1. Introduction

To extract a dominant mode of variability from a meteorological field, Empirical Orthogonal Function (EOF) analysis is commonly used. To extract common variability from two fields, Singular Value Decomposition (SVD) analysis, also called Maximum Covariance Analysis (MCA), is employed as an extension of EOF analysis (e.g., von Storch and Zwiers 1999), which maximizes the covariance between two fields. Using the correlation matrix (instead of the covariance matrix used in SVD) results in the Canonical Correlation Analysis (CCA) method (e.g., Bretherton et al. 1992), which maximizes the correlation between two fields. These analysis methods are frequently used in meteorological research (e.g., Wallace et al. 1992).

Some studies have sought to extend the CCA theory to more than two fields (e.g., Kettenring 1971; Gifi 1990; Chen et al. 1994). For example, Kettenring (1971) proposed an analysis method that maximizes the sum of correlations among fields; this could include four other possible criteria to select pattern vectors. However, this analysis method requires solving the eigenvalue problem of a huge matrix whose row and column sizes are equal to the sum of the spatial dimensions of all fields (i.e., if the spatial size of each field is $N_i$, the eigenvalue problem to be solved involves a $\sum \sum_i N_i \times \sum_i N_i$ square matrix) (Gifi 1990). Note that solving the eigenvalue problem requires $O(N^3)$ floating point calculations for an $N \times N$ matrix (e.g., Golub and Van Loan 1996). The calculation thus becomes increasingly difficult as the number
of fields increases. In another approach, Chen et al. (1994) proposed a method to maximize a kind of product of correlations. Their method involves solving eigenvalue problems of matrices of the size of each field. This approach largely reduces the difficulty of solving problems with huge matrices; however, it still requires solving eigenvalue problems with matrices whose spatial dimensions are equal to those of each field, which are sometimes very large.

Here, we propose new basic methods as simple extensions of SVD analysis to include multiple fields. They do not require the calculation of eigenvalue problems, as they only require iteration of the products of some matrices. An important feature of this approach is that our methods are little affected by the spatial dimensions of the fields, even when they are large, thus offering a great advantage when dealing with various types of meteorological problems when using limited computational resources.

Before addressing the problem of multiple fields, let us review how EOF and SVD analyses are calculated. Normally, these calculations are performed by matrix algebra, as in solving eigenvalue problems. However, when dealing with large spatial datasets typical of climate problems, matrix algebra is computationally impractical. In this regard, Hirose and Kutzbach (1969) and Kuroda (1998) proposed alternative schemes that convert the EOF or SVD calculation to a matrix problem with a temporal array to reduce the size of the matrix array. Kuroda (1998) also proposed a method of simple iteration called the power method (e.g., Golub and Van Loan 1996). The conversion from a spatial to a temporal array is possible because the basic equations relating pattern vectors and time coefficients have complete symmetry, or a dual relationship (Kuroda 1998). Baldwin et al. (2009) have more recently proposed performing EOF analyses by this iteration technique. As EOF is a special case of SVD analysis, it is sufficient for us to start from the theory of SVD analysis and construct a new method to extract common variability from multiple datasets, called Multivariable Maximum Covariance Analysis (MMCA).

The outline of the paper is as follows. Section 2 presents the mathematical structure and techniques of SVD analysis. Section 3 offers two possible extensions of SVD analysis to more than two fields, one using the extreme of a summation of covariances and the other using the product of covariances. Section 4 offers a concrete example of MMCA, and Section 5 presents further discussion of the new method.

