EOF: Empirical Orthogonal Function
SVD: Singular Value Decomposition analysis
MCA: Maximum Covariance Analysis
PCA: Principal Component Analysis

(Factor analysis, Combined EOF, Combined PCA, Canonical Correlation Analysis, Complex EOF, Conditional MCA, etc.)

Shane Elipot
selipot@rsmas.miami.edu
MSC 316H
(305) 421-4609

A few references

Books:

Preisendorfer and Mobley (1988): *Principal Component Analysis in Meteorology and Oceanography*, Elsevier, New York. (RSMAS Library)

von Storch and Zwiers (2002): *Statistical Analysis in Climate Research*, Cambridge University Press, Cambridge (UM Library Online)

Emery and Thomson (2001): *Data analysis methods in physical oceanography*. Elsevier, second and revised edition (RSMAS Library)

Menke (2012), *Environmental data analysis with MatLab, (UM Library Online)*

Bendat and Piersol (1986), Random data. Analysis and measurements procedures, Third Edition, New York, (UM Library Online)

Papers:

Bretherton C, Smith C, Wallace J, *et al.* (1992). *An intercomparison of methods for finding coupled patterns in climate data*. J. Clim. 5(6): 541–560. (cited 725 times according to Web of Knowledge)

Venegas S, Mysak L, Straub D. (1996). *Evidence for interannual and interdecadal climate variability in the South Atlantic*. Geophys. Res. Lett. 23(19): 2673–2676.

Barnett T. 1983. *Interaction of the monsoon and pacific trade wind system at interannual time scales. I- the equatorial zone*. Monthly Weather Review 111: 756–773.

Merrifield M, Guza R. 1990. *Detecting propagating signals with complex empirical orthogonal functions: A cautionary note*. J. Phys. Oceanogr. 20(10): 1628–1633.

Introduction

The Preisendorfer approach:

Principal Component Analysis of a "multivariate" dataset

multivariate = N-variable dataset

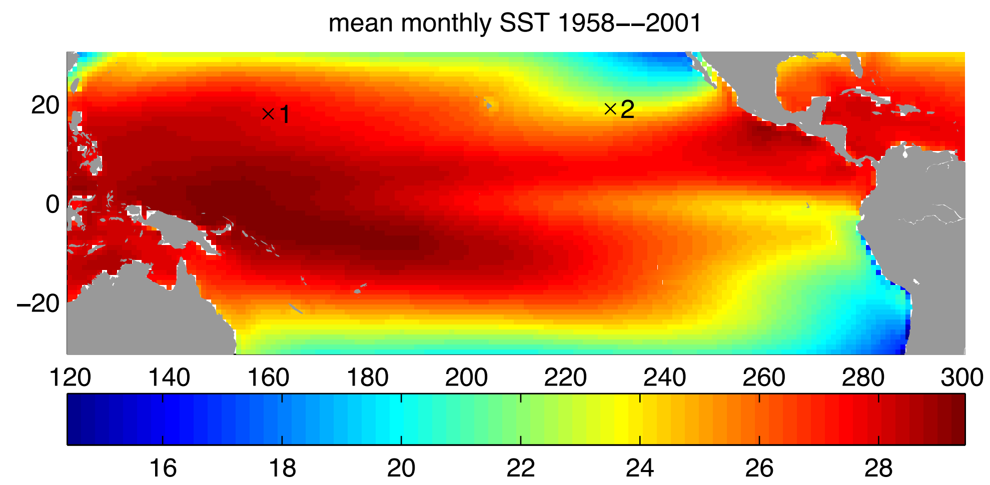
The Menke approach:

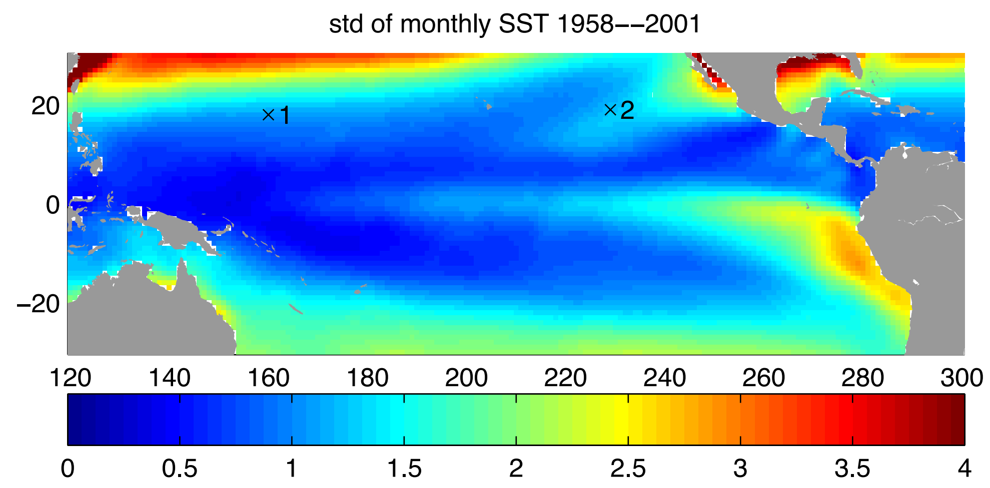
Factor Analysis

a "sample" is a linear "loading" of "factors"
or
The "data" can be decomposed into "modes"

Example

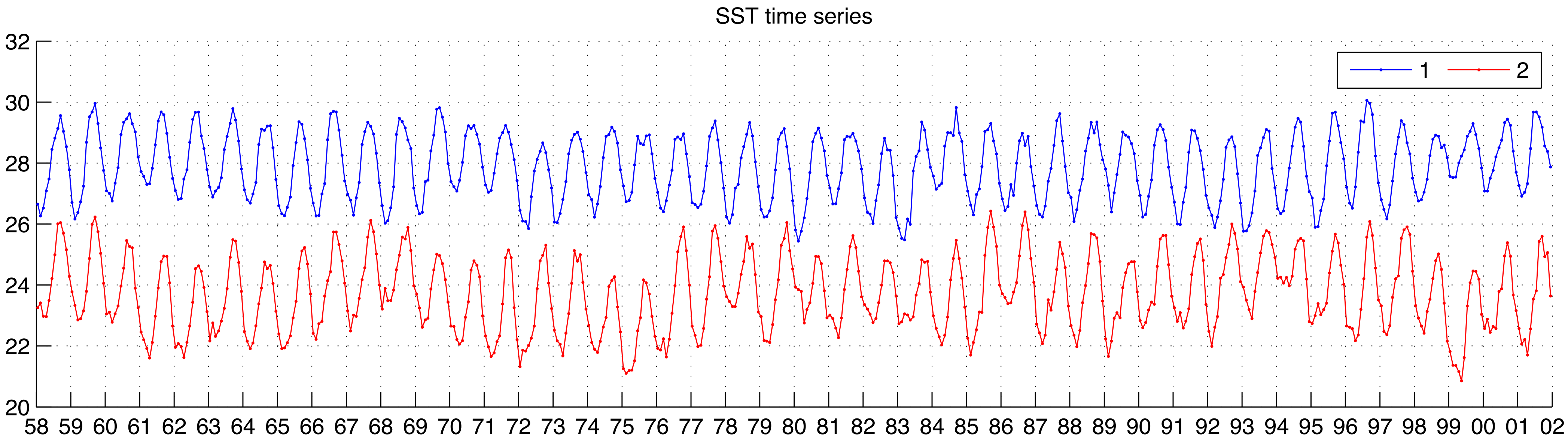
Tropical Pacific Sea Surface Temperature
ECMWF ERA-40 Reanalyses, 1° resolution, monthly values
Two locations; 1:(160°E,18°N) and 2:(131°W,19°N)



