients $a_k(t)$ and $b_k(t)$ and by temporally correlating these expansion coefficients with the field variables at each grid point. As noted in BSW, the patterns on the heterogeneous correlation map for the kth mode in the SVD expansion indicate how well the pattern of anomalies in the 500-mb height (SST) field can be specified on the basis of a knowledge of the kth expansion coefficient of the SST (500-mb height) field. The maps for a given field (SST or 500-mb height) are mutually orthogonal in the space domain.

Once $a_k(t)$, and $b_k(t)$ have been obtained, it is equally straightforward to construct homogeneous correlation maps $r[a_k(t), s_i(t)]$, and $r[b_k(t), z_j(t)]$, relating the expansion coefficient of the kth mode to the time series of the *same* field variable at each grid point, as explained in section 2a of BSW. The homogeneous correlation maps do not have such a direct relation to the singular vectors, and they are not mutually orthogonal. However, they are of interest because they show the spatial patterns whose polarity and amplitude the corresponding expansion coefficients are representing. It is expected that the heterogeneous correlations will tend to be weaker than the homogeneous correlations. Differences in the shapes of the patterns on the heterogeneous and homogeneous correlation maps for the same field, if significant, could conceivably provide insights into the nature of cause/effect linkages between the fields.

As in PCA, there is no need to restrict either kind of correlation map to the domain or grid spacing that was used in performing the singular value decomposition. For display purposes, the SST analysis is extended into the tropics and the 500-mb patterns are presented on the full-resolution, 1977-point hemispheric NMC grid. The above computations are summarized in Fig. 1 of BSW.

The correlation coefficient between the expansion coefficients $r[a_k(t), b_k(t)]$ provides a measure of how strongly the left and right fields are related to one another. It is this quantity that is maximized in canonical correlation analysis discussed in the following section. Table 2 shows selected statistics for the five leading modes in the SVD expansion. It is evident that the squared covariance is strongly concentrated in the first three modes. Consistent with the discussion in section 5a of BSW, the squared covariance fraction for the first mode is much larger than the variance fraction for the first mode in either of the principal component expansions considered in the previous section. As noted in BSW, much of the difference is simply a reflection of

TABLE 1. Percentages of the variance of the two fields accounted for by the various principal components of the individual fields, as indicated, and the squared covariance fraction computed as explained in BSW.

	Field		
	SST	Z_{500}	SCF
E_1(SST)	20.8	13.4	45
E_2(SST)	10.3	7.2	11
$E_1(Z_{500})$	8.7	32.3	39
$E_2(Z_{500})$	9.0	23.7	33

TABLE 2. The squared covariance fraction for the leading modes of a direct SVD expansion, expressed as a percent, and the temporal correlation coefficient between the expansion coefficients of the SST and 500-mb height fields associated with the same mode.

k	SCF	CSCF	$r(a_k, b_k)$
1	52	52	0.81
2	23	75	0.80
3	12	87	0.86
4	4	91	0.70
5	4	95	0.77

the fact that the covariances are squared in the accounting scheme, whereas the variances are not.

The heterogeneous correlation patterns for the two leading SVD modes are shown in Figs. 6 and 7. In each figure the identifier of the reference time series used to create each map is printed on the upper-left corner (S for SVD). Between the two maps, $r(a_k, b_k)$ is printed on the left-hand side of the page, and SCF, expressed as a percent, on the right in bold type. The fractions of the variances of the respective fields explained by the time series of the expansion coefficient of the other field is printed on the upper right-hand corner of each map. Figures 8 and 9 show the homogeneous correlation maps for the same two modes in the same format, where the label S^* is used to distinguish them from the heterogeneous correlation maps. In the latter figures the numbers printed on the upper right-hand corner of each map, in lighter type, refer to the fractions of the variances of the respective fields explained by the time series of the expansion coefficient of the *same* field.

As expected, the "centers of action" over the Pacific sector tend to be stronger in the homogeneous correlation maps than in the heterogeneous correlation

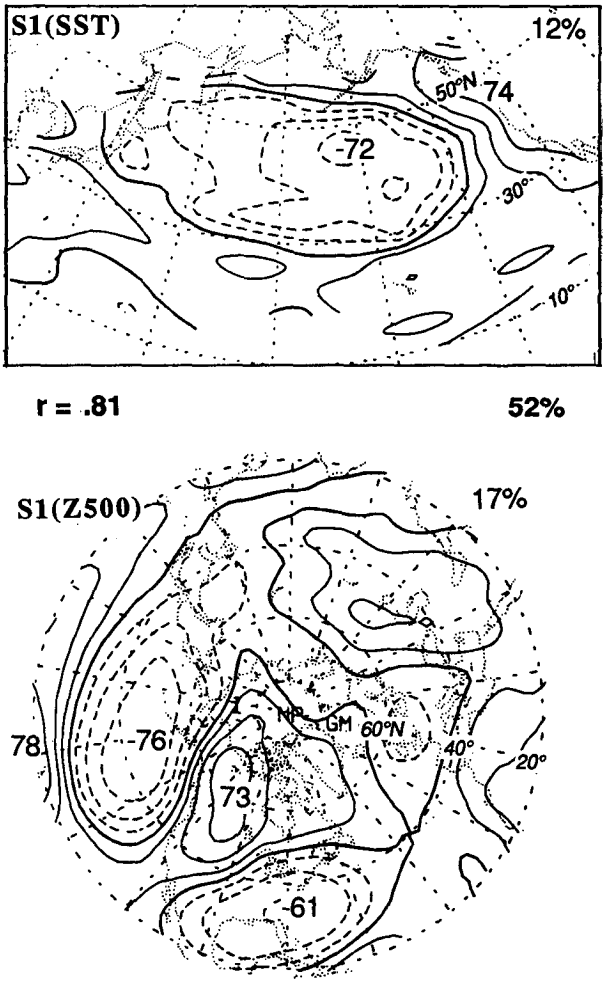


FIG. 6. Heterogeneous correlation pattern for the first mode in the direct SVD expansion. The temporal correlation coefficient between the corresponding expansion coefficients $r(a_1, b_1)$, expressed as a decimal, is printed on the left, and *SCF*, expressed as a percent, is printed on the right. The percentages of the variances of the respective fields explained by the time series of the expansion coefficient of the other field is printed on the upper right-hand corner of each map, in lighter type.

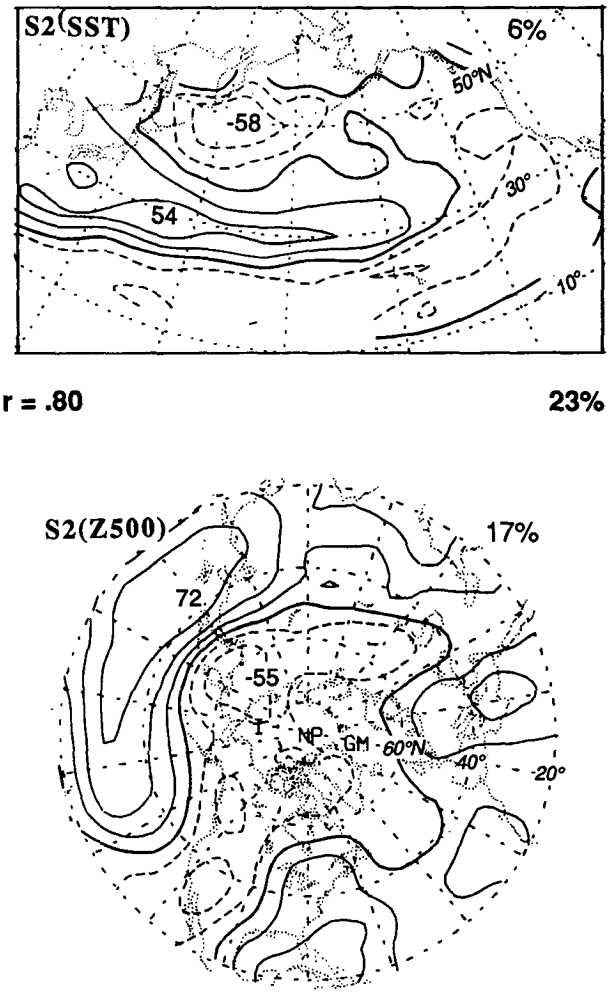
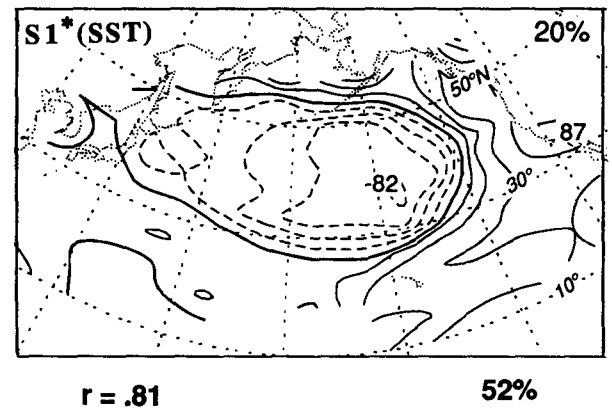


