

Methods for finding coupled patterns

* Designed to introduce the concepts, but no HW on these

These methods are extensions from basic EOF/PC analysis

MCA: Maximum covariance analysis (also called SVD analysis)

EEOF: extended EOF analysis

CCA: Canonical correlation analysis.

MCA, SVD analysis

Case 1: SVD of the data matrix $A_{M \times N}$ (M -time, N -space)

$$A = U \Sigma V^T$$

The columns in U & V are the eigenvectors of AA^T and $A^T A$, respectively, or the PCs and EOFs of A .

The diagonal elements in Σ are the singular values. They are the square roots of the eigenvalues of the covariance matrix (C) where $C = A^T A$ or $C = AA^T$.

Case 2: SVD of a dispersion, or covariance, matrix ($A^T A$ or AA^T)

Before $\rightarrow A = U \Sigma V^T$

$C = A^T A$ or AA^T

In this case $\rightarrow C = V \Sigma^2 V^T$

The columns in V are the orthonormal eigenvectors of $A^T A$. Note: eigenvectors of C are the same as eigenvectors of $C^T C$.

The diagonal elements in Σ^2 have units of variance - they are the eigenvalues of $A^T A$ and AA^T

In MCA, apply SVD to the dispersion matrix between 2 different data sets.

$\rightarrow$ datasets have different state space, but a common sampling dimension (e.g. time)

Result of MCA

Patterns in one dataset whose time series explains the largest fraction of variance in the other dataset, or vice versa.

MCA used to isolate coupling between 2 physical ~~fields~~ fields.

MCA Mathematics

* Note:

Slightly
different
notation
in Hartmann.

2 data sets $X_{M \times N}$: left field
 $Y_{M \times L}$: right field

M = shared sampling dimension (e.g. time)
 N & L = structure dimensions.

1) Compute the dispersion matrix C

$$C = \frac{1}{M} X^T Y \rightarrow (C_{xy})$$

$$\frac{1}{M} \begin{matrix} & X^T & & & Y & & C \\ \begin{matrix} N \\ \downarrow \end{matrix} & \left[\begin{array}{c} \text{---} \\ \text{---} \\ \text{---} \end{array} \right] & \begin{matrix} M \\ \downarrow \end{matrix} & \left[\begin{array}{c} \text{---} \\ \text{---} \\ \text{---} \end{array} \right] & = & \begin{matrix} N \\ \downarrow \end{matrix} & \left[\begin{array}{c} \text{---} \\ \text{---} \\ \text{---} \end{array} \right] \\ & & M & & & & L \end{matrix}$$

Dispersion matrix

Inner dimension M (e.g. time)
is consumed by the multiplication.

2) Apply matrix operation SVD to C_{xy}

$$C_{xy} = U \Sigma V^T$$

C_{xy} is $N \times L$ U is $N \times N$ V^T is $L \times L$

- Columns in $U_{N \times N}$ (referred to as singular vectors) are the column space of C_{xy} . Correspond to structures in dataset X .
- Columns in $V_{L \times L}$ are the row space of C_{xy} and correspond to structures in Y .
- Time series of each pattern in U is found by projecting X onto each pattern in U . Similarly, Y is projected onto patterns in V .
- In contrast to EOF analysis, leading singular vector in U corresponds to regression of the left field (X) onto the leading time series in Y , and vice versa.
- Diagonal elements in Σ have units of covariance.
- Expansion coefficient time series are not mutually orthogonal.

Displaying the results

Two types of regression maps can be used:

- Homogeneous regression maps: regress (or correlate) the input data in the left field onto the expansion coefficient time series for the left field. Ditto for right field.

→ Show how singular vectors do in explaining the variance of their own dataset.

- Heterogeneous regression maps: regress (or correlate) ~~the~~ input data in the left field onto the exp. coef. time series for the right field, or vice versa.

→ The resulting maps are characteristic of MCA, as they are the patterns that explain covariance of the 2 data sets.

*⇒ Usually show heterogeneous regression maps!

If the spatial patterns that explain variance and covariance are similar, the maps should have similar patterns.

Statistical Significance of MCA

Question is similar to problem with EOFs.