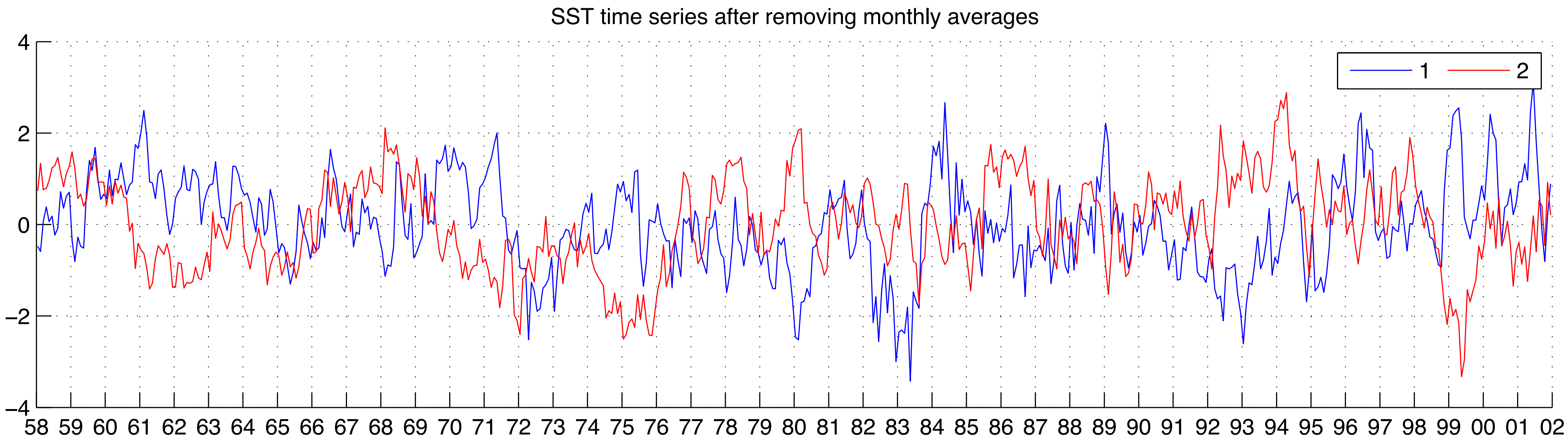


Example: “bivariate” dataset

$X_1 = SST(160°E, 18°N)$; $X_2 = SST(131°W, 19°N)$



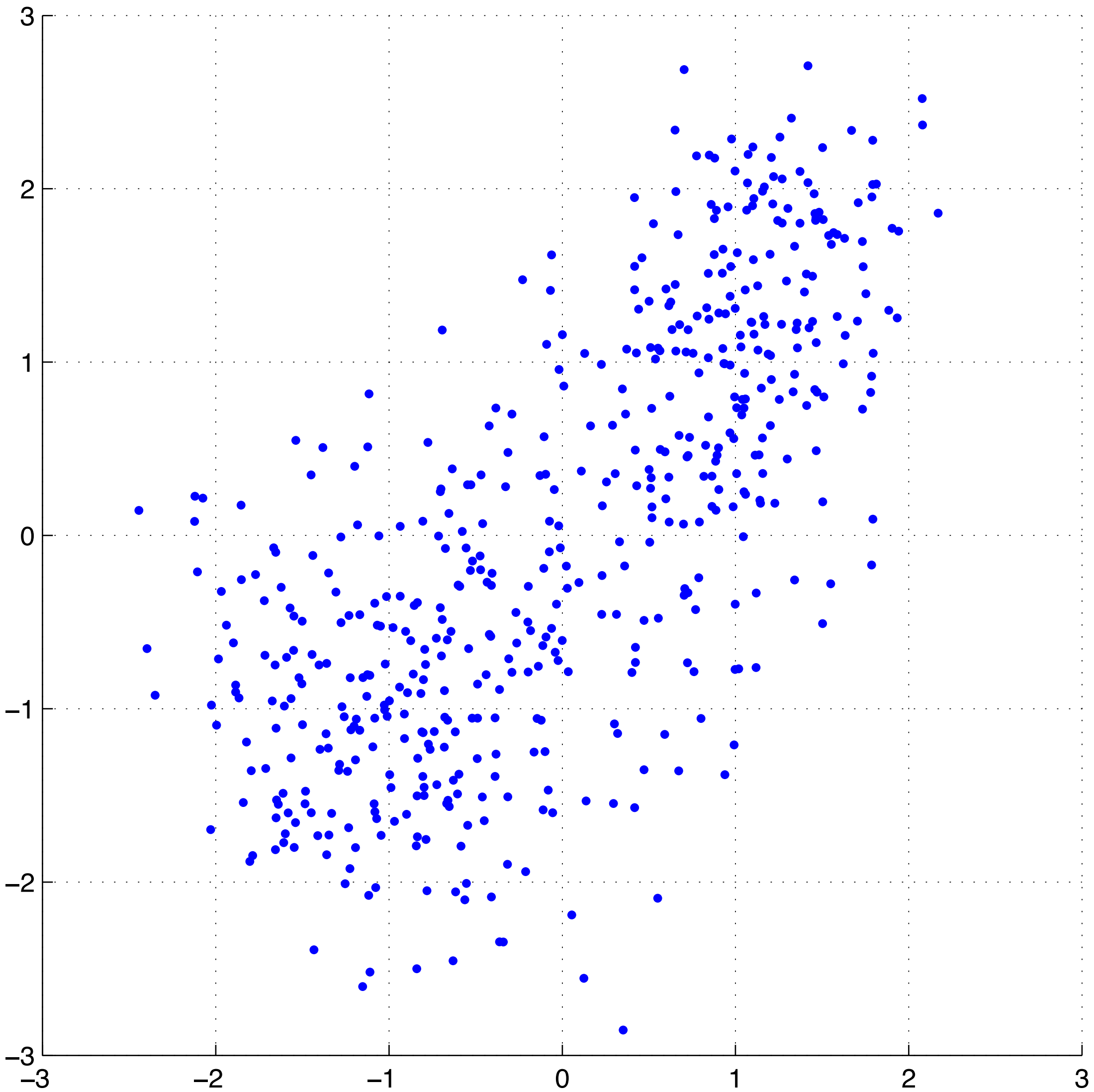
$$\rho = 0.70$$



$$\rho = -0.20$$

$N = 528$ data points

Scatter & statistics



Mean

$$\overline{X_j} \equiv \frac{1}{N} \sum_{n=1}^{N} X_j(n)$$

Anomaly

$$X_j'(n) = X_j(n) - \overline{X_j}$$

Covariance

$$S_{X_j X_k} \equiv \frac{1}{N-1} \sum_{n=1}^{N} X_j'(n) X_k'(n)$$

Auto-covariance or simply “variance”:

$$S_{X_1 X_1} \quad S_{X_2 X_2}$$

cross-covariance or simply “covariance”:

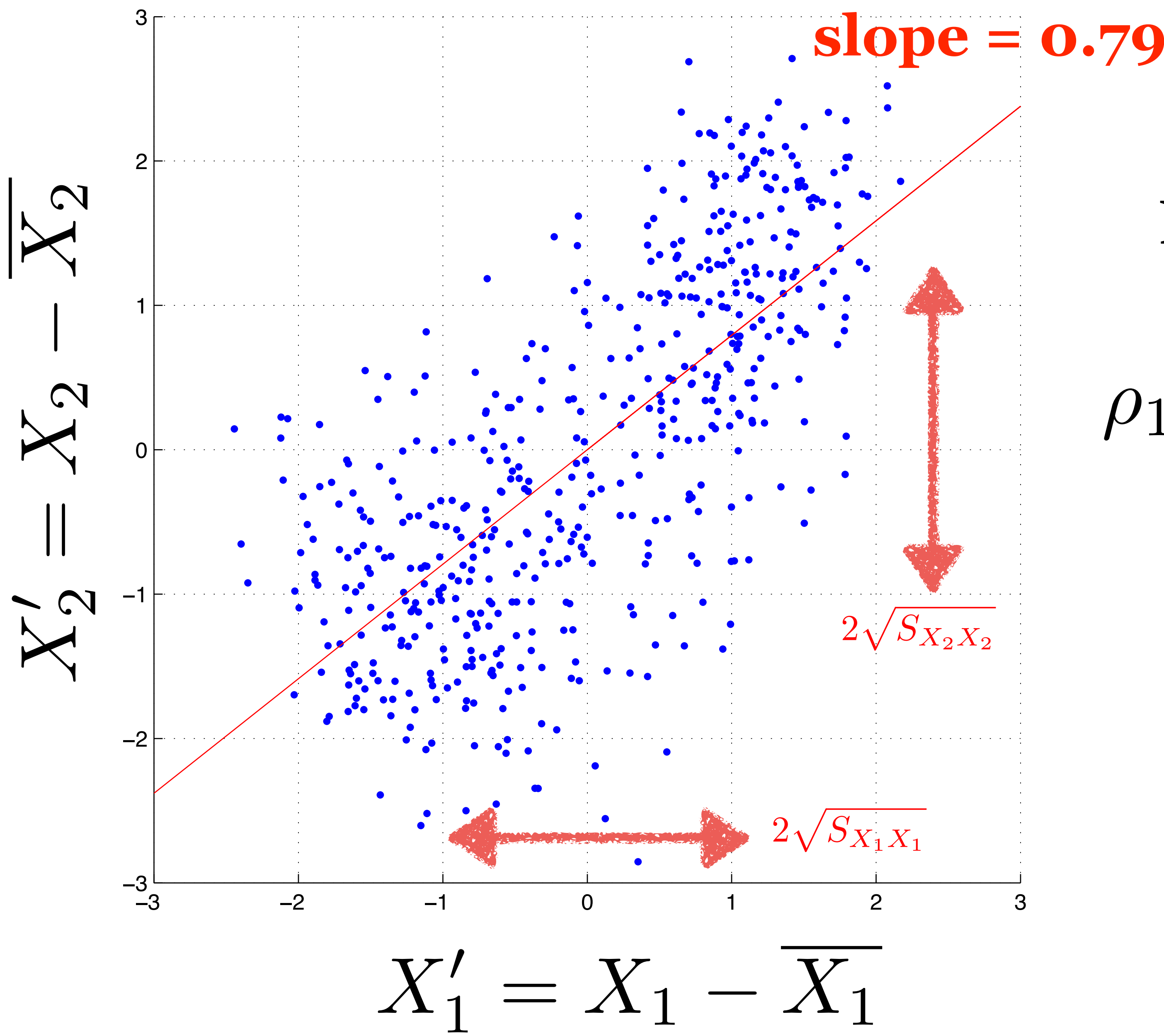
$$S_{X_1 X_2}$$

Scatter & statistics

$$\sqrt{S_{X_1X_1}} = 1.09$$
$$\sqrt{S_{X_2X_2}} = 1.23$$

Standard deviations

$$S_{X_1X_2} = 0.97^2$$ covariance



Note! correlation coefficient:

$$\rho_{12} = \frac{S_{X_1X_2}}{\sqrt{S_{X_1X_1}S_{X_2X_2}}} = 0.70$$

Note! regression coefficient of X_2 onto X_1:

$$\widehat{X_2'} = \alpha X_1'$$

"estimate" of X_2 by X_1

$$\alpha = \frac{S_{X_1X_2}}{S_{X_1X_1}} = 0.79$$

Matrix notation (forgetting the "primes")

Data matrix

$$\mathbf{X}_1 = \begin{bmatrix} X_1(1) \\ X_1(2) \\ \vdots \\ X_1(N) \end{bmatrix} \quad \mathbf{X}_2 = \begin{bmatrix} X_2(1) \\ X_2(2) \\ \vdots \\ X_2(N) \end{bmatrix} \quad \mathbf{X} = [\mathbf{X}_1\ \mathbf{X}_2] = \begin{bmatrix} X_1(1) & X_2(1) \\ X_1(2) & X_2(2) \\ \vdots & \vdots \\ X_1(N) & X_2(N) \end{bmatrix}$$

Note! Transpose = "switch" dimensions: $\mathbf{X}_1^T = \begin{bmatrix} X_1(1) & X_1(2) & \cdots & X_1(N) \end{bmatrix}$

covariance matrix

$$\mathbf{C}_{XX} = \frac{1}{N-1}\mathbf{X}^T\mathbf{X} = \frac{1}{N-1}[\mathbf{X}_1\ \mathbf{X}_2]^T[\mathbf{X}_1\ \mathbf{X}_2]$$

$$= \frac{1}{N-1}\begin{bmatrix} X_1(1) & X_1(2) & \cdots & X_1(N) \\ X_2(1) & X_2(2) & \cdots & X_2(N) \end{bmatrix} \begin{bmatrix} X_1(1) & X_2(1) \\ X_1(2) & X_2(2) \\ \vdots & \vdots \\ X_1(N) & X_2(n) \end{bmatrix}$$

$$= \frac{1}{N-1}\begin{bmatrix} \sum_{n=1}^N X_1(n)X_1(n) & \sum_{n=1}^N X_1(n)X_2(n) \\ \sum_{n=1}^N X_2(n)X_1(n) & \sum_{n=1}^N X_2(n)X_2(n) \end{bmatrix}$$