FIG. 7. As in Fig. 6 but for the second mode in the SVD expansion.

maps. The shapes of the patterns on corresponding maps are similar, particularly over the domain of the SVD analysis, but there are some notable differences in the 500-mb height patterns for the first mode: The positive correlations over western Canada are stronger in the heterogeneous correlation map than in the homogeneous correlation map, and the PNA pattern, as defined in WG, is more prominent. The strong resemblance between respective heterogeneous and homogeneous correlation maps for E_1(SST) and S_1 (Fig. 4 versus Fig. 6, lower panel, and Fig. 2 versus Fig. 8, upper panel) and the strength of the temporal correlation (0.97) between the expansion coefficients of S_1(SST) and E_1(SST) confirm our intuitive impression in the previous section that the single field principal component analysis based on SST provides more information on the coupling between the fields than its



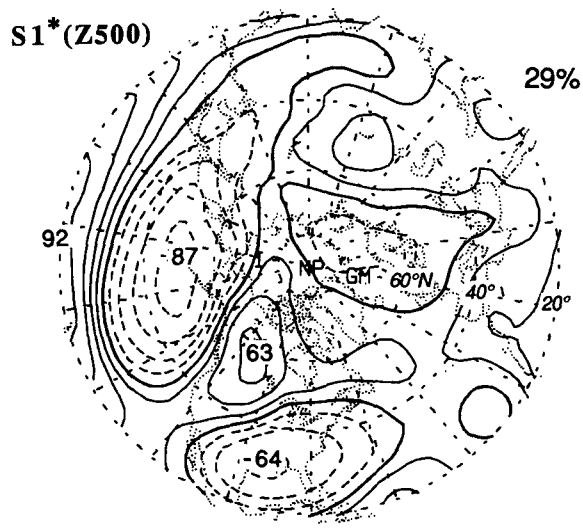
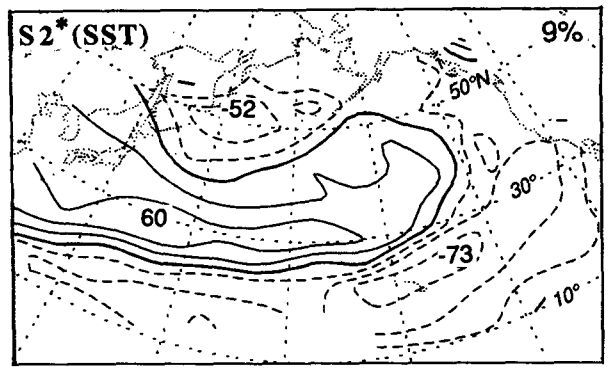


FIG. 8. Homogeneous correlation pattern for the first mode in the direct SVD expansion. The numbers printed on the figure are analogous to those in Fig. 6, except that the numbers printed on the upper right-hand corner of each map in lighter type refer to the percentages of the variances of the respective fields explained by the time series of the expansion coefficient of the *same* field.



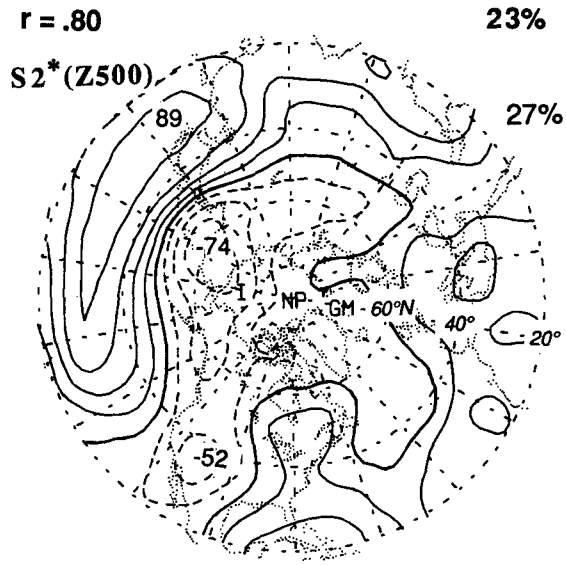


FIG. 9. As in Fig. 8 but for the second mode in the SVD expansion.

counterpart, based on 500-mb height.[2] From a comparison of percentage variances in Fig. 8 with corresponding statistics in Figs. 2 and 3, it is evident that the expansion coefficients associated with S_1 account for nearly as much of the variance of their own fields as the leading EOFs of those fields (19.9% versus 20.8% for SST and 29.2% versus 32.3% for Z_{500}). The patterns for the second mode are not very closely related to any of the EOFs displayed in the previous section.

The SST expansion coefficients for S_1 and S_2 are moderately successful at specifying the PNA and WP patterns [for the 39-winter sample, $r(S_1, \text{PNA}) = 0.77$ and $r(S_2, \text{WP}) = 0.65$, where PNA and WP are the pattern indices defined by WG]. It is interesting to note that the correlation maps for S_1 and S_2 more

[2] Here $S_1(Z_{500})$ is not as closely related to any of the modes in the EOF expansion of the 500-mb height field. Its expansion coefficient exhibits a temporal correlation of 0.81 with that of $E_1(Z_{500})$ and 0.56 with that of $E_2(Z_{500})$.

closely resemble patterns derived from rotated principal component analysis of the hemispheric 500-mb height field (Horel 1981; Barnston and Livezey 1987; Kushnir and Wallace 1989) than the EOF patterns for the Pacific domain shown in Fig. 3. Evidently, some information on the hemispheric structure of the 500-mb height field is implicit in the patterns of North Pacific SST anomalies.

The S_2(SST) and S_2^*(SST) fields exhibit a distinctive "sandwich" pattern, with anomalies in the western Pacific along 30°N flanked by anomalies of the opposite sign along 50°N and in the trade-wind belt. Note the strong resemblance to E_3(SST) shown in Fig. 2. Wallace et al. (1990) noted a similar structure in the first EOF of midwinter SST tendency in both the North Pacific and the North Atlantic.

The results presented in this section are not strictly comparable with those of Lanzante (1984) because, as a final step in his analysis, Lanzante performed a varimax rotation of his spatial patterns, which was carried out independently for the SST and 500-mb height fields. This procedure is capable of destroying the unique pairing that exists between the patterns in the two fields. In addition, the two studies are based on quite different domains: Lanzante incorporated SST anomalies in both the North Pacific and the North Atlantic and 700-mb height on a 10° latitude–longitude grid extending from 15°W to 155°E, and his winter season is a month later than ours. Despite the differences in methodology, our first SVD mode strongly resembles Lanzante's leading mode for his winter season.