Basically, does one mode stand out from another?

Fraction of the square covariance explained:

$$\text{1st mode} \quad \frac{\sigma_1^2}{\sum_{i=1}^n \sigma_i^2} = \text{Frac. exp.}$$

σ = diagonal elements of Σ

How well coupled are data in general?

Root mean square covariance (RMSC)

the total squared covariance is the square of the diagonal elements in Σ .

$$\text{RMSC} = \left(\frac{\sum_{i=1}^N \sum_{j=1}^L (\overline{x_i y_j})^2}{\sum_{i=1}^N \overline{x_i^2} \sum_{j=1}^L \overline{y_j^2}} \right)^{\frac{1}{2}} \rightarrow \sigma_i^2 \text{ from MCA}$$

$\rightarrow \sigma_i^2 \text{ from } \text{basic SVD}$

$\sim$ should be of order 0.1 for well correlated fields.

Other useful numbers: correlation between the leading time series $\bar{z}$

Caveats of MCA

- Significance is hard to assess
- MCA finds coupled patterns, but these do not necessarily reflect physical modes in the data.
- probably best to examine the EOFs of the datasets and their corresponding PCs for strong correlations.

CCA: Canonical Correlation Analysis

Basic idea: perform EOF analysis and MCA in sequence.

- EOF analysis is used to truncate the data sets (X & Y)
- MCA is used to find coupling patterns within the truncated data.

Advantage: Reduces the "noise" in the data, and hence less likely to be impacted by sampling variability.

* Here use Barnett and Preisendorfer (1987) method - see Wilks & von Storch-Zwiers for more details on others.

Treatment of input data

1st step : EOF analysis

- Perform EOF analysis of the original data for both the left and right fields
- Construct the time series of the PCs (EOF amplitudes) at each sampling time.

→ Essentially an orthogonal rotation.

2nd step : Truncation of data.

- Reduce # of degrees of freedom by truncating the data to a small number of EOF/PCs.
- The choice of the number of modes to be retained is tradeoff between explaining as much variance as possible vs. statistical significance. More PCs retained → more variance explained.

3rd step : Normalization

- Normalize each of the retained PCs for left and right fields so the $\hat{P}_c$ s have unit variance. ~~the $\hat{P}_c$ s~~ (i.e. $\hat{P}_c^2 = 1$)

4th step : Construct correlation matrix

- Using the truncated matrices of normalized PCs for the right and left fields, Construct covariance matrix

- Covariance matrix = correlation matrix since PCs have been normalized

5th step :

- Perform MCA on correlation matrix

- Singular values ^(σ) may be interpreted as correlation coefficients, or "canonical correlations"

- Singular vectors (U and V) correspond to dimensions $N \times N$ and $L \times L$, where N and L are ~~in the~~ ~~at~~ the truncated time values. — not the original time value.

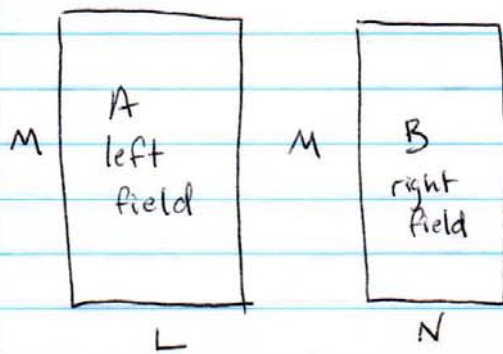
Step 6

- To get canonical correlation vectors, project the matrix of truncated PCs onto the singular vectors.

- Correlate canonical correlation vectors with input data to generate heterogeneous correlation maps. $\rightarrow$ Canonical patterns

CCA Schematics

Steps 2/3 - Data truncation, normalization



A & B = input
data matrices
of truncated PCs
(normalized)