2. Theory of SVD analysis

SVD analysis, or MCA (von Storch and Zwiers 1999), is a statistical method to extract pattern vectors \( e_I (I = 1, 2) \), represented by column vectors of length \( N_I \), from two datasets represented by \((N_I \times N_J)\) matrices \( X_I (I = 1, 2) \) whose time coefficients have the largest covariance. The procedure is as follows. If we define time coefficients \( a_j \) associated with \( e_I \), represented by column vectors of length \( N_I \), as
\[
a_j = X_I e_I,
\]
then the requirement of the largest covariance is equivalent to the extreme condition of the cost function \( S \), defined by the expression
\[
S = e_I^T X_I^T X_J e_J - \frac{\lambda_1}{2} (e_I^T e_I - 1) - \frac{\lambda_2}{2} (e_J^T e_J - 1),
\]
where \((\cdot)^T\) represents the transpose of the matrix. The first term indicates the covariance of two time series of two fields, and the second and third terms are constraints with Lagrange multipliers of \( \lambda_I \), which require a unit vector condition of \( e_I \). The Euler–Lagrange equations for \( e_I \) are given by
\[
X_I^T X_J e_J = \lambda_I e_I,
\]
where \( I \neq J \). The solution of Eq. (3) can be obtained through matrix decomposition of \( X_I^T X_J \), hence the name SVD (e.g., Golub and Van Loan 1996). See Kuroda (1998) for more detail.

From Eq. (1), Eq. (3) can be rewritten as
\[
e_I = \frac{1}{\lambda_I} X_I^T \hat{a}_I,
\]
where associated time coefficients \( \hat{a}_I \) are defined by
\[
\hat{a}_I \equiv \sum_{j=1}^{2} C_{I,j} a_j,
\]
with \( C_{I,2} = C_{2,I} = 1 \) and \( C_{1,1} = C_{2,2} = 0 \). Note that Eqs. (1) and (4) indicate the dual relationship between pattern vectors and time coefficients; indeed, the two equations can be written identically if \( e_I \) and \( a_I \) are suitably scaled.

The dual relationship leads to a solution of the equations in the following way: Starting from arbitrary non-zero column vector \( A_I(1) \) and following the recurrence equation
\[
(1) \text{ Covariance is usually defined as the first term divided by } N_I, \text{ but we modify it as shown here for simplicity.}
\]
\[
(2) \text{ The unit vector can be a constant, but we set it at unity for simplicity.}
\]
\[ 
\hat{A}_2(k) = A_1(k) \\
E_2(k) = X^I_2 \hat{A}_2(k) \\
E_2(k)' = N[E_2(k)] \\
A_1(k) = X_1 E_1(k)' \\
\hat{A}_1(k) = A_1(k) \\
E_1(k) = X_1^T \hat{A}_1(k) \\
E_1(k)' = N[E_1(k)] \\
A_1(k+1) = X_1 E_1(k)', 
\]

with \( N[E] \equiv E/\sqrt{EE^T} \) being a normalization operator, then \( A_1(k) \) and \( E_1(k)' \) converge to \( a_1 \) and \( e_1 \), respectively, for \( k \to \infty \). In fact, if the normalization operation is ignored and \( E_1(k) \) is neglected in Eq. (6), we have

\[ 
A_1(k+1) = X_1 X_1^T X_1 X_1^T A_1(k). 
\]

The theory of the power method (e.g., Golub and Van Loan 1996) shows that the vectors \( A_1(k) \) approach the eigenvectors associated with the largest eigenvalue of the matrix \( X_1 X_1^T X_1 X_1^T \). This is equivalent to \( E_1(k)' \) in Eq. (6) converging to \( e_1 \), corresponding to the largest covariance \( \lambda_1 = \lambda_1' \).

Solving Eq. (7) using the power method is the same procedure that Kuroda (1998) proposed. As the array size of time is usually far smaller than that of space in typical climate problems, solving Eq. (7) is far more economical than solving the spatial equation of Eq. (3) or carrying out the SVD operation.

The recurrence equation of Eq. (6) can be constructed from the relation satisfied by exact solutions of Eqs. (1) and (4). The technique used is shown schematically in Fig. 1. Here the arrows indicate a matrix product from the dataset with the corresponding suffix, and the lines indicate simple replacement. By sequential counterclockwise operations, \( A_1 \) and \( E_1 \) approach their solutions except for their scaling.

Once \( e_1 \) and \( a_1 \) associated with the largest eigenvalue are obtained, components that incorporate them may be removed from the original data \( X_1 \), so that

\[ 
X_1' = X_1 - a_1 e_1^T, 
\]

and we can repeat the procedure on the deflated dataset. In such a way, higher modes may also be obtained. Here however, we focus on obtaining the solution with the gravest eigenmode for simplicity.