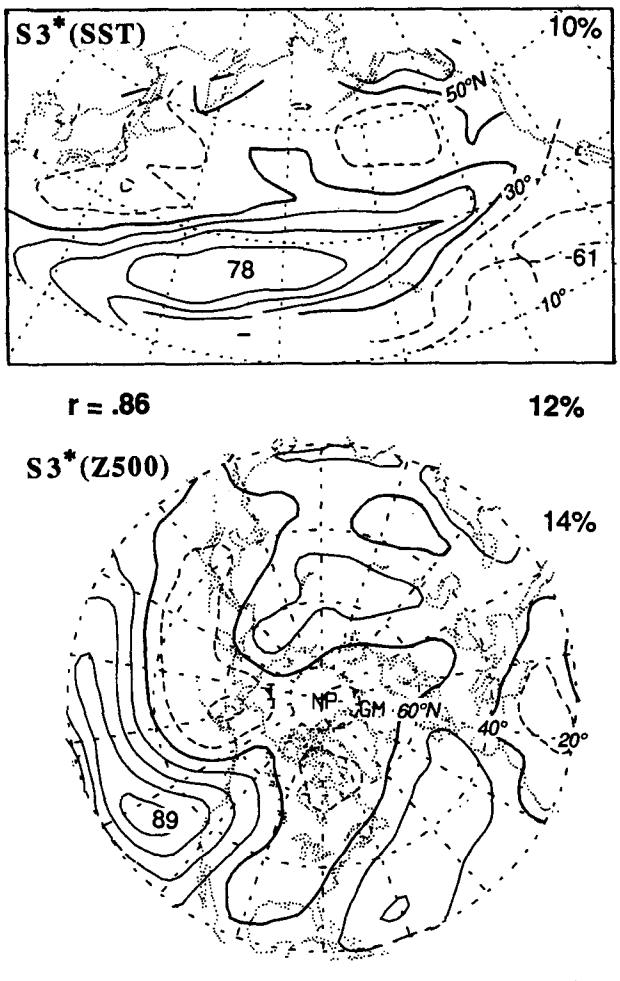


FIG. 10. As in Fig. 8 but for the third mode in the SVD expansion.

The S_3 mode accounts for another 12% of the squared covariance between the SST and 500-mb height fields: Its homogeneous correlation patterns are shown in Fig. 10. The existence of this mode indicates that SST in a broad band across the subtropical central Pacific is positively correlated with 500-mb height anomalies to the northeast of Hawaii. The SST pattern strongly resembles E_2(SST), while the 500-mb height pattern resembles Barnston and Livezey's "east Pacific" (EP) pattern. The expansion coefficients associated with this mode are more strongly correlated with one another than those associated with the first two modes, but they account for much smaller fractions of the variance of their respective fields. Corresponding patterns for S_4 and S_5, shown in Smith (1989), tend to be smaller in horizontal scale than the leading modes, and the correlations associated with them tend to be weaker. Together they account for ~7% of the squared covariance, leaving only ~6% to be accounted for by modes higher than the fifth.

As a test of the statistical robustness of these results, we repeated the preceding analysis on subsets of the data for odd and even years. Results for S_1, shown in Fig. 11, are in good agreement. In the results that we inspected, the degree of reproducibility appears to be comparable to that of rotated component analysis as applied to similar datasets.

As an additional test, following one of the procedures suggested by Barnett and Preisendorfer (1987), a similar analysis was performed on the full 39-year dataset, but with the 39 observed 500-mb charts ordered randomly in the time domain so that most of them were paired with the wrong SST field. The results of 28 such Monte Carlo SVD expansions are summarized in Table 3. The squared covariance between the two scrambled fields averages about 40% as large as the observed covariance, and in none of the runs does it closely approach the observed. Hence, there is little question as to the existence of statistically significant relationships between the two fields. The relatively large squared covariance fractions and $r(a_1, b_1)$ of the leading modes in these scrambled runs illustrate the need for significance testing. In this particular case, the observed value of $r(a_1, b_1)$ stands out sufficiently from the corresponding population for the scrambled data; hence, we can be quite confident that this mode is statistically significant. However, squared covariance fractions on

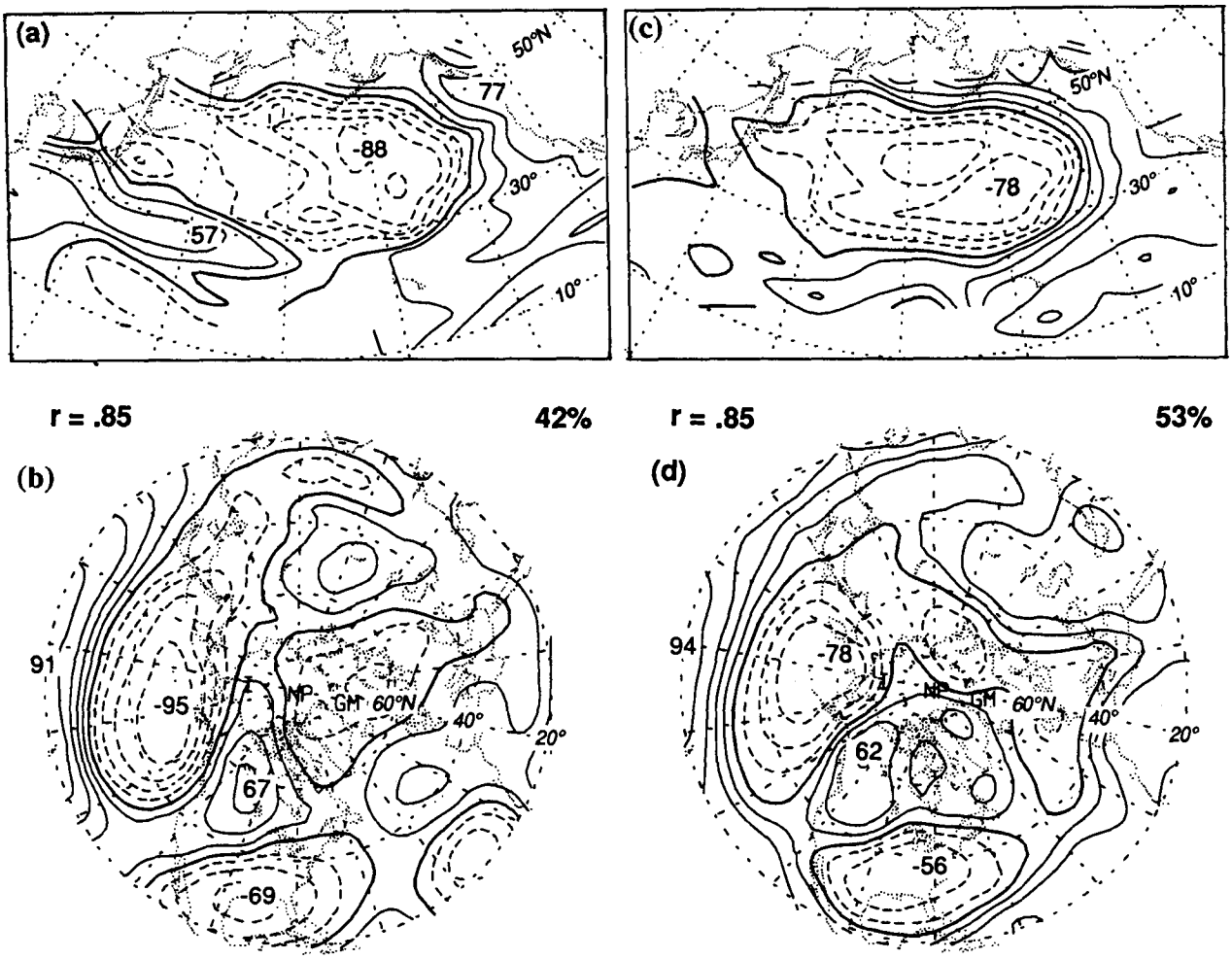


FIG. 11. As in Fig. 8 but for winter seasons whose Januarys fall in odd (a,b) and even numbered (c,d) calendar years, as indicated.

the order of the observed value are not uncommon among the random runs. These high values of SCF are probably a consequence of the dominance of each individual field by the leading EOFs. For instance, if both fields were dominated by their own leading EOFs and if the expansion coefficients of those EOFs were unrelated to one another, then the squared covariance between the two fields would be small and entirely due to sampling fluctuations, but what little there was would be explained by a single SVD mode whose left singular vector was the left EOF and whose right singular vector was the right EOF.

The structure of the leading SVD modes for the scrambled datasets is revealed by the temporal correlations between the SVD expansion coefficients and the principal components of the two fields. (Note that the spatial structure of the EOFs is unaffected by the scrambling of the data in the time domain.) In 18 of the 28 runs, $S_1(Z_{500})$ is correlated with $E_1(Z_{500})$ at a level of 0.90, in absolute value or higher, and in 12 runs it is correlated at a level of 0.95 or higher. In only 8 of the runs is the relation with $E_1(\mathrm{SST})$ above 0.90 and in only 3 is it above 0.95. Hence, in about half of the scrambled runs, the structure of the SST field is apparently being adjusted so as to optimize the "fit" with $E_1(Z_{500})$. We suspect that the 500-mb height field may be dominant because its leading EOF accounts

TABLE 3. Summary of results of 28 SVD expansions, each based on the 39 observed SST and 500-mb height fields but with the temporal order of the 500-mb height fields randomized. Here $\|\mathbf{C}\|_F^2$ represents the squared covariance between the two scrambled fields; SCF and r refer to the leading mode in the expansion.