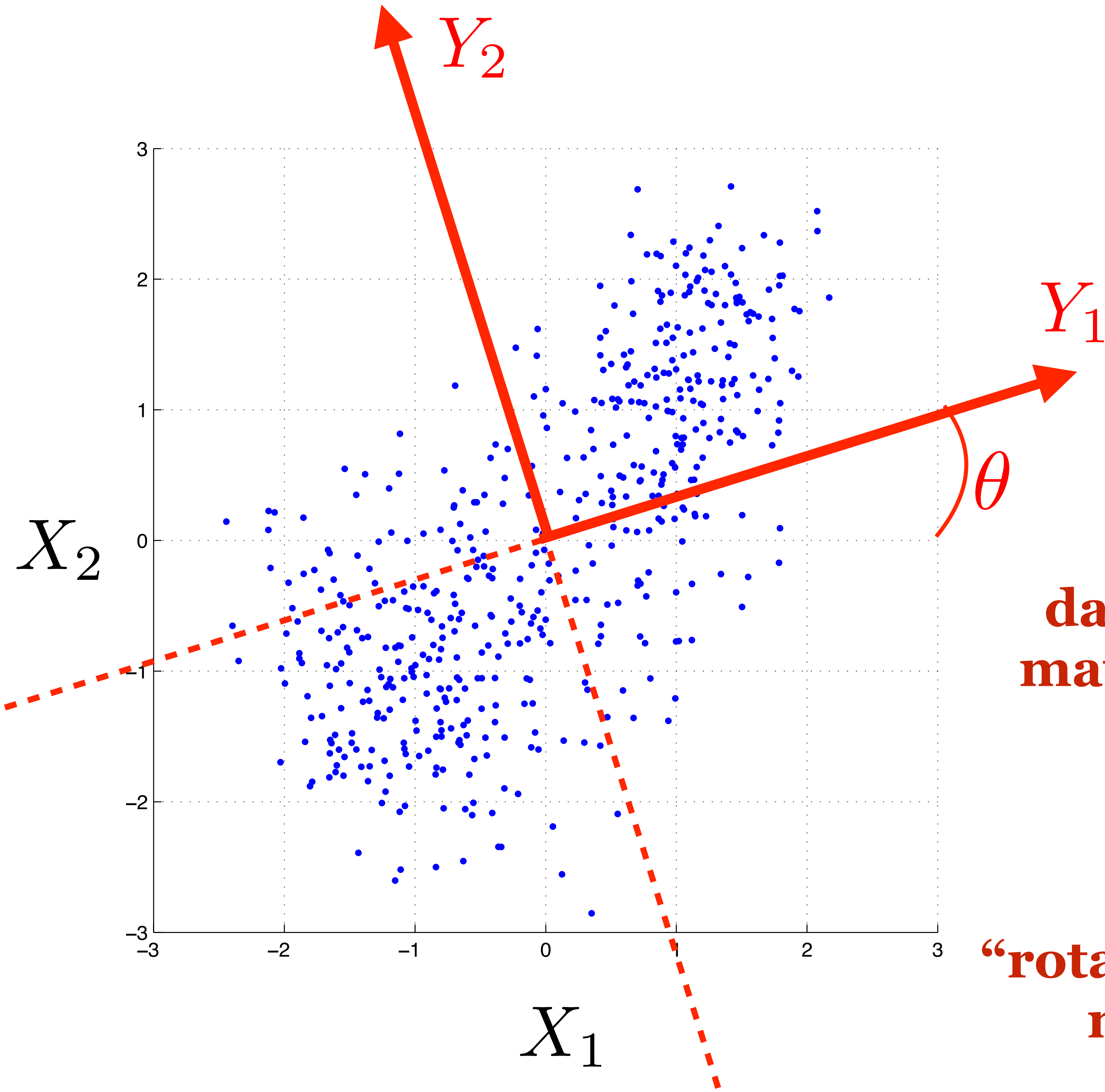
$$= \begin{bmatrix} S_{X_1X_1} & S_{X_1X_2} \\ S_{X_2X_1} & S_{X_2X_2} \end{bmatrix}$$

The covariance matrix
has a lot of
information!

scatter matrix $\mathbf{X}^T\mathbf{X}$

In Matlab from N×1 <u>column</u> vectors X1, X2: `C = cov(X1,X2);`

Rotation of frame of reference



New "coordinates"

$$Y_1(n) = X_1(n)\cos\theta + X_2(n)\sin\theta$$
$$Y_2(n) = -X_1(n)\sin\theta + X_2(n)\cos\theta$$

$$[\mathbf{Y}_1\ \mathbf{Y}_2] = [\mathbf{X}_1\ \mathbf{X}_2]\begin{bmatrix}\cos\theta & -\sin\theta \\ \sin\theta & \cos\theta\end{bmatrix}$$

data matrix

$$\mathbf{Y}(\theta) = \mathbf{X}\mathbf{R}(\theta)$$

"rotated" data matrix

"rotation" matrix

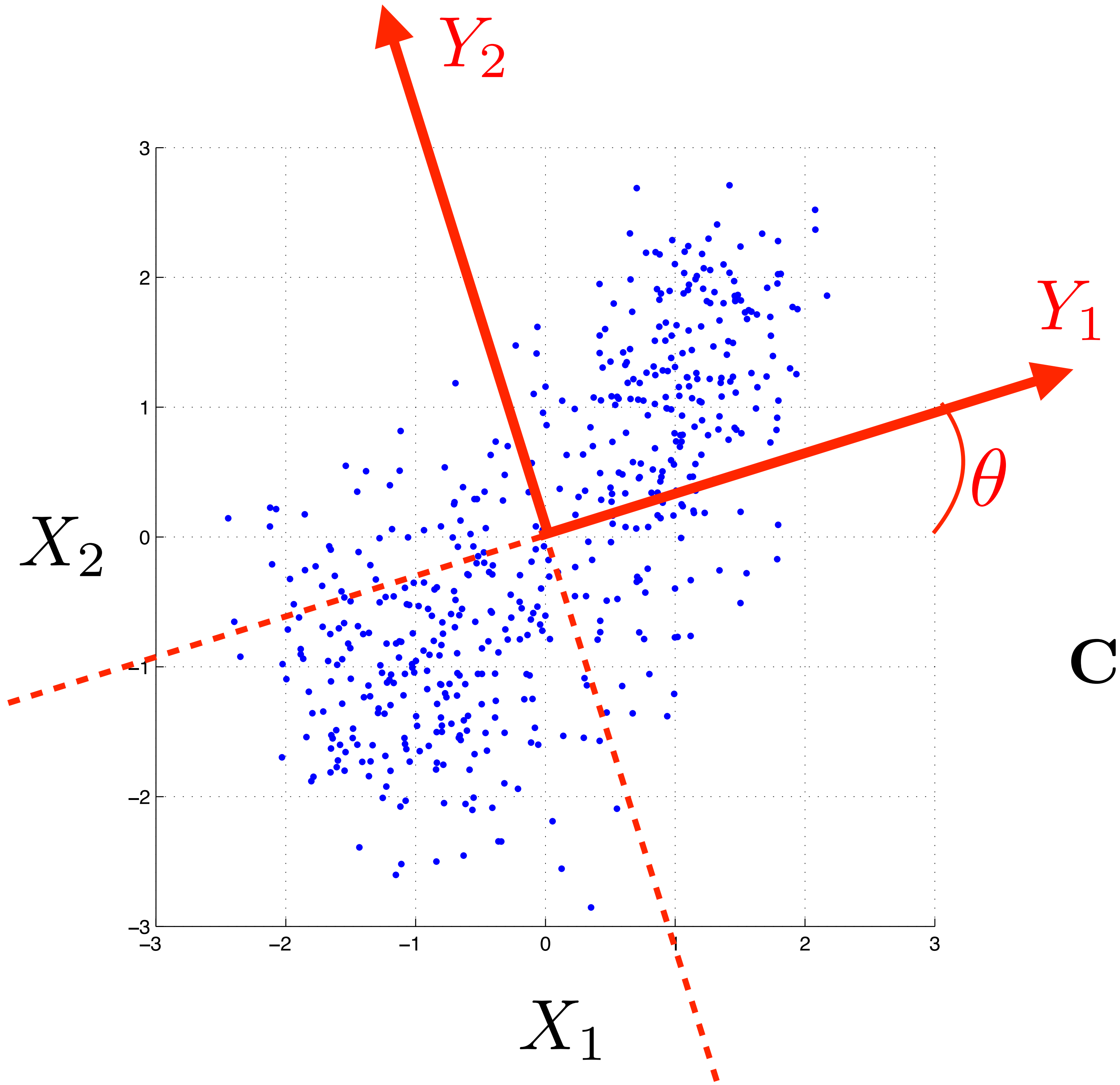
Rotation matrix property:

$$\mathbf{R}^{-1} \equiv \mathbf{R}^T$$

$$\mathbf{R}\mathbf{R}^{-1} = \mathbf{R}^{-1}\mathbf{R} = \mathbf{I} \equiv \begin{bmatrix}1 & 0 \\ 0 & 1\end{bmatrix}$$

Matlab: inv(R) or R^-1 NOT R.^{-1};

rotation of frame of reference



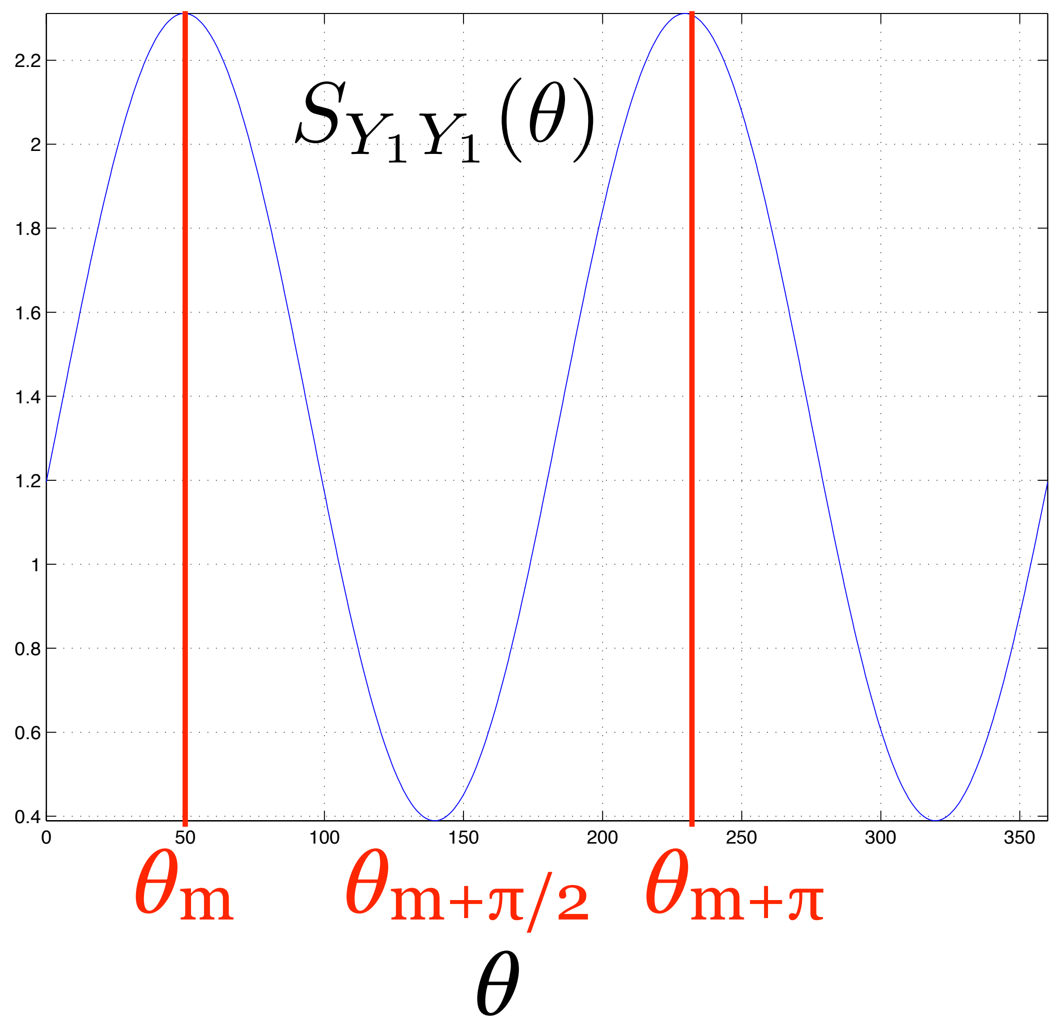
Covariance matrix for rotated data matrix:

$$\mathbf{C}_{YY}(\theta) = \frac{1}{N-1}\mathbf{Y}^T(\theta)\mathbf{Y}(\theta)$$
$$= \begin{bmatrix} S_{Y_1Y_1}(\theta) & S_{Y_1Y_2}(\theta) \\ S_{Y_2Y_1}(\theta) & S_{Y_2Y_2}(\theta) \end{bmatrix}$$

Auto-covariance of Y_1:

$$S_{Y_1Y_1}(\theta) = S_{X_1X_1}\cos^2\theta + 2S_{X_1X_2}\sin\theta\cos\theta + S_{X_2X_2}\sin^2\theta$$

Can we find a value θ_{m} of θ that maximizes $S_{Y_1 Y_1}$?