3. Extension of SVD analysis theory to multiple fields

3.1 The sum MMCA method

One possible method to extend SVD analysis is to extend the cost function expression of Eq. (2) to multiple fields. One natural extension to multiple fields is

\[ 
S = \sum_{i,j=1}^{n} \frac{1}{2} C_{i,j} e_i^T X_j e_j - \sum_{i=1}^{n} \frac{\lambda_i}{2} (e_i^T e_i - 1), 
\]

where \( C_{i,j} \) is a symmetric matrix and \( n \) is the total number of fields. The meaning of the cost function is a summation over the covariances between fields \( I \) and \( J \) with a strength of \( C_{i,j} \), and the second term is a constraint for unit vectors \( e_j \) (\( I = 1, 2 \ldots n \)). The Euler–Lagrange equations for \( e_j \) are given by

\[ 
\sum_{j=1}^{n} C_{i,j} X_j^T X_j e_j = \lambda_i e_i. 
\]

If time coefficients \( a_i \), associated with \( e_i \), are defined in the same form as in Eq. (1),

\[ 
a_i \equiv X_j e_j, 
\]

then the equation for \( a_i \) is obtained from Eq. (10) as

---

\[ \text{Footnote 1:} \] The normalization operation is optional and is introduced only to help restrict the length of vectors when desirable.

\[ \text{Footnote 2:} \] From Eq. (3), we have \( \lambda_i = a_i^T a_i \). Thus \( \lambda_i \) is a covariance of two fields.

\[ \text{Footnote 3:} \] The antisymmetric components of \( C_{i,j} \) in Eq. (9) vanish in the Euler–Lagrange equations due to the symmetries between \( e_i \) and \( e_j \).
where the associated time coefficient $\hat{a}_j$ is defined by

$$\hat{a}_j \equiv \sum_{j=1}^n C_{j,j} a_j.$$  

(13)

So the relationship between $e_I$ and $a_I$ is of the same form as in the SVD analysis in Section 2.

By analogy with the SVD analysis, the solution is obtained through the same type of recurrence equation:

$$\hat{A}_j(k) = \sum_{j=1}^n C_{j,j} A_j(k)$$

$$E_j(k) = X_j^T \hat{A}_j(k)$$

$$E_j(k)' = N[E_j(k)]$$

$$A_j(k+1) = X_j E_j(k)'$$

starting the calculation from arbitrary non-zero column vectors $A_j(1)$ ($I = 1, 2 \ldots n$). If this recurrence equation converges for $k \to \infty$, we obtain a solution to Eq. (10) that corresponds to the largest value of $S$. See Appendix for more detail. Note from Eq. (10) that the cost function in this solution has the value

$$2S = \sum_{j=1}^n \lambda_j = \sum_{j=1}^n C_{j,j} a_j^T a_j = \sum_{j=1}^n a_j^T \hat{a}_j.$$  

(15)

Thus, we seek solutions that have the largest summation of covariance between time coefficients and their associated coefficients for each field. We refer to this method as the sum MMCA method.

A schematic diagram for the application of Eq. (14) to three fields is shown in Fig. 2 with a typical simple choice of $C_{1,1} = C_{2,2} = C_{3,3} = 0$ and $C_{I,J} = 1$ for other matrices. Here the black arrows indicate matrix products from the dataset with the corresponding suffix and the gray arrows indicate linear calculation of associated time coefficients $\hat{A}_I$ from $A_J$, such as $\hat{A}_I = A_2 + A_3$. In this case, the operation of a black arrow must be performed before the operation of a gray arrow. By sequential two-step operations, $A_I$ and $E_I$ approach a solution except for their scaling.

For best use of this method, care should be taken to select values of $C_{I,J}$ so that covariances of every field have a similar scale. Otherwise, the largest contributor of some specific covariance will determine the solution.