	$\|C\|_F^2/N_s \cdot N_z$	SCF (%)	$r(a_1, b_1)$
28 scrambled runs:			
Highest	0.034	57	0.78
Mean	0.027	40	0.60
Lowest	0.020	28	0.46
Std. dev.	0.004	7	0.07
Observations:	0.063	52	0.81

for a substantially larger fraction of its variance than in the case of SST.

5. Canonical correlation analysis (CCA)

In view of the large percentage of the squared covariance between the SST and 500-mb height fields explained by the first few modes in the direct SVD expansion, we decided to apply the methodology for computing the canonical correlation vectors proposed in BSW, starting with the first two modes in the SVD expansion and proceeding by adding one mode at a time and watching how the results evolve. This procedure is analogous to the strategy often used to determine the sensitivity of rotated principal component analysis to the number of modes retained in the expansion. The computational procedure is outlined schematically in Fig. 3 of BSW and in the accompanying discussion therein. Proceeding as in the previous section, these vectors or weighting functions were projected onto the corresponding anomaly maps for the corresponding field for each of the 39 seasons to obtain expansion coefficients $a_k(t)$ and $b_k(t)$, which were temporally correlated with their respective field variables at each grid point to derive homogeneous correlation maps analogous to those shown in the previous section.

In some of the idealized examples that we have examined, the canonical correlation vectors were much more localized in the space domain than the homogeneous correlation patterns derived from them. However, in our analysis, the patterns in the two types of maps proved to be quite similar. In the interest of brevity, the following discussion is based primarily on the homogeneous correlation maps, which provide information on the strength of the relationships between the fields represented by the canonical correlation vectors, as well as the shapes of the patterns. Only a few examples of the canonical correlation vectors themselves will be shown.

The labeling conventions for the reference time series are analogous to those in previous figures, where C stands for canonical correlation vector. The labels contain two subscripts: The first denotes the number of SVD modes retained in the expansion, and the second stands for the order of the mode in the expansion. In CCA the modes are ordered on the basis of the correlation coefficient between the expansion coefficients of SST and the 500-mb height: The leading mode in the expansion is not necessarily the one that explains the largest fraction of the squared covariance between the SST and 500-mb height fields.

Homogeneous correlation maps derived from the canonical correlation vectors based on the first two SVD modes are shown in Figs. 12 and 13. The homogeneous correlation patterns for 500-mb height bear a striking similarity to the EOFs shown in Fig. 3. The temporal correlations between their respective expansion coefficients are 0.98 for the first mode and 0.99 for the second mode. The homogeneous correlation pattern for SST for $C_{2,1}$ strongly resembles E_3(SST), while that for $C_{2,2}$ bears a weaker resemblance to E_1(SST). The relations between the SST and Z_{500} fields are reminiscent of those obtained from the EOF expansion based on the 500-mb height field shown in Figs. 3 and 5. Consistent with the optimization criterion for CCA, the correlation coefficient between the expansion coefficients of the first mode (0.85) is stronger than that associated with either of the singular vectors on which the solution is based.

Corresponding results based on the first three singular vectors are shown in Figs. 14–16. The homogeneous correlation patterns for $C_{3,2}$ resemble those for $C_{2,1}$, though the patterns have been degraded somewhat by the addition of another SVD mode, as evidenced by the declines in the strength of the various centers of action and in the drops in the percentage of the total variance of the two fields that they explain. The patterns

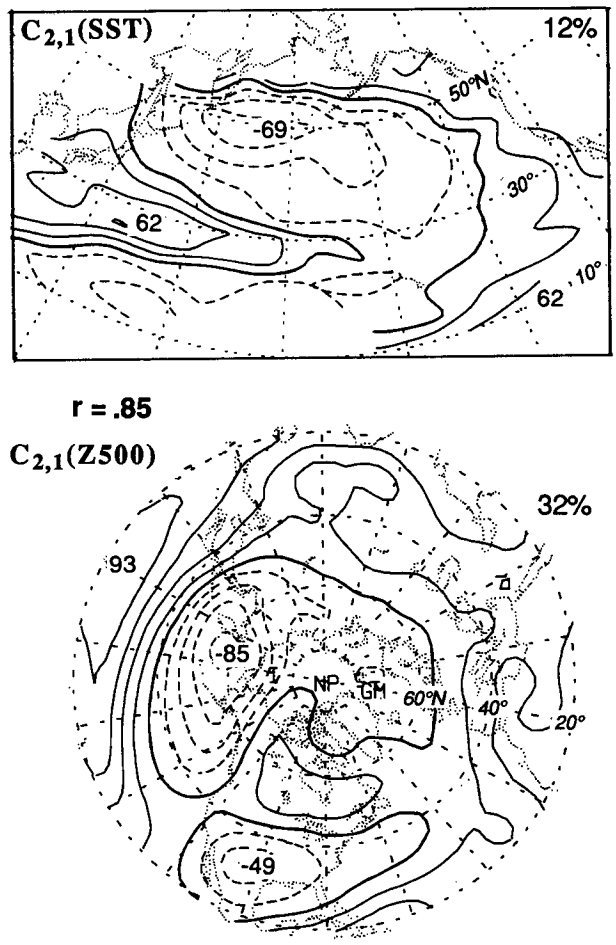


FIG. 12. Homogeneous correlation map for the first canonical correlation vector based on the first two SVD modes. Labeling conventions as in Fig. 8.

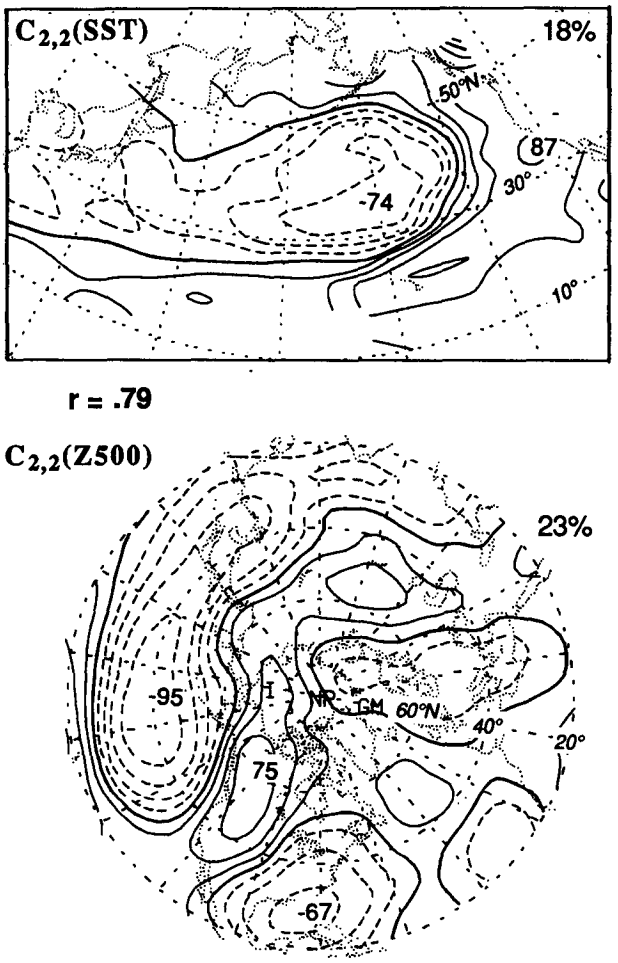


FIG. 13. As in Fig. 12 but for the second canonical correlation vector.

for $C_{3,3}$ closely resemble those for $C_{2,2}$, and they have suffered little, if any, degradation. A notable consequence of the additional degree of freedom is the appearance of a new mode $C_{3,1}$, whose 500-mb height signature resembles the Southern Oscillation signature identified by Horel and Wallace (1981), Wallace and Jiang (1987), Graham (1990), and others, and it contains elements of what Mo and Livezey (1986) referred to as the "Tropical-Northern Hemisphere (TNH) Pattern." Its expansion coefficient for the Z_{500} field is correlated with equatorial Pacific SST (an index based on SST averaged from 6°N to 6°S and from the date line to the South American coast for the corresponding months) at a level of 0.79. The strong contribution from S_3 is evident in the primary center of action over the subtropical southwest Pacific in SST and northeast of Hawaii in Z_{500}. This pattern assumes first place in the ranking by virtue of the strength of the correlations between its expansion coefficients (0.89). It accounts for somewhat less of the squared covariance than the other modes, and the centers of action in the correlation patterns are not particularly strong in either field.