YES! The solution is such that:

$$\tan 2\theta_m = \frac{S_{X_1 X_2}}{S_{X_1 X_1} - S_{X_2 X_2}}$$

And the covariances are such that:

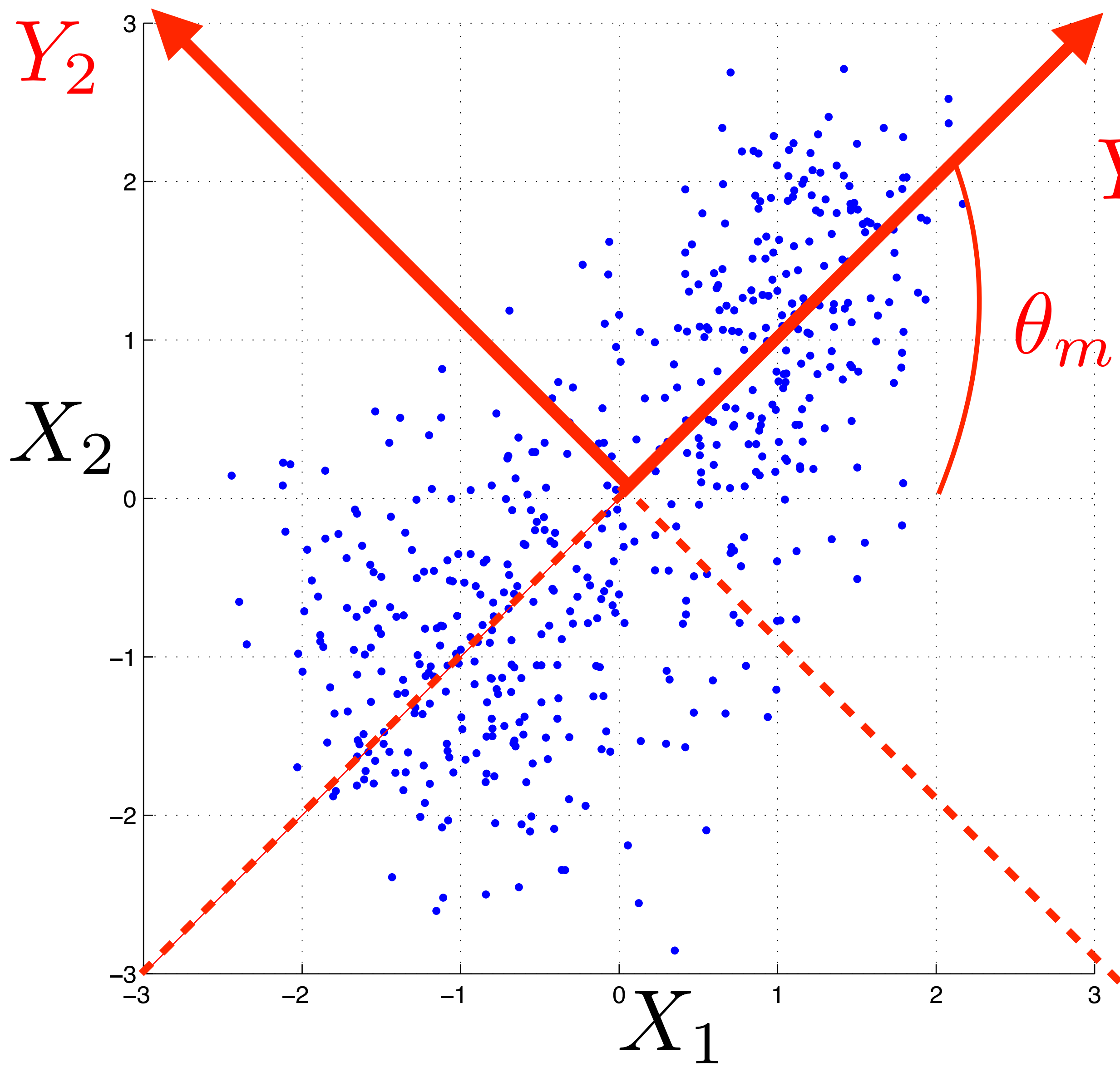
$$S_{Y_1 Y_1}(\theta_m) = \frac{1}{2}\left\{(S_{X_1 X_1} + S_{X_2 X_2}) + [(S_{X_1 X_1} + S_{X_2 X_2})^2 + 4S^2_{X_1 X_2}]^{1/2}\right\}$$ **maximum!**

$$S_{Y_2 Y_2}(\theta_m) = \frac{1}{2}\left\{(S_{X_1 X_1} + S_{X_2 X_2}) - [(S_{X_1 X_1} + S_{X_2 X_2})^2 + 4S^2_{X_1 X_2}]^{1/2}\right\}$$ **minimum!**

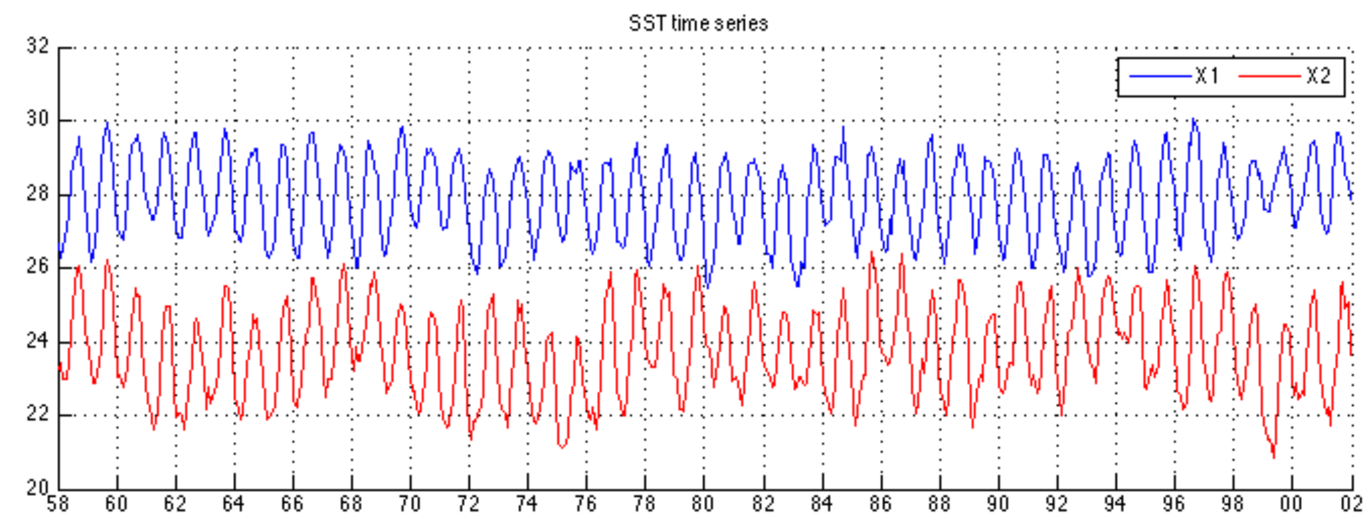
$$S_{Y_1 Y_2}(\theta_m) = 0$$ **zero!**

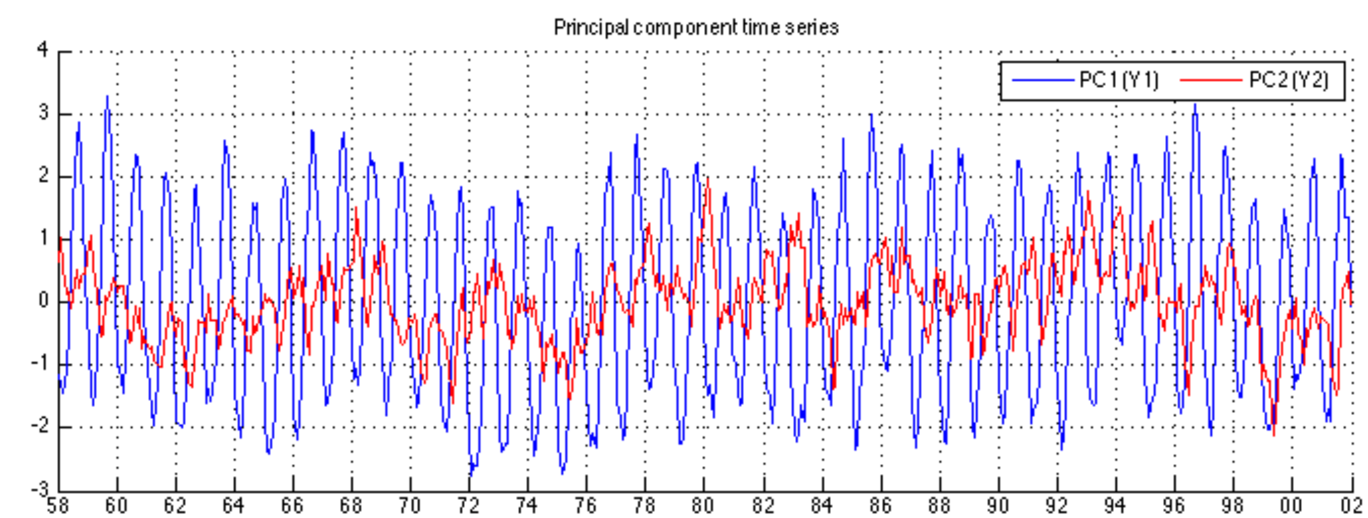
$$0 \le S_{Y_2 Y_2}(\theta_m) \le S_{X_1 X_1}, S_{X_2 X_2} \le S_{Y_1 Y_1}(\theta_m)$$

What do we get?