### 3.2 The product MMCA method

Another expansion of the SVD analysis method starts from Eqs. (11) to (13), in which $C_{I,J}$ is a cyclic matrix with the form $C_{I+1,J} = 1$ for $I = 1$ to $I = n - 1$, $C_{1,n} = 1$, and $C_{I,J} = 0$ for other values, where $n$ is the number of fields. The recurrence equation in this case takes the form

$$\hat{A}_I(k) = A_n(k)$$

$$E_I(k) = X_I^T \hat{A}_I(k)$$

$$E_I(k)' = N[E_I(k)]$$

$$A_I(k) = X_I E_I(k)'$$

$$\hat{A}_I(k+1) = A_n E_I(k)'$$

(16)

where the calculation is started from an arbitrary non-zero column vector $A_1(1)$. In this case, we can show that $A_I(k)$ converges to the solution of $a_I$ for $k \to \infty$, because the equation for $A_I(k)$ can be ren-
dered as an eigenvalue problem if normalization is ignored:
\[ A_j(k+1) = X_jX_j^T \ldots X_jX_j^T X_{j+1}X_{j+1}^T \ldots X_{j+l}X_{j+l}^T A_j(k). \]  
(17)

This shows that \( A_j(k) \) approaches the direction of the eigenvector associated with the largest eigenvalue of matrix \( V_I \equiv X_IX_I^T \ldots X_IX_I^T X_{I+1}X_{I+1}^T \ldots X_{I+k}X_{I+k}^T \) for \( k \to \infty \) in accordance with the power method theory\(^6\).

The eigenvalue of matrix \( V_I \) is equal to \( U \equiv (a_I^Ta_I)(a_I^Ta_I)\ldots(a_I^Ta_I) \) for all \( V_I \) regardless of \( I \). We refer to this scheme as the product MMCA method as it selects the largest product of the covariances of time coefficients \( U_i \), in contrast to the summation method described in the previous section. Because it selects the maximum product, this method does not depend on the scaling of each dataset and may be easier to handle than the sum MMCA method.

A schematic diagram for Eq. (16) for the case of three fields is shown in Fig. 3. Here the arrows indicate a matrix product from the dataset with the corresponding suffix, and the lines indicate simple replacement. By sequential counterclockwise operations, \( A_j \) and \( E_j \) approach a solution except for their scaling.

4. Example

In this section, we demonstrate the use of the two MMCA methods with three actual meteorological datasets.

Although for simplicity we neglected spatial and temporal weighting in constructing the theory in Section 3, it is straightforward to include them, as Kuroda (1998) and Baldwin et al. (2009) have done. Inclusion of spatial weighting, in particular, is important to compensate for inhomogeneity of a dataset. For inclusion of spatial weighting, the only requirement is to include the diagonal matrix \( W \) for the product of two column vectors \( e \) and \( f \) with spatial array size \( N_S \):

\[ e^T f \rightarrow e^T W f = \sum_{i=1}^{N_S} e_i f_i w_i, \]
(18)

where \( w_i \) represents the spatial area represented by label \( i \). Note that spatial weighting can set the relative scale throughout the calculation (see Appendix of Kuroda 1998). The following example introduces spatial weighting to compensate for inhomogeneity of the dataset.

Our example involves the Arctic Oscillation (Thompson and Wallace 1998), a natural variability of sea-level pressure between the polar cap and middle-latitude belt in the Northern Hemisphere winter. It is produced through the interaction between zonal wind and eddies, and the observed surface pressure change is produced through meridional circulation driven by eddies (e.g., Kuroda 2005).

The three datasets we use are the zonal-mean zonal wind and the vertical and horizontal components of the E-P flux, which are calculated from daily data. We use monthly mean NCEP-DOI reanalysis data (Kanamitsu et al. 2002) for extended winter months (November to April) of each year from 1979 to 2007 (162 monthly values in total), rendered as anomalies by first removing monthly climatological values from the data. The domain of the dataset extends in latitude from 20°N to 80°N with 5° intervals and extends in irregular vertical levels from 1000 to 10 hPa. The data are not weighted according to latitude, but for the vertical direction, the data are weighted in proportion to the thickness of the levels. All data are then normalized by their climatological standard deviation, with the result that all data are made non-dimensional.