Figure 17 shows the weight (canonical correlation) vectors for the 500-mb height field in the three modes discussed in the previous paragraph. The similarities with the corresponding homogeneous correlation maps are quite obvious, especially in the second and third modes. However, the amplitude in the first mode is somewhat more concentrated in the negative center of action to the west of California in the canonical correlation vector than in the homogeneous correlation map. The corresponding SST maps (not shown) are virtually identical in shape to the corresponding homogeneous correlation maps in Figs. 14–16.

In view of the marked decline of statistical significance of canonical correlation analysis as the number of "predictors" is increased, there is little point in carrying the analysis beyond the this point, since 87% of the squared covariance between the fields has already been accounted for and each of the next two SVD modes accounts for only another 4%. For the sake of

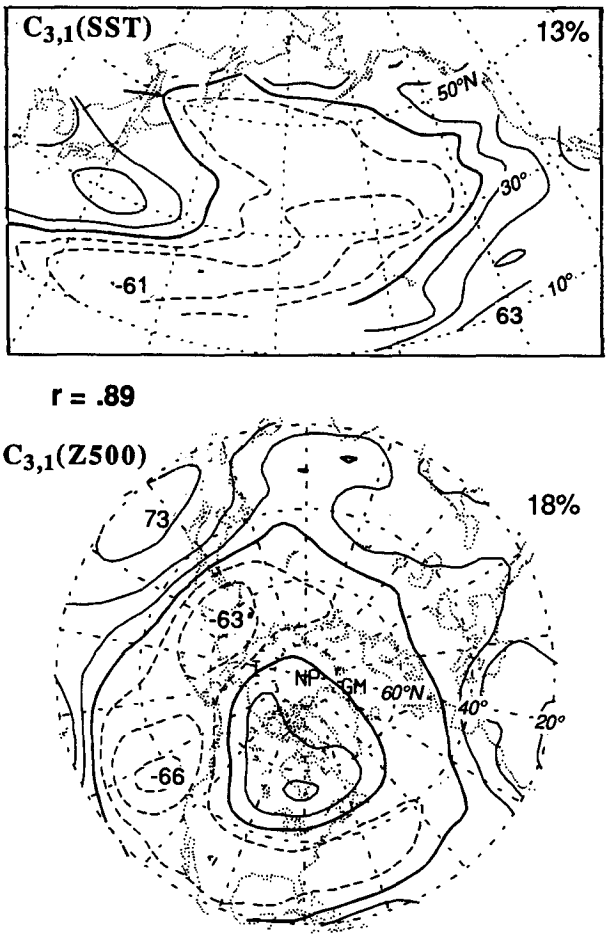


FIG. 14. Homogeneous correlation map for the first canonical correlation vector based on the first three SVD modes.

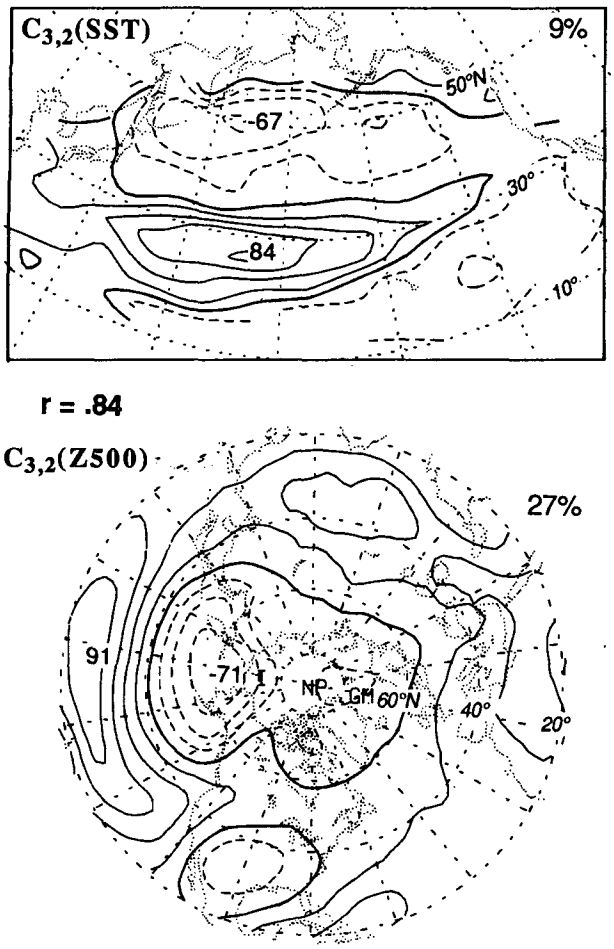


FIG. 15. As in Fig. 14 but for the second canonical correlation vector.

completeness, we extended the analysis to include these modes and found that the results are fairly stable. Additional documentation is given in Smith (1989).

Table 4 summarizes the CCA results. In the interest of brevity we have included only the fractions of the variances of the fields explained by their own expansion coefficients (i.e., the percentages printed in the homogeneous correlation maps). SCF and CSCF are computed as explained in BSW, where the canonical correlation ρ_k corresponds to $r(a_k, b_k)$. As expected, the correlation coefficient $r(a_1, b_1)$ is higher than in the SVD modes and increases monotonically with the number of spatial degrees of freedom in the expansion, while the squared covariance fraction is lower than in the leading SVD mode, and it does not tend to increase with the inclusion of additional SVD modes. Hence, which method is "better" depends on which measure of performance one uses as a basis for deciding.

Whether the CCA modes are dynamically more significant than the SVD modes on which they are based is an open question. The ENSO-like mode is intriguing, but is it of sufficient value to warrant transforming S_1 and S_2 into CCA patterns that look less like the patterns that have appeared in previous studies based on other analysis techniques? As with most questions involving the choice of analysis techniques, the verdict is likely to be strongly influenced by the interests and intuition of the analyst.

6. Combined principal component analysis (CPCA)

Combined principal component analysis (CPCA) was performed on the two normalized fields. No attempt was made to compensate for the fact that there are 157 SST grid points and only 125 for Z_{500}, and therefore, the analysis is slightly biased toward recovering the dominant patterns in the SST field. Results were similar, in many respects, to those obtained from the SVD expansion described in section 4. A summary of the performance characteristics is presented in Table 5.

SCF and CSCF for the leading CPCA modes tend to run only slightly lower than for the corresponding

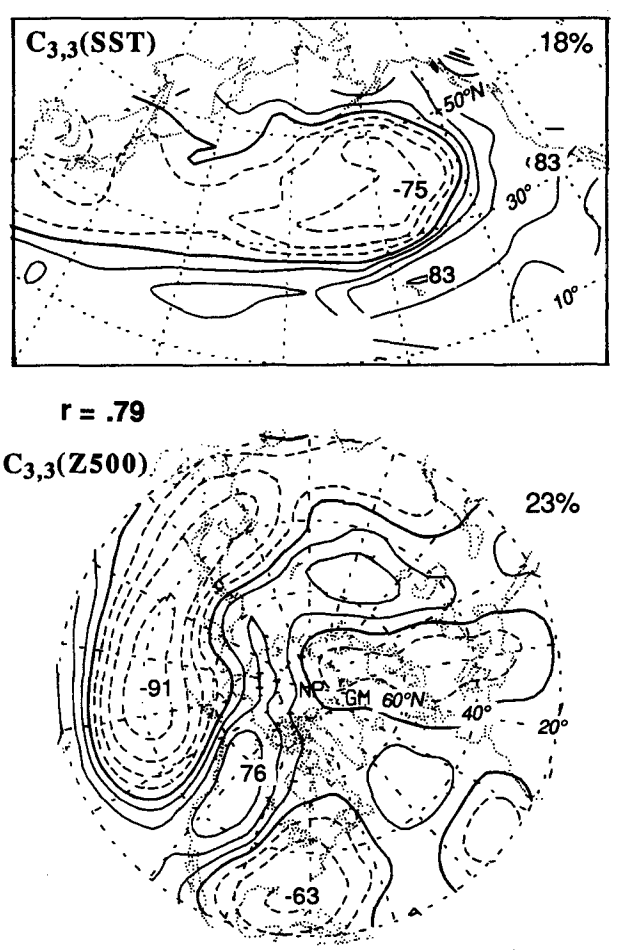


FIG. 16. As in Fig. 14 but for the third canonical correlation vector.