Principal component analysis provides an alternate representation of the variance of your data:





$$
\begin{aligned}
[\mathbf{Y}_1\ \mathbf{Y}_2] &= [\mathbf{X}_1\ \mathbf{X}_2]\,\mathbf{R}(\theta_m) \\
&= [\mathbf{X}_1\ \mathbf{X}_2] \begin{bmatrix} \cos\theta_m & -\sin\theta_m \\ \sin\theta_m & \cos\theta_m \end{bmatrix}
\end{aligned}
$$

$$
\mathbf{R}(\theta_m) = \begin{bmatrix} 0.65 & -0.76 \\ 0.76 & 0.65 \end{bmatrix}
$$

$$
\begin{aligned}
X_1(n) &= 0.65Y_1(n) - 0.76Y_2(n) \\
X_2(n) &= 0.76Y_1(n) + 0.65Y_2(n)
\end{aligned}
\qquad \rho_{X_1X_2} = 0.70
$$

$$\Leftrightarrow$$

$$
\begin{aligned}
Y_1(n) &= 0.65X_1(n) + 0.76X_2(n) \\
Y_2(n) &= -0.76X_1(n) + 0.65X_2(n)
\end{aligned}
\qquad \rho_{Y_1Y_2} = 0!
$$

Let's define:

$$\mathbf{R}(\theta_m) = \begin{bmatrix} \cos\theta_m & -\sin\theta_m \\ \sin\theta_m & \cos\theta_m \end{bmatrix} \equiv \mathbf{U}$$

2 "Eigen vectors" of Cxx

$$\mathbf{Y} = \mathbf{X}\mathbf{U}$$

2 "Principal components"

Covariance matrix for rotated data matrix:

$$\mathbf{C}_{YY} = \frac{1}{N-1}\mathbf{Y}^T\mathbf{Y} = \frac{1}{N-1}(\mathbf{X}\mathbf{U})^T(\mathbf{X}\mathbf{U}) = \frac{1}{N-1}\mathbf{U}^T(\mathbf{X}^T\mathbf{X})\mathbf{U} = \mathbf{U}^T\mathbf{C}_{XX}\mathbf{U}$$

or rewrite:

2 "Eigen values" of Cxx

$$\mathbf{C}_{XX} = \mathbf{U}\mathbf{C}_{YY}\mathbf{U}^T$$

$$= \mathbf{U}\begin{bmatrix} S_{Y_1Y_1} & 0 \\ 0 & S_{Y_2Y_2} \end{bmatrix}\mathbf{U}^T$$

"Diagonalization" of matrix or "eigen" decomposition

$$\mathbf{X} = \mathbf{Y}\mathbf{U}^{-1} = \mathbf{Y}\mathbf{U}^{T} \quad \Leftrightarrow \quad \mathbf{Y} = \mathbf{X}\mathbf{U}$$

principal components

PC1 PC2

$$\mathbf{Y} = [\mathbf{Y}_1\ \mathbf{Y}_2] = \begin{bmatrix} Y_1(1) & Y_2(1) \\ Y_1(2) & Y_2(2) \\ \vdots & \vdots \\ Y_1(N) & Y_2(N) \end{bmatrix}$$

modes

mode #1 mode #2

$$\mathbf{U} = \begin{bmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{bmatrix}$$

$$\mathbf{U}\mathbf{U}^{-1} = \mathbf{U}^{-1}\mathbf{U} = \mathbf{I} \equiv \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$X_1(n) = Y_1(n)U_{11} + Y_2(n)U_{12}$$

$$X_2(n) = Y_1(n)U_{21} + Y_2(n)U_{22}$$

Data is decomposed

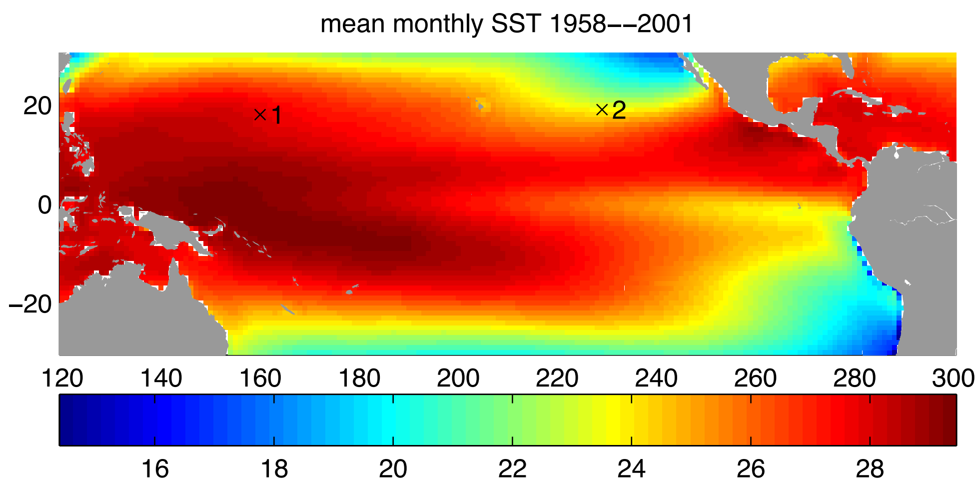
into loadings (the Ys) of factors (the Us)

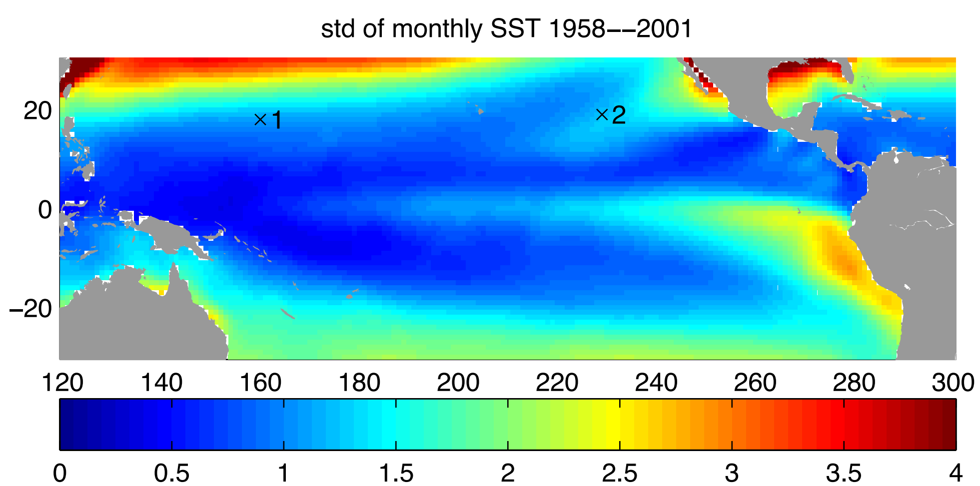
or

in principal components (the Ys) times modes (the Us)

Generalization to P data points: EOF!

P = 9560 (without land)





Data matrix is N×P

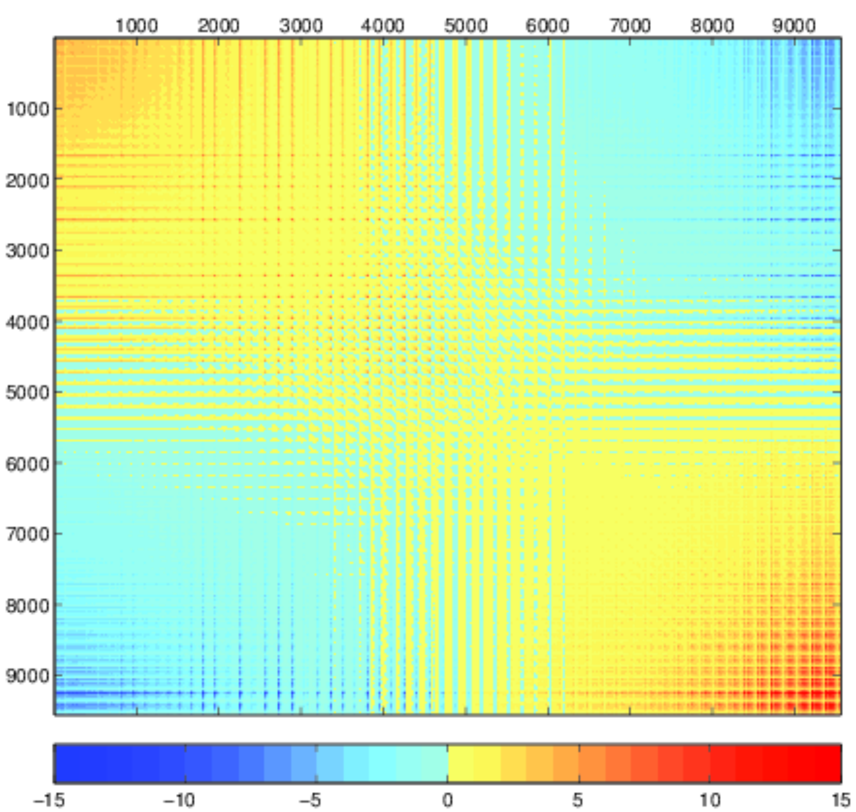
$$\mathbf{X} = [\mathbf{X}_1 \ \mathbf{X}_2 \ \cdots \mathbf{X_P}] = \begin{bmatrix} X_1(1) & X_2(1) & \cdots & X_P(1) \\ X_1(2) & X_2(2) & \cdots & X_P(2) \\ \vdots & \vdots & \cdots & \vdots \\ X_1(N) & X_2(N) & \cdots & X_P(N) \end{bmatrix}$$

covariance matrix is P×P symmetric

$$\mathbf{C}_{XX} = \frac{1}{N-1}\mathbf{X}^T\mathbf{X}$$

position index from 1 to P

position index from 1 to P



Calculate eigen decomposition:

$$
\begin{aligned}
\mathbf{C}_{XX} &= \mathbf{U}\boldsymbol{\Gamma}\mathbf{U}^T \\
&= \sum_{k=1}^{k=P} \gamma_k \mathbf{U}_k \mathbf{U}_k^T \\
&= \gamma_1 \mathbf{U}_1 \mathbf{U}_1^T + \gamma_2 \mathbf{U}_2 \mathbf{U}_2^T + \cdots + \gamma_P \mathbf{U}_P \mathbf{U}_P^T
\end{aligned}
$$

$$
\mathbf{U} = \begin{bmatrix} U_{11} & U_{12} & \cdots & U_{1P} \\ U_{21} & U_{22} & \cdots & U_{2P} \\ \vdots & \vdots & \cdots & \vdots \\ U_{P1} & U_{P2} & \cdots & U_{PP} \end{bmatrix} = \begin{bmatrix} \mathbf{U}_1 & \mathbf{U}_2 & \cdots & \mathbf{U}_P \end{bmatrix}
$$