Figure 4a shows the result calculated from the sum...
Fig. 4. (a) Heterogeneous correlation maps of zonal wind $U$ (top), meridional E-P flux $EP_Y$ (lower left), and vertical E-P flux $EP_Z$ (lower right) showing the dominant mode of variability in the winter season extracted by the sum MMCA method. Numbers at the upper right indicate the percentage of variance explained by this mode. Numbers between panels indicate the correlation between the main time coefficients of each field ($r$) and the squared covariance fraction between them (percent). Contour interval is 0.1, and dashed lines indicate negative values. (b) Regression pattern of the dominant mode of variability. Arrows indicate E-P flux and contours indicate zonal-mean zonal wind. Contour interval is 1 m s$^{-1}$. Shading indicates area of 95% significance (correlation greater than 0.16).
MMCA method of Eq. (14) with \( C_{1,1} = C_{2,2} = C_{3,3} = 0 \) and \( C_{i,j} = 1 \) for other matrices, as in Fig. 2. The zonal wind pattern shows a meridional dipole structure with positive (negative) centers around 60°N (35°N) at 100 hPa and around 65°N (35°N) at the surface, with a node around 45°N. The meridional component of the E-P flux (EPY) is mostly negative throughout the troposphere with a peak around 55°N at 400 hPa. The vertical component of the E-P flux (EPZ) has a meridional dipole structure, with a positive center around 65°N in the lower troposphere and negative centers around 45°N in the lower troposphere and around 40°N at 100 hPa.

The regression pattern of the mode, which is obtained by multiplying the standard deviation of each field by each pattern vector in Fig. 4a, is shown in Fig. 4b. Zonal winds at high latitudes increase with height and peak at 5 m s\(^{-1}\) around 10 hPa; at lower latitudes, they peak around 1 m s\(^{-1}\) from the upper troposphere to 10 hPa. Enhanced upward propagation of waves occurs around 60°N near the surface, and equatorward propagation is enhanced around 50°N at 300 hPa. The overall pattern is very similar to those obtained in previous studies (e.g., Fig. 1 of Kuroda 2005).

The analysis also shows that the extracted components are very well correlated with each other: the correlation between the zonal wind and EPZ is 0.68, that between the zonal wind and EPY is 0.67, and that between EPY and EPZ is 0.76. The result shows that the mode explains a large part of the squared covariance fraction between the zonal wind and components of the E-P flux.

Figure 5 shows that the results of the product MMCA method using Eq. (16) are very similar to those of the sum MMCA method (Fig. 4). However, the variances explained by the mode are smaller than those obtained by the sum MMCA method, as are the squared covariance fractions between modes.

Except for these relatively small differences, both MMCA methods capture the important characteristics of Arctic Oscillation-type variability with very similar results. Both methods appear suitable for extracting common variability from multiple datasets.

5. Discussion

Our proposed analytical methods are natural extensions of SVD analysis or the MCA method. The results of trials with real datasets suggest that our methods are useful for extracting common variability from three or more fields.

Note that the result for the special case of our sum MMCA method (\( n \) standardized fields with \( C_{i,j} = 1 \) for \( i \neq j \) and \( C_{i,i} = 0 \)) should be precisely equivalent to that from the correlations summation method of Kettenring (1971). Our product MMCA method should also be equivalent to the method of Chen et al. (1994) if all considered fields are standardized before each calculation. Therefore, the greatest advantage of our proposed analysis methods compared with previous ones lies in their easier calculations rather than their proposing completely new analytical criteria.

There are \( n(n - 1)/2 \) combination pairs from \( n \) fields. The sum MMCA method can account for all of these by stipulating non-zero \( C_{i,j} \) matrices except for the diagonal components (which should be zero to exclude contributions from each field alone). However, the product MMCA method as presented in Section 3.2 includes only \( n \) pairs of combination. This can be far less than the total number of combinations if \( n \) is large. However, it is possible to include more covariance in the theory. If we consider an \( n \)-sided polygon, the original scheme of Eq. (16) shown in Fig. 3 corresponds to a circuit around the polygon apexes, \( 1 \rightarrow \cdots \rightarrow n \rightarrow 1 \). However, by simple extension, the circuit can include additional diagonal paths, such as \( 1 \rightarrow 3 \rightarrow 5 \cdots \). The problem becomes determining a circuit that connects all apexes of the polygon with one stroke. This is possible if \( n \) is an odd number according to the graph theory (e.g., Wilson 1996). Figure 6 shows the case of five fields. In this sense, a complete set of the product MMCA can be applied only for an odd number of datasets. Practically, however, the use of the complete set will not always be needed and only the simplest set of the product MMCA (3.8) will be sufficient for most of the analysis.