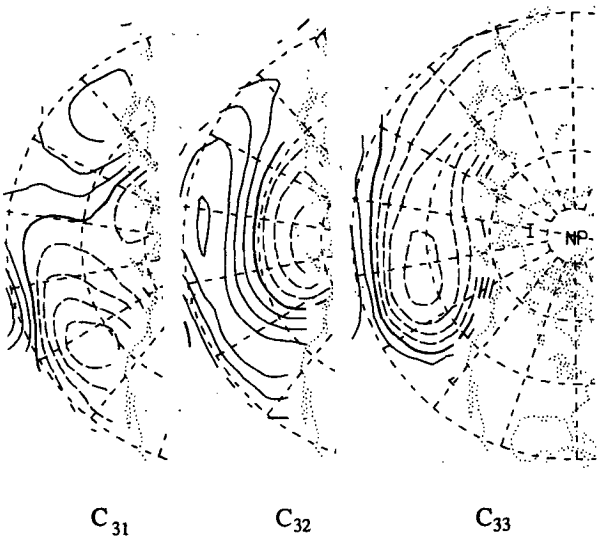


FIG. 17. Canonical correlation vectors or weighting functions for Z_{500} in the expansion based on the first three SVD modes. Positive and zero contours are solid; negative contours, dashed; contour interval, arbitrary.

TABLE 4. Summary of the CCA results. The correlation $r(a_k, b_k)$ is between the two expansion coefficients; the next two columns give the percentage of the variance of the two fields explained by their respective expansion coefficients; and SCF and CSCF are computed as explained in BSW.

	Resembles	$r(a_k, b_k)$	SST field	Z_{500} field	SCF	CSCF
$C_{2,1}$	$E_1(Z_{500})$	0.85	11.7	31.8	35	35
$C_{2,2}$	$E_2(Z_{500})$	0.79	17.5	23.0	33	67
$C_{3,1}$	ENSO	0.89	11.9	15.5	24	24
$C_{3,2}$	$E_1(Z_{500})$	0.84	10.4	27.2	27	47
$C_{3,3}$	$E_2(Z_{500})$	0.79	16.6	22.9	31	79
$C_{4,1}$	ENSO	0.90	12.2	16.8	27	27
$C_{4,2}$	$E_1(Z_{500})$	0.86	10.1	21.8	23	48
$C_{4,3}$	$E_2(Z_{500})$	0.79	16.0	22.8	28	78
$C_{4,4}$	—	0.70	7.8	13.0	8	83
$C_{5,1}$	ENSO	0.91	11.4	14.3	25	25
$C_{5,2}$	$E_1(Z_{500})$	0.86	10.3	23.7	27	48
$C_{5,3}$	$E_2(Z_{500})$	0.81	12.6	18.2	22	73
$C_{5,4}$	—	0.76	10.3	11.9	15	81
$C_{5,5}$	—	0.70	7.8	13.3	8	86

modes in the SVD expansion. The leading three modes (not shown) closely resemble their SVD counterparts: for example, the correlation coefficients between the time series of the expansion coefficients of the first mode in the CPCA expansion and S_1(SST) and $S_1(Z_{500})$ are 0.91 and 0.96, respectively.

7. SVD and CCA results for an expanded 500-mb height domain

Smith (1989) performed a series of experiments in which the Atlantic and Pacific SST domains were held fixed while the Z_{500} domains were extended in longitude in three discrete steps beginning with sectors 80° of longitude in width over the oceanic sectors and culminating with the full, coarse resolution (23 × 23 point) hemispheric grid. Results will be shown here for only the fourth and final pair of experiments with the full hemispheric domain for Z_{500}.

Figure 18 shows the first and second EOFs of the hemispheric 500-mb height field based on this 39-year dataset. The patterns are similar to those presented in WG (Fig. 27a,b) based on monthly mean wintertime data for two shorter datasets. The first, which accounts for 19% of the variance of the hemispheric 500-mb height field, is dominated by the WP pattern, but much of the variance that it explains is associated with 500-mb height anomalies in the subtropics, not only in the Pacific but throughout most of the hemisphere. The second mode, which accounts for 15% of the variance, is dominated by the PNA and WA patterns over the oceanic sectors.

Figure 19 shows the homogeneous correlation maps for the leading mode of two different direct SVD expansions based on this enlarged 500-mb height domain. The maps on the left are based on Pacific SST on the same grid as in the previous sections, and those on the right are based on North Atlantic SST on a 4° lat by 10° long grid poleward of 10°N. The SST pattern in the Pacific is virtually identical to the one for the first mode based on the more limited 500-mb height domain shown in Fig. 8, which, in turn, resembles E_1(SST) shown in Fig. 2. The corresponding 500-mb height pattern bears a strong resemblance to the second EOF of the hemispheric 500-mb height field shown in the previous figure (the correlation between their expansion coefficients is 0.97).

Corresponding results based on Atlantic SST (Fig. 19, right) are remarkably analogous. The SST pattern is virtually identical to the one obtained in an SVD expansion in which the 500-mb height domain is restricted to the Atlantic sector, and it bears a strong resemblance to E_1 of Atlantic SST (Smith 1989). Like its Pacific counterpart, the corresponding 500-mb height pattern bears a strong resemblance to the second EOF of the hemispheric 500-mb height field shown in the previous figure (the correlation between their expansion coefficients is 0.96).

In the second mode for Pacific SST and hemispheric 500-mb height (not shown), the SST and 500-mb height patterns bear a strong resemblance to their

TABLE 5. Summary of the CPCA results; SCF and CSCF.

k	Percent of total variance	Percent of SST variance	Percent of Z_{500} variance	SCF	CSCF
1	21.0	16.5	26.7	46	46
2	15.5	8.5	24.4	17	71
3	9.6	9.6	10.0	9	83
4	—	6.4	7.3	2	87
5	—	5.6	5.9	2	91

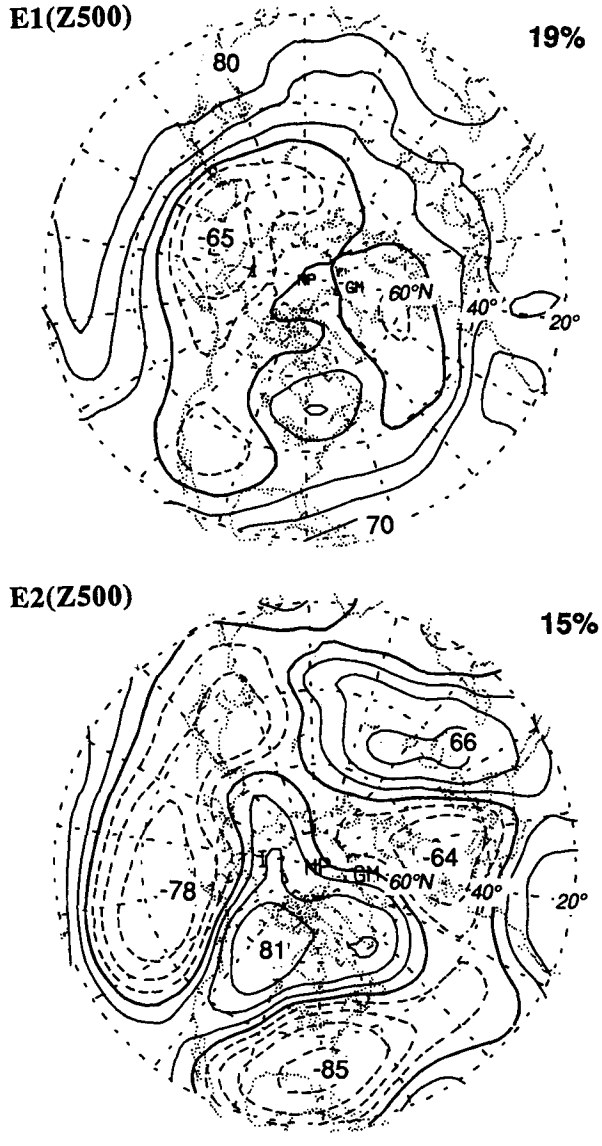


FIG. 18. The first and second EOFs of the normalized, wintertime (December–February) hemispheric 500-mb height field based on the 39-year dataset 1946–85.

counterparts in Figs. 7 and 9. The 500-mb height pattern resembles E_1 of the hemispheric 500-mb height field shown in Fig. 18 (the correlation between their expansion coefficients is 0.97).