P “Eigen vectors” of Cxx or EOF

$$
\boldsymbol{\Gamma} = \begin{bmatrix} \gamma_1 & 0 & \cdots & 0 \\ 0 & \gamma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \gamma_P \end{bmatrix}
$$

$$
\gamma_1 \geq \gamma_2 \geq \cdots \geq \gamma_p > 0
$$

P “Eigen values” of Cxx

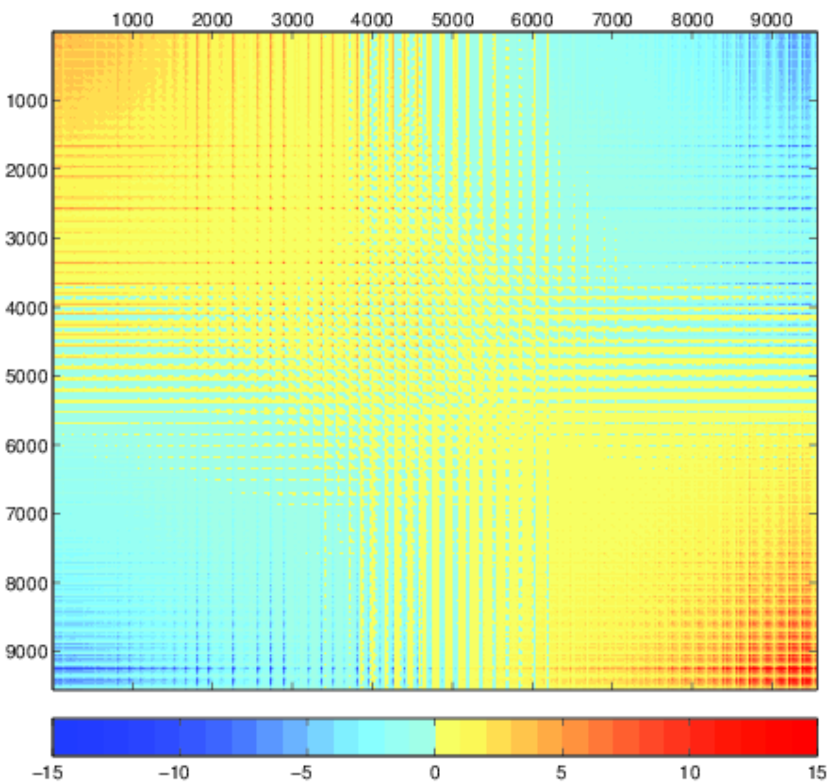
$$
\mathrm{scf}_m = \frac{\gamma_m}{\sum_{k=1}^{P} \gamma_k}
$$

“fraction” of variance explained (<1)

covariance matrix eigen decomposition:

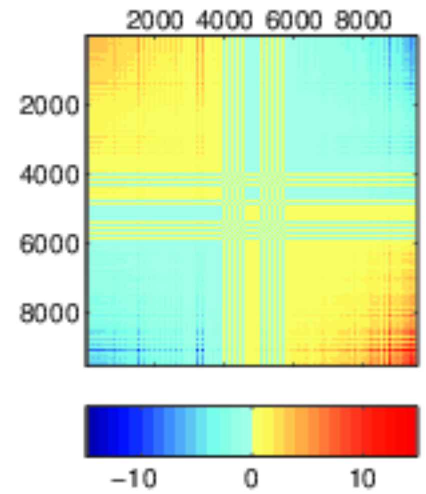
position index from 1 to P

position index from 1 to P



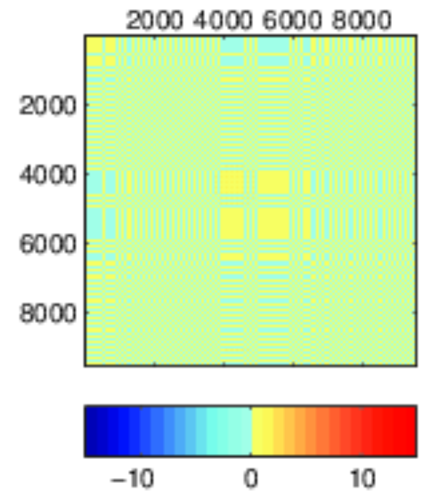
=

scf = 0.77

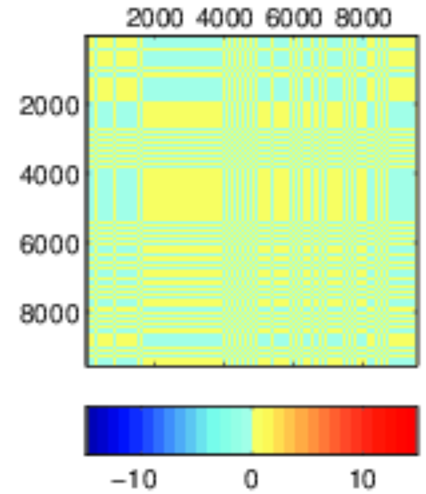


+

scf = 0.09

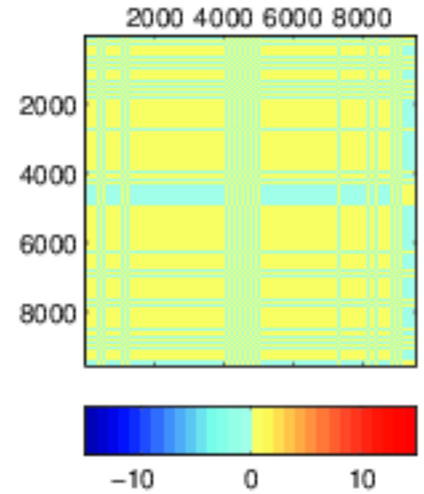


+...



scf = 0.03

+

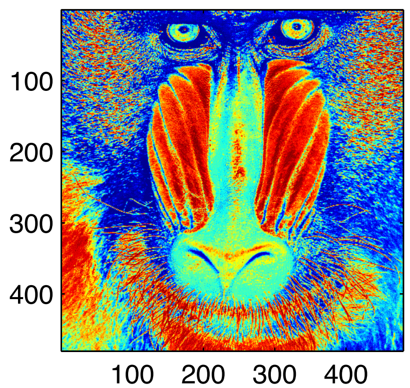


scf = 0.017

+9556 others...

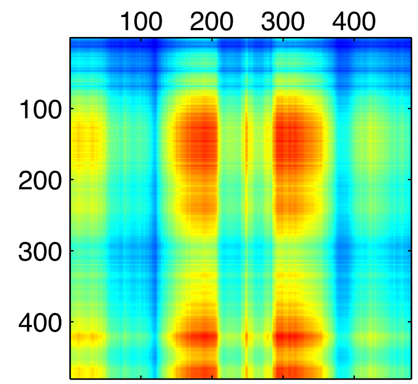
Digression: eigen decomposition of the Mandrill

load mandrill;
imagesc(X(:,1:480));

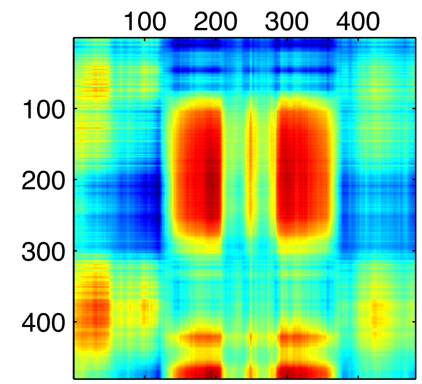


=

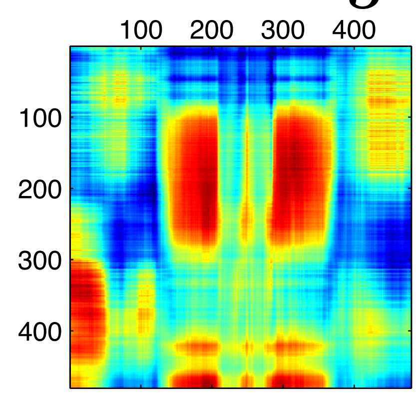
C1



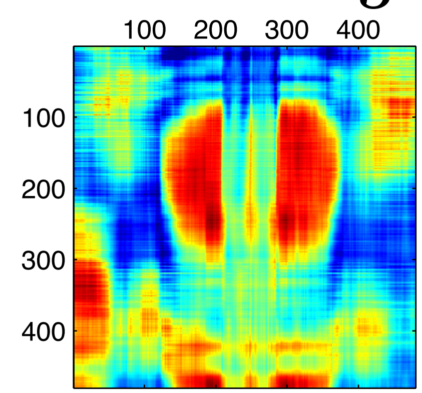
C1+C2



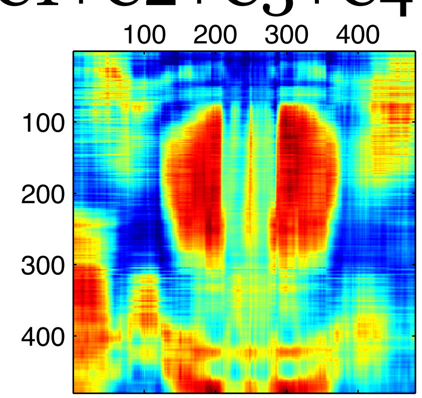
C1+C2+C3



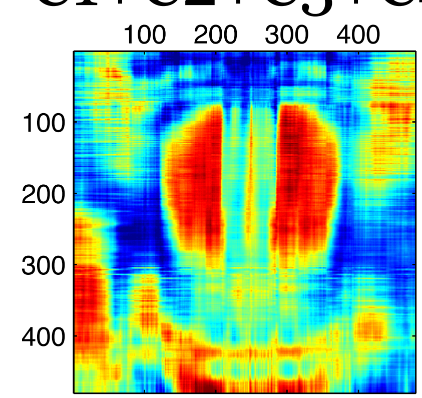
C1+C2+C3+C4



C1+C2+C3+C4+C5



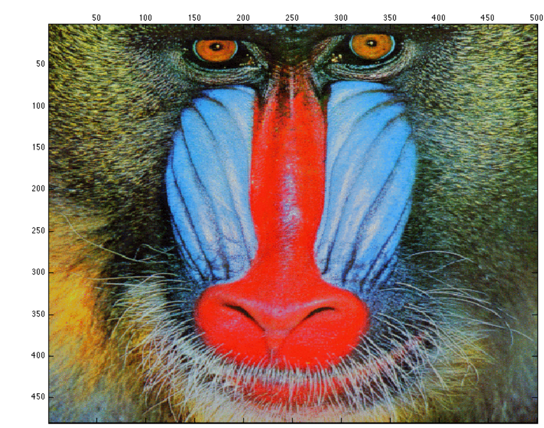
C1+C2+C3+C4+C5+C6



+ ... (474 others)

Try this:

load mandrill;
image(X);
colormap(map);



Back to SST in the Pacific:

Calculate eigen decomposition: $\mathbf{C}_{XX} = \mathbf{U}\boldsymbol{\Gamma}\mathbf{U}^T$

Calculate principal components: $\mathbf{A} = \mathbf{X}\mathbf{U}$

$$\mathbf{U} = \begin{bmatrix} U_{11} & U_{12} & \cdots & U_{1P} \\ U_{21} & U_{22} & \cdots & U_{2P} \\ \vdots & \vdots & \cdots & \vdots \\ U_{P1} & U_{P2} & \cdots & U_{PP} \end{bmatrix} = \begin{bmatrix} \mathbf{U}_1 & \mathbf{U}_2 & \cdots & \mathbf{U}_P \end{bmatrix}$$

P "Eigen vectors" of Cxx or EOF

$$\boldsymbol{\Gamma} = \begin{bmatrix} \gamma_1 & 0 & \cdots & 0 \\ 0 & \gamma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \gamma_P \end{bmatrix}$$

$$\mathbf{A} = \begin{bmatrix} A_1(1) & A_2(1) & \cdots & A_P(1) \\ A_1(2) & A_2(2) & \cdots & A_P(2) \\ \vdots & \vdots & \cdots & \vdots \\ A_1(N) & A_2(N) & \cdots & A_P(N) \end{bmatrix}$$

P "Eigen values" of Cxx

P principal components

$$\mathrm{scf}_m = \frac{\gamma_m}{\sum_{k=1}^{P} \gamma_k}$$

"fraction" of variance explained (<1)

Note! PCA property:

$$\mathbf{A}^T\mathbf{A} = (N-1)\boldsymbol{\Gamma}$$

means the PC are uncorrelated!