We have examined only the gravest eigenmode in this paper. However, we can also calculate higher eigenmodes by using the technique illustrated in Eq. (8). Calculations using both the sum and product MMCA methods show that eigenvectors associated with different eigenmodes, as well as time coefficients between different fields, are not orthogonal in general. This contrasts with the case of SVD analysis, in which these eigenvectors and time coefficients have complete orthogonality. In that respect, mode decomposition of climate data using the MMCA method will not be easy to handle. We believe that the MMCA method will be useful only for extracting the largest common variability from multiple datasets.

For the product MMCA method, we did not start from the cost function expression of Eq. (2), but from Eqs. (11) and (12). However, it is possible to create a cost function expression with the general form of \( C_{i,j} \) as
Fig. 5. Same calculation as in Fig. 4, using the product MMCA method.
choose between these methods in light of the details of their research problems.

**Appendix: Meaning of the solution of Eq. (14)**

If we set $E_j(k) = \beta_j(k)E_j(k)$ and $E_j(k)$ is deleted from Eq. (14), we have the following recursion equation:

$$A_j(k+1) = \beta_j(k)X_jX_j^T \sum_{j=1}^{n} C_{j,j}A_j(k).$$

(A1)

This equation can be put into the matrix form with a column (and row) size of $nN_i$ as

$$\bar{A}(k+1) = B(k)UC\bar{A}(k),$$

(A2)

where $\bar{A}(k) \equiv (A_1(k), A_2(k), \ldots, A_s(k))^T$. $B(k)$ and $U$ are diagonal block matrices with the form $B(k) \equiv \text{diag}(\beta_1(k)1, \beta_2(k)1, \ldots, \beta_s(k)1)$ with 1 being the $N_i \times N_i$ unit matrix and $U \equiv \text{diag}(X_1^T, X_2^T, \ldots, X_s^T)$, and $C$ is a block matrix in which its $(I, J)$ component is $C_{I,j}$. Let $\Lambda$ be diagonal block matrices with the form $\Lambda \equiv \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_s)$, where $\lambda_i$ is the $i$th largest eigenvalue of $A(k)$. If we take the power method theorem, for a large $k$ then $\bar{A}(k)$ will approach the eigenvector associated with the largest eigenvalue of $A(k)$, thus the largest eigenvalue $\lambda_1$ should be unity. On the other hand, we can write an equation for $a_j$ in the form

$$X_j^T \sum_{j=1}^{n} C_{j,j}a_j = \lambda_ja_j,$$

(A3)

which is derived from Eqs. (11) and (12). This can be put into the matrix form as

$$UC\tilde{a} = \Lambda\tilde{a},$$

(A4)

where $\tilde{a} \equiv (a_1, a_2, \ldots, a_s)^T$ and $\Lambda$ is a diagonal block matrix with the form $\Lambda \equiv \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_s)$. If we make a comparison between the eigenvalue equation of the matrix $B(\infty)UC$ and Eq. (A4), we have

$$\lambda_i = \sigma_m/\beta_i(\infty).$$

(A5)

Equation (A5) shows that the covariance $\lambda_i$ is proportional to the eigenvalue $\sigma_m$ and has the largest value when $m = 1$.

In considering the physical meaning of this solu-
tion, it should be noted that the basic Eq. (10) was derived from the extreme condition of the cost function expression Eq. (9). Thus, the solution should occupy one of the extreme values including the maximum of the functional Eq. (9). Recalling that the meaning of the cost function is the sum of covariances, and the iteration of Eq. (A1) converges to the solution with the largest eigenvalue $\sigma_1$, the solution should have the largest possible value of the cost function. This leads to the physical meaning of the solution, that is, the largest sum of covariances.

Acknowledgments

The authors are grateful to M. Inatsu of Hokkaido University and anonymous reviewers for their useful comments. This research was supported in part by Grants-in-Aid for Science Research (26287115, 26281016, 16H01184) from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

References