The similarity between the 500-mb height patterns for the leading SVD modes based on Pacific and Atlantic SST is noteworthy. Apparently the dominant SST pattern in the two oceans is linked to a common 500-mb height pattern that extends over both oceans. By virtue of this common bond, they are temporally correlated with one another. Based on the statistics presented in Fig. 19 and in the text, it is readily verified that the automatic correlation between the SST expansion coefficients of S_1 of Pacific SST and S_1 of Atlantic SST (both analyzed with the same hemispheric 500-mb height field) is $0.82 \times 0.71 \times 0.97 \times 0.96 = 0.54$. The actual correlation is 0.61. Both values are statistically significant at the 99% level, even when one takes into account the autocorrelation inherent in the time series.

Returning to the statistics of the SVD expansion, we note that the correlation coefficient between the SST and 500-mb height fields is not particularly sensitive to the size of the 500-mb height domain, and the squared covariance fraction accounted for by the first mode drops off only from 52% to 44% as the domain is expanded to hemispheric. The fraction of the variance of the 500-mb height field accounted for by the mode drops from 29% to 15%, and the normalized squared covariance $\|C\|_F^2/(\mathbf{N}_s \cdot \mathbf{N}_z)$ drops by almost the same ratio.

Canonical correlation analysis was performed on the North Pacific SST field and the hemispheric 500-mb height field in the same manner as described in section 5, starting with the first two modes of the SVD expansion and terminating with the first five modes. Results based on the first two modes (not shown) are virtually identical to their counterparts in the SVD expansion, but the order is reversed (i.e., the WP-like pattern is the leading mode). When additional SVD modes are retained in the prefiltering, the patterns change only slightly.

8. Discussion and conclusions

In the applications explored in this study, the patterns derived from PCA, direct SVD, CCA, and CPCA exhibit many common elements. The leading modes in the expansions appear to be statistically significant and representative of real patterns of large-scale atmosphere–ocean interaction, as discussed in many previous papers on this subject.

In the scrambled SVD runs described at the end of section 4, the leading EOF of the 500-mb height field tends to recur in the SVD solutions with only slight modifications, while the patterns in SST exhibit more variability from one run to the next. We interpret this result as indicating that, in the absence of any real relationship between the two fields, the EOFs of the field with fewer spatial degrees of freedom (i.e., the field whose variance tends to be more strongly concentrated in the leading EOFs of its temporal variance matrix) tend to dominate the SVD solutions. The fact that the observed S_1 in section 4 exhibits a structure distinct from this "default solution" lends further credence to its statistical and dynamical significance.

In the SVD expansion based on the hemispheric 500-mb height field, discussed in section 8, the two leading modes are virtually identical to the leading EOFs of the hemispheric 500-mb height field and the SST fields over the North Pacific and North Atlantic. The strong correspondence suggests that the dominant spatial

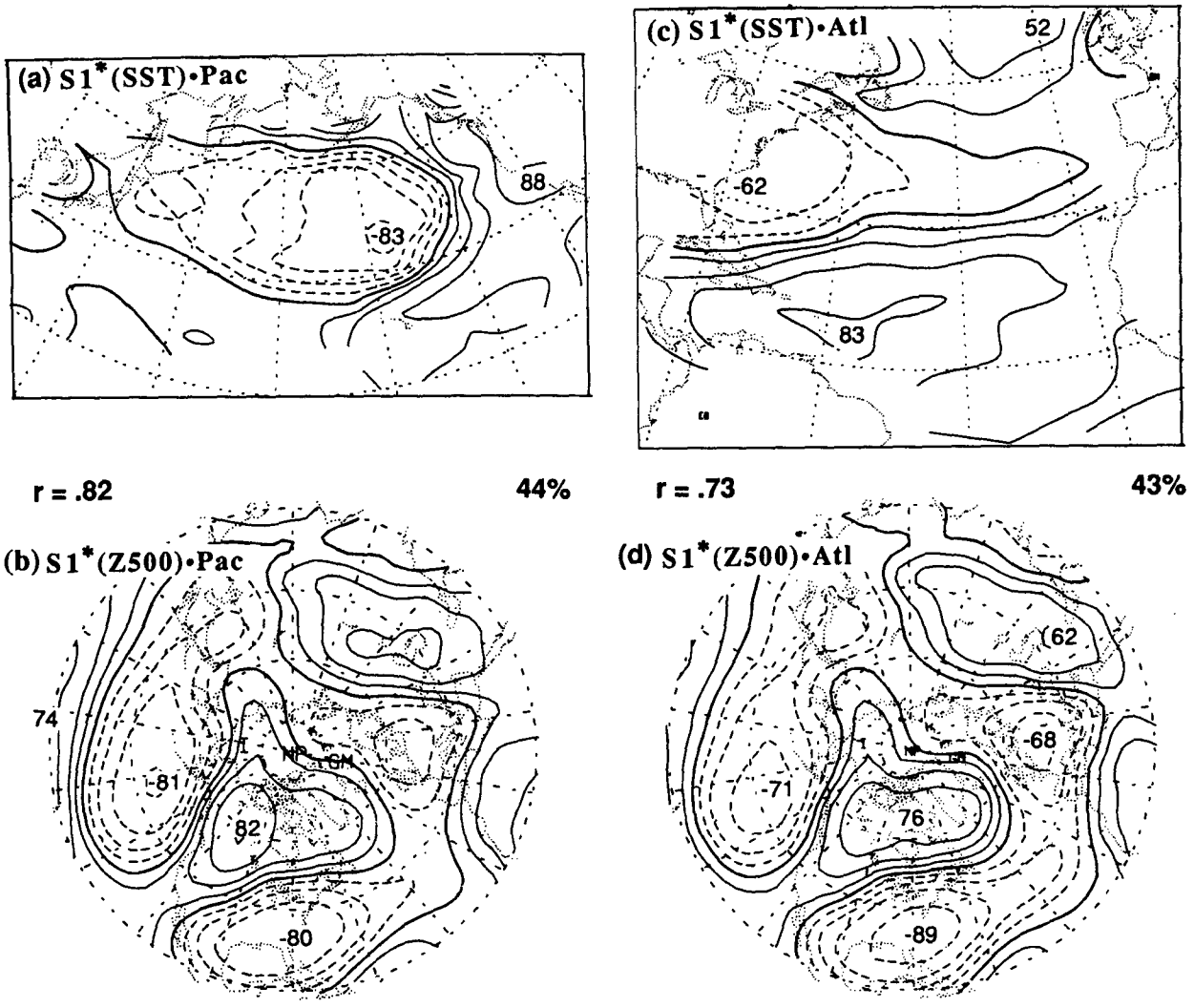


FIG. 19. Homogeneous correlation maps for the leading mode of two different direct SVD expansions based on the hemispheric 500-mb height field. The maps on the left are based on Pacific SST on the same grid as in the previous sections, and those on the right are based on North Atlantic SST on a 4° lat by 10° long grid poleward of 10°N. Labeling conventions as in Fig. 8.

structures within the fields are a reflection of large-scale atmosphere–ocean interaction.