Back to SST in the Pacific:

Data matrix → $$\mathbf{X} = \mathbf{A}\mathbf{U}^T$$

$$= \sum_{k=1}^{P} \mathbf{A}_k \mathbf{U}_k^H$$ ← Sum of P modes

Time series at one location is the sum of P time series

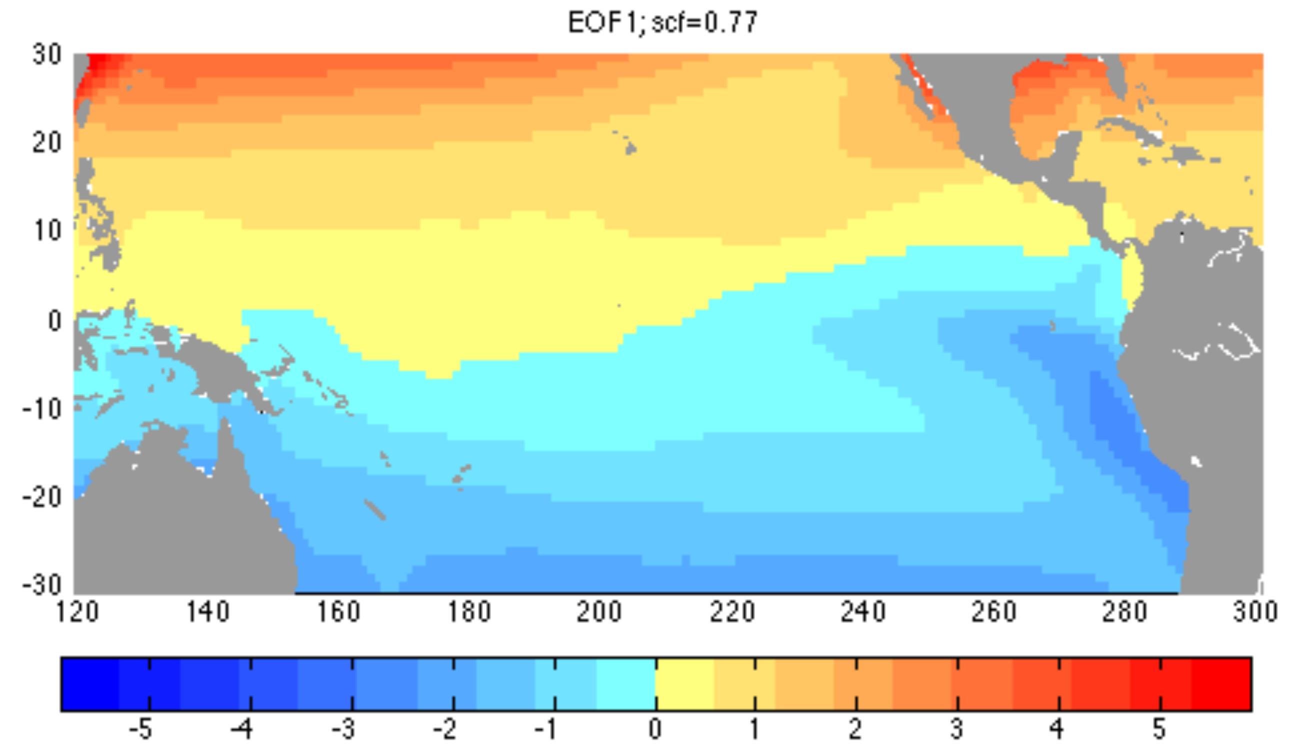
$$X_j(t) = \sum_{k=1}^{P} X_j^k(t) = \sum_{k=1}^{P} A(n) U_{jk}$$

One can study one single mode time series ...

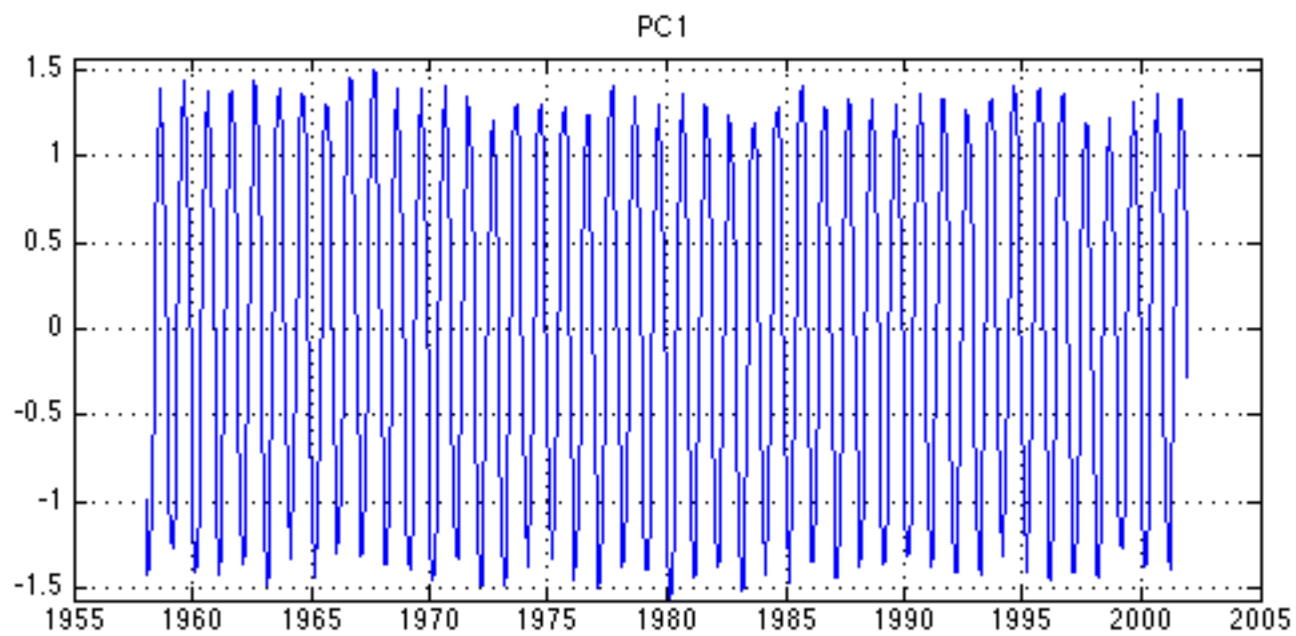
$$X_j^k(t) = A_k(n) U_{jk}$$

EOF1 "explains" 77% of the variance

First eigen vector U_1 or "EOF1" reorganized into a map:



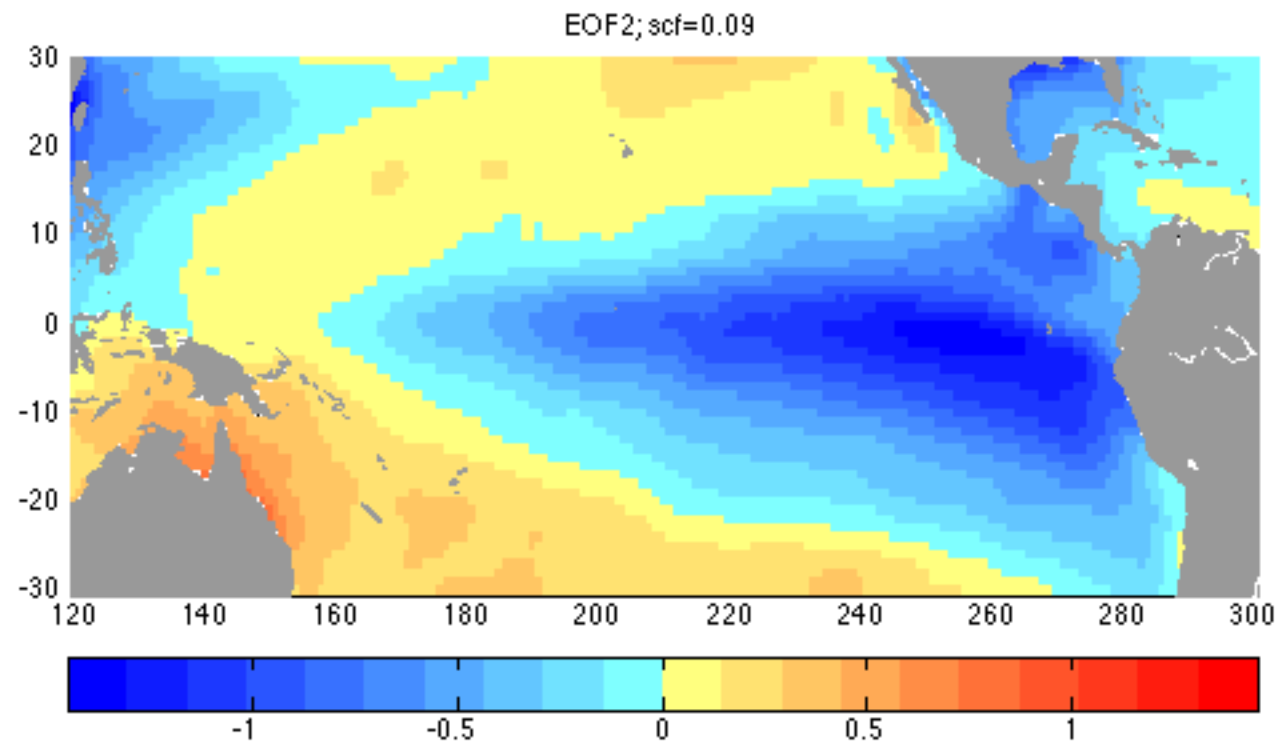
First principal component time series (PC1)