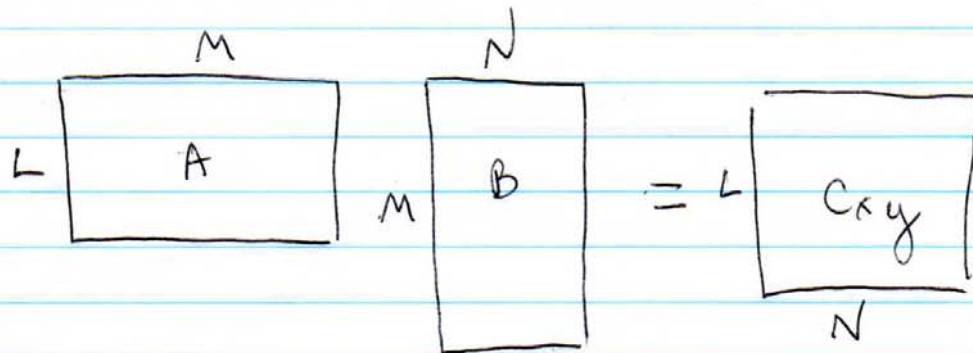
L = # of truncated PCs
in A

N = # of truncated
PCs in B .

$$N < L$$

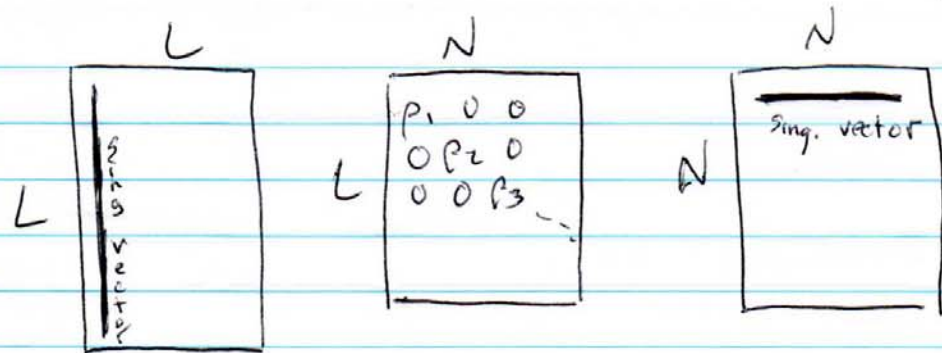
M = time dimension

Step 4 - Construct Correlation matrix



$$A^T B = C_{xy}$$

Step 5 - MCA on correlation Matrix C_{xy}



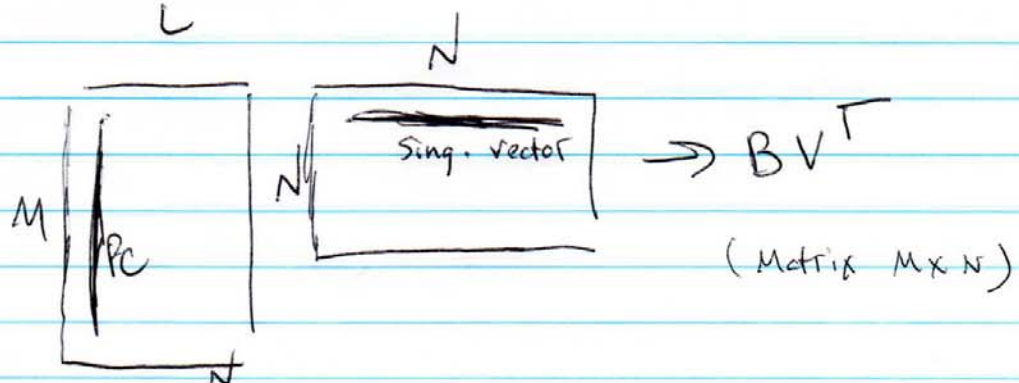
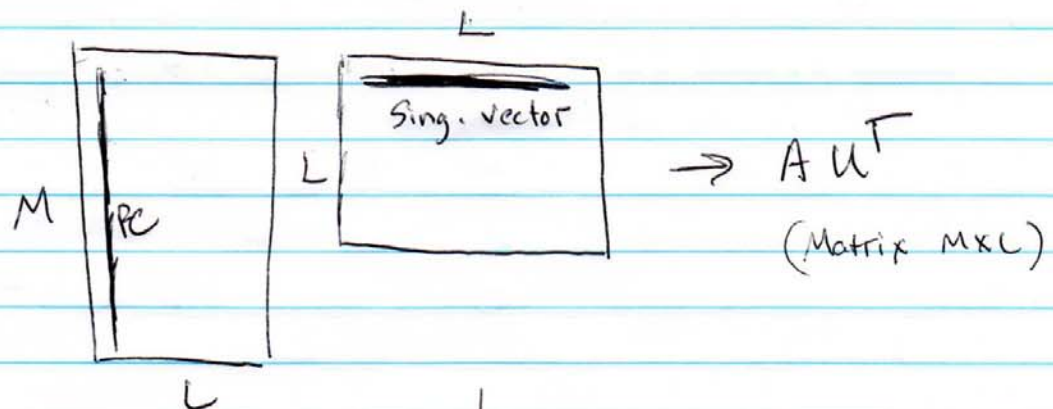
U matrix
Singular vectors
of A