In the expansions based on the hemispheric 500-mb height field, the SVD and CCA solutions proved to be similar, but for the ordering of the modes. For the more limited 500-mb height domain, the SVD and CCA solutions are somewhat different. The SVD solution for the Pacific sector is dominated by two of the leading hemispheric teleconnection patterns (PNA and WP), which are apparent in the first two EOFs of the hemispheric 500-mb height field, while the CCA modes more closely resemble the EOFs of the 500-mb height field for the Pacific sector, which are somewhat more distinct from the hemispheric patterns. The 500-mb height pattern in the leading mode of the CCA solution based on the three leading SVD modes bears an intriguing similarity to the pattern of anomalies observed in association with ENSO, even though the strongest and most coherent part of the ENSO "signal" in SST lies outside the domain of the analysis.

The trade-offs between the performance of SVD are evident from the tables presented in the paper. In the CCA solution based on the first three SVD modes, which appears to be the optimal number in the present analysis, $r(a_1, b_1)$ is substantially higher than in the first SVD mode (0.89 versus 0.80), but it explains only about half as much of the squared covariance between the fields as the first SVD mode (24% versus 52%).

The spatial patterns derived from the direct SVD and CPCA solutions are much more similar, and it is evident from the summary statistics presented in Table 6 that the CPCA expansion coefficients account for almost as much of the squared covariance between the two fields as the SVD expansion coefficients. As dis-

TABLE 6. Squared covariance fraction and cumulative squared covariate fraction for the five leading modes in the expansions based on singular value decomposition, canonical correlation analysis, and combined principal component analysis of wintertime SST and Z_{500} within the domains shown in Fig. 1.

	Individual modes (SCF)			Cumulative (CSCF)		
k	SVD	CCA	CPCA	SVD	CCA	CPCA
1	52	25	46	52	25	46
2	23	27	17	75	48	71
3	12	22	9	87	71	83
4	4	15	2	91	81	87
5	4	8	2	95	86	91

cussed in BSW, the solutions derived from the two methods differ in form: SVD yields expansion coefficients for each field, whereas CPCA yields only a single time series of expansion coefficients for each mode. The orthogonality characteristics of the modes are also somewhat different. It is conceivable that these differences could be exploited to learn more about the relationships between the two fields, but that is beyond the scope of the present paper.

A distinct advantage of CPCA and single field PCA relative to SVD and CCA is the fact that the solutions can be rotated in order to increase the degree of regional localization. As discussed by Richman (1986), Horel (1981), and Barnston and Livezey (1987), this practice can, in some situations, substantially improve the stability of the solutions and their resemblance to the patterns derived from one-point correlation maps. The left and right patterns derived from CCA and SVD cannot be rotated independently without losing the unique pairing that exists between them, and we are not aware of any method of rotating the patterns together in such a way that the pairing is retained.

Acknowledgments. We would like to thank Dr. Valentin Dymnikov of the Center for Numerical Mathematics of the USSR Academy of Sciences for bringing SVD to our attention and Ms. Yuan Zhang for her many helpful suggestions. This work was supported by the National Science Foundation through the Climate Dynamics Program under Grant ATM 8822872 and the Meteorology Program under Grant ATM 8858846.

REFERENCES

Barnston, A., and R. E. Livezey, 1987: Classification, seasonality and persistence of low-frequency atmospheric circulation patterns. *Mon. Wea. Rev.*, **115,** 1083–1126.

Bartlett, M. S., 1941: The statistical significance of canonical correlations. *Biometrika,* **32,** 29–38.

——, 1947: Multivariate analysis. *Suppl. J. Roy. Stat. Soc.,* **9,** 176–197.

Bretherton, C. S., C. Smith, and J. M. Wallace, 1992: An intercomparison of methods for finding coupled patterns in climate data. *J. Climate,* **5,** 541–560.

Davis, R., 1976: Predictability of sea surface temperature and sea-level pressure anomalies over the Northern Hemisphere. *J. Phys. Oceanogr.,* **6,** 249–266.

Esbensen, S. K., 1984: A comparison of intermonthly and interannual teleconnections in the 700 mb geopotential height field during the Northern Hemisphere winter. *Mon. Wea. Rev.,* **112,** 2016–2032.

Fletcher, J. O., R. J. Slutz, and S. D. Woodruff, 1983: Towards a comprehensive ocean–atmosphere data set. *Trop. Ocean–Atmos. Newslett.,* **20,** 13–14.

Graham, N. E., 1990: Seasonal relations between tropical Pacific SSTs and Northern Hemisphere 700-mb heights. Proc. of Fourteenth Annual Climate Diagnostics Workshop. NOAA Climate Analysis Center, 184–191.

Horel, J. D., 1981: A rotated principal component analysis of the interannual variability of the Northern Hemisphere 500-mb height field. *Mon. Wea. Rev.,* **109,** 2080–2092.

Hsu, H.-H., and J. M. Wallace, 1985: Vertical structure of wintertime teleconnection patterns. *J. Atmos. Sci.,* **40,** 1595–1612.

Kushnir, Y., and J. M. Wallace, 1989: Low-frequency variability in the Northern Hemisphere winter: Geographical distribution, structure and time dependence. *J. Atmos. Sci.,* **46,** 3122–3142.

Leith, C. E., 1973: The standard error of time averaged estimates of climatic means. *J. Appl. Meteor.,* **12,** 1066–1069.

Iwaska, N., K. Hanawa, and Y. Toba, 1987: Analysis of SST anomolies in the North Pacific and their relation to 500 mb height anomalies over the Northern Hemisphere during 1969–1979. *J. Meteor. Soc. Japan,* **65,** 103–113.

Kawamura, R., 1984: Relation between atmospheric circulation and dominant sea surface temperature anomaly patterns in the North Pacific during the Northern Winter. *J. Meteor. Soc. Japan,* **62,** 910–916.

Lanzante, J. R., 1984: A rotated eigenanalysis of the correlation between 700-mb heights and sea surface temperatures in the Pacific and Atlantic. *Mon. Wea. Rev.,* **112,** 2270–2280.

Mo, K. C., and R. E. Livezey, 1986: Tropical–extratropical geopotential height teleconnections during the Northern Hemisphere winter. *Mon. Wea. Rev.,* **114,** 2488–2515.

Namais, J., 1972: Space scales of sea-surface temperature anomalies and their causes. *Fish. Bull.,* **70,** 611–617.

——, 1973: Thermal communication between the sea surface and lower troposphere. *J. Phys. Oceanogr.,* **3,** 373–378.

——, 1975: Stabilization of atmospheric circulation patterns by sea surface temperature. *J. Mar. Res.,* **33**(Suppl.), 53–60.

——, 1978: Multiple causes of the North American abnormal winter 1976–77. *Mon. Wea. Rev.,* **106,** 279–295.

——, 1981: Teleconnections of 700 mb height anomalies for the Northern Hemisphere. *CalCOFI Atlas* No. 29, [Available from Scripps Institute of Oceanography, La Jolla, CA 92093.]

——, X. Yuan, and D. R. Cayan, 1988: Persistence of North Pacific sea surface temperature and atmospheric flow pattern. *J. Climate,* **1,** 682–703.

North, G. T., Bell, R. Cahalan, and F. Moeng, 1982: Sampling errors in the estimation of empirical orthogonal functions. *Mon. Wea. Rev.,* **110,** 699–706.

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

Smith, C., 1989: Singular value decomposition as a statistical technique in the study of the low frequency relationship between the atmosphere and ocean. M.S. thesis, University of Washington, 92 pp. [Available from University of Washington, Dept. of Atmospheric Sciences, Seattle, WA 98195.]

Wallace, J. M., and D. S. Gutzler, 1981: Teleconnections in the geopotential height field during the Northern Hemisphere winter. *Mon. Wea. Rev.,* **109,** 784–812.

——, and Q. Jiang, 1987: On the observed structure of the interannual variability of the atmosphere/ocean system. *Variability of the Atmosphere/Ocean System,* Roy. Meteor. Soc., 182 pp.

——, C. Smith, and Q. Jiang, 1990: Spatial patterns of atmosphere–ocean interaction in the northern winter. *J. Climate,* **3,** 990–998.

Weare, B. C., A. Navato, and R. E. Newell, 1976: Empirical orthogonal analysis of Pacific sea surface temperatures. *J. Phys. Oceanogr.* **6,** 671–678.