P matrix
Canonical
correlations

V^T matrix
Singular Vectors
of B.

$$C_{xy} = U P V^T$$

Step 6 - Project singular vectors onto original data of truncated PCs (A & B matrices) to get canonical correlation vectors.



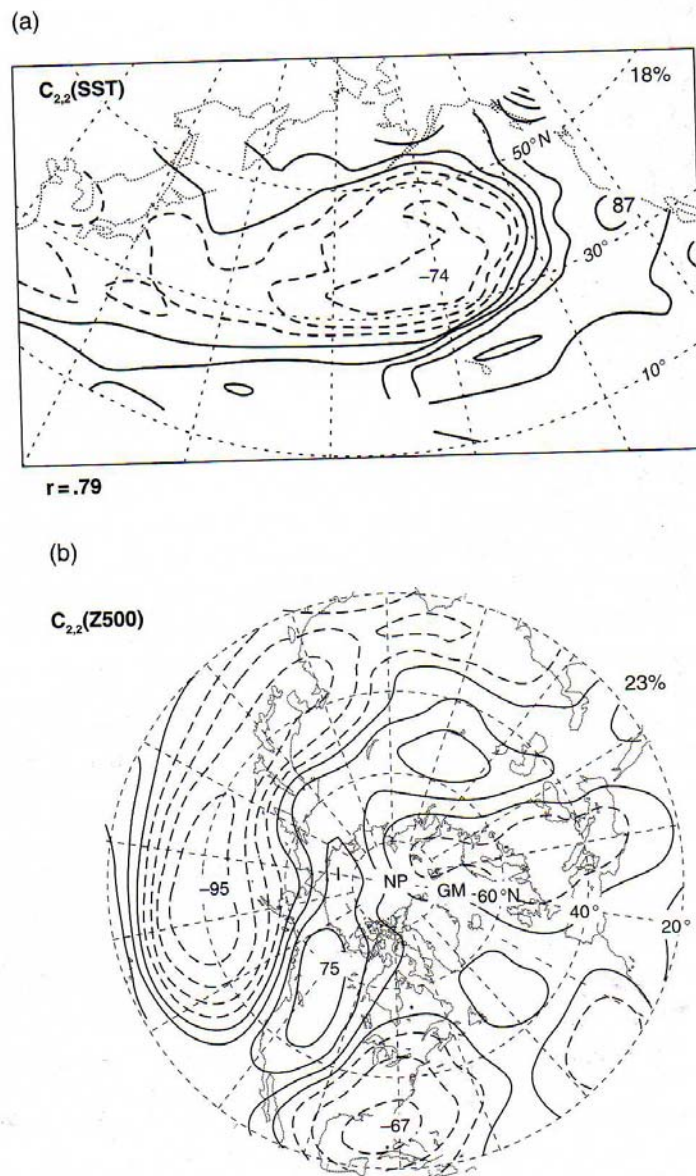


FIGURE 12.1 Homogeneous correlation maps for a pair of canonical variables pertaining to (a) average winter sea-surface temperatures (SSTs) in the northern Pacific Ocean, and (b) hemispheric winter 500 mb heights. The pattern of SST correlation in the left-hand panel (and its negative) are associated with the PNA pattern of 500 mb height correlations shown in the right-hand panel. The canonical correlation for this pair of canonical variables is 0.79. From Wallace *et al.* (1992).