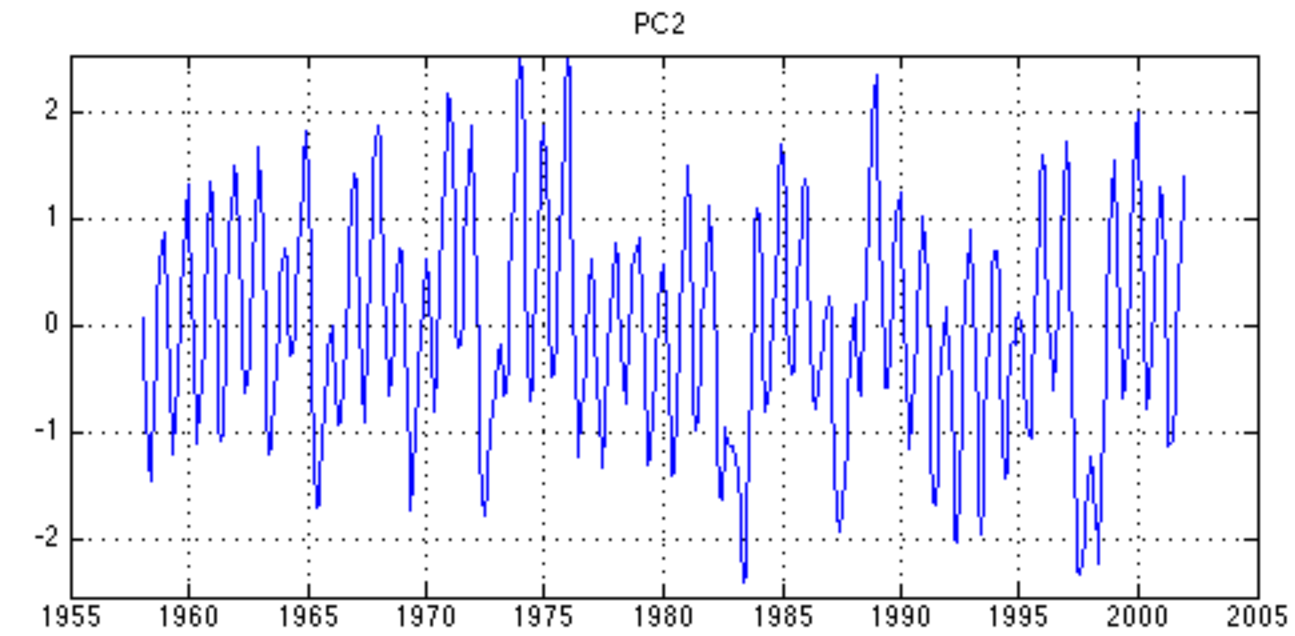
for all j's $X_j^k(t) = A_k(n)U_{jk}$ for k = 1

EOF2 “explains”9% of the variance

First eigen vector U_2 or “EOF2” reorganized into a map:

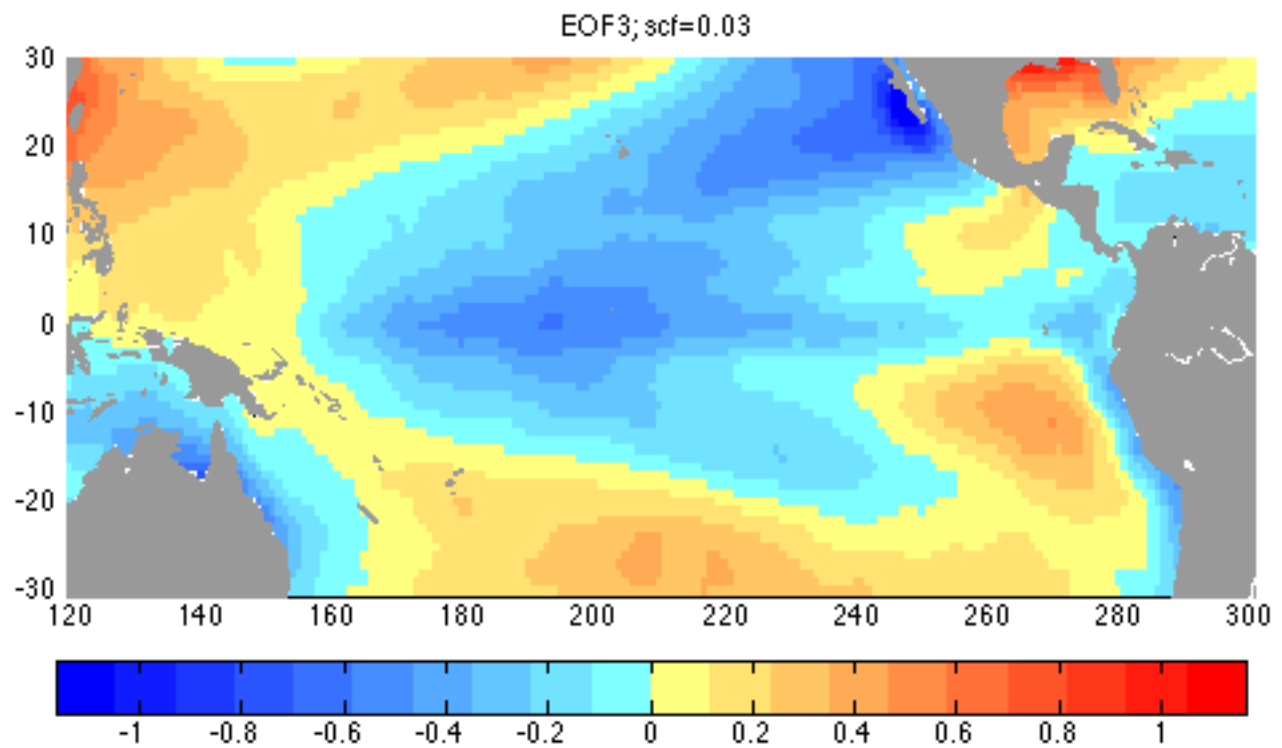


First principal component time series (PC2)

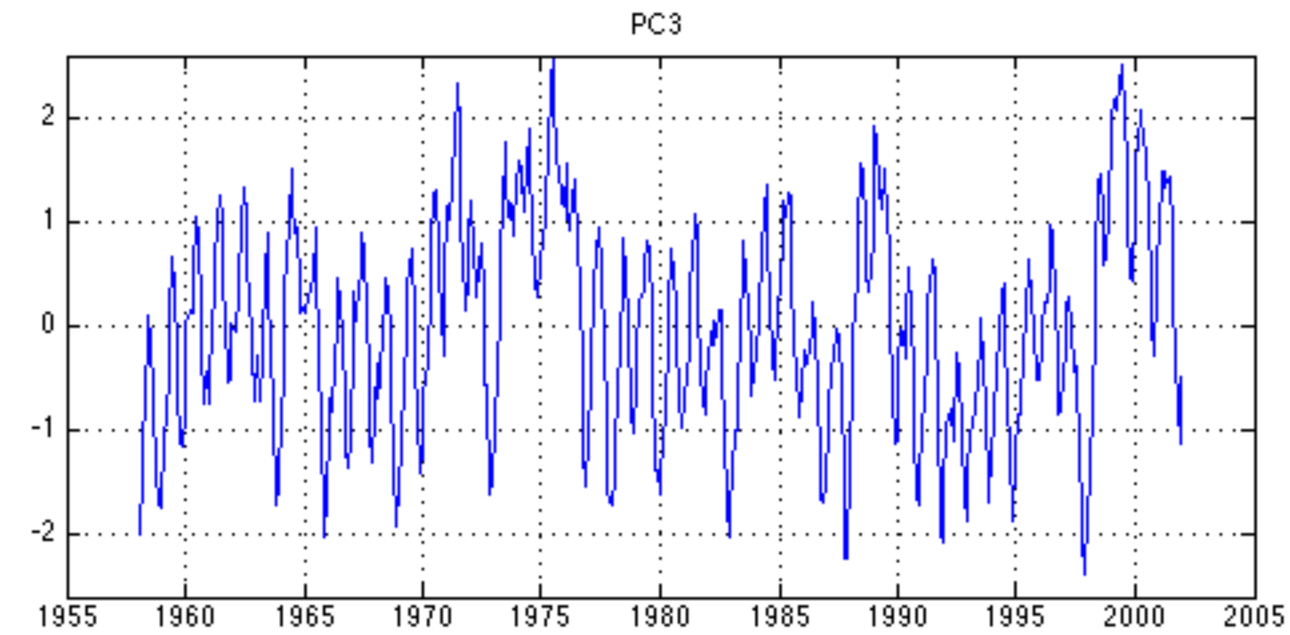


First eigen vector U_3 or "EOF3" reorganized into a map:

EOF3 "explains"3% of the variance



First principal component time series (PC3)

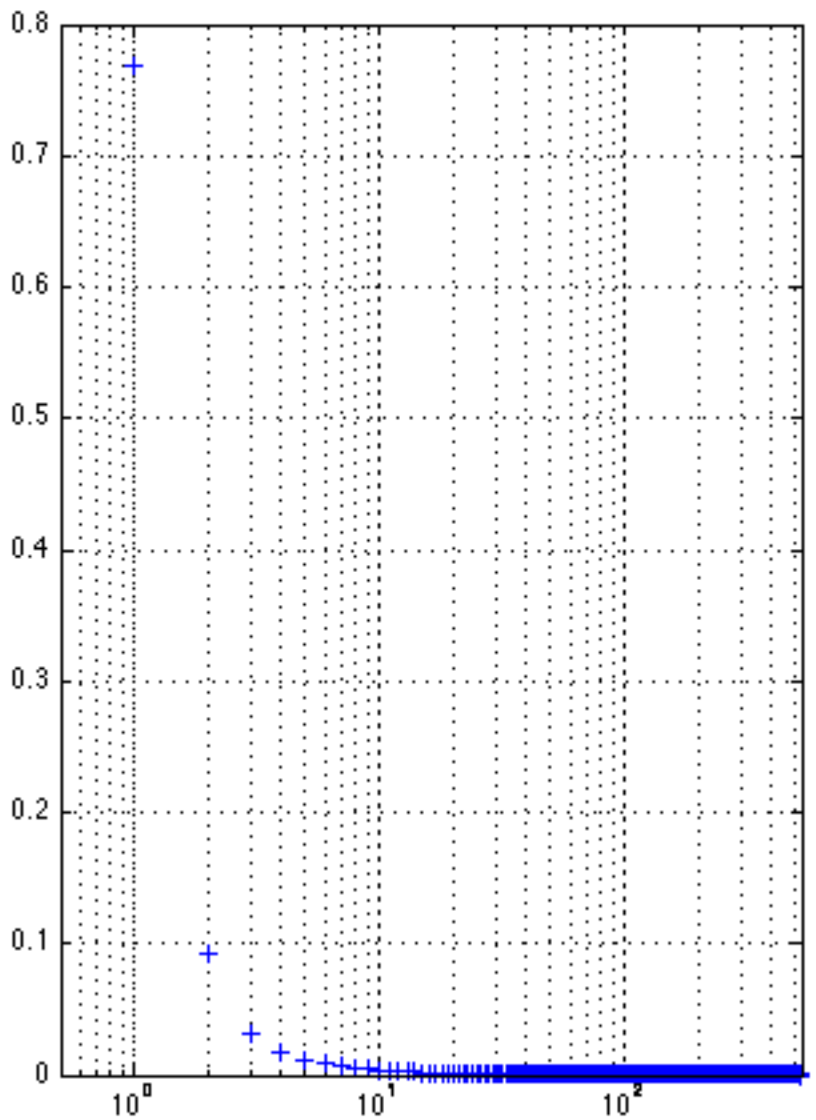


How many EOFs do we need?

$$\mathrm{scf}_m = \frac{\gamma_m}{\sum_{k=1}^{P} \gamma_k}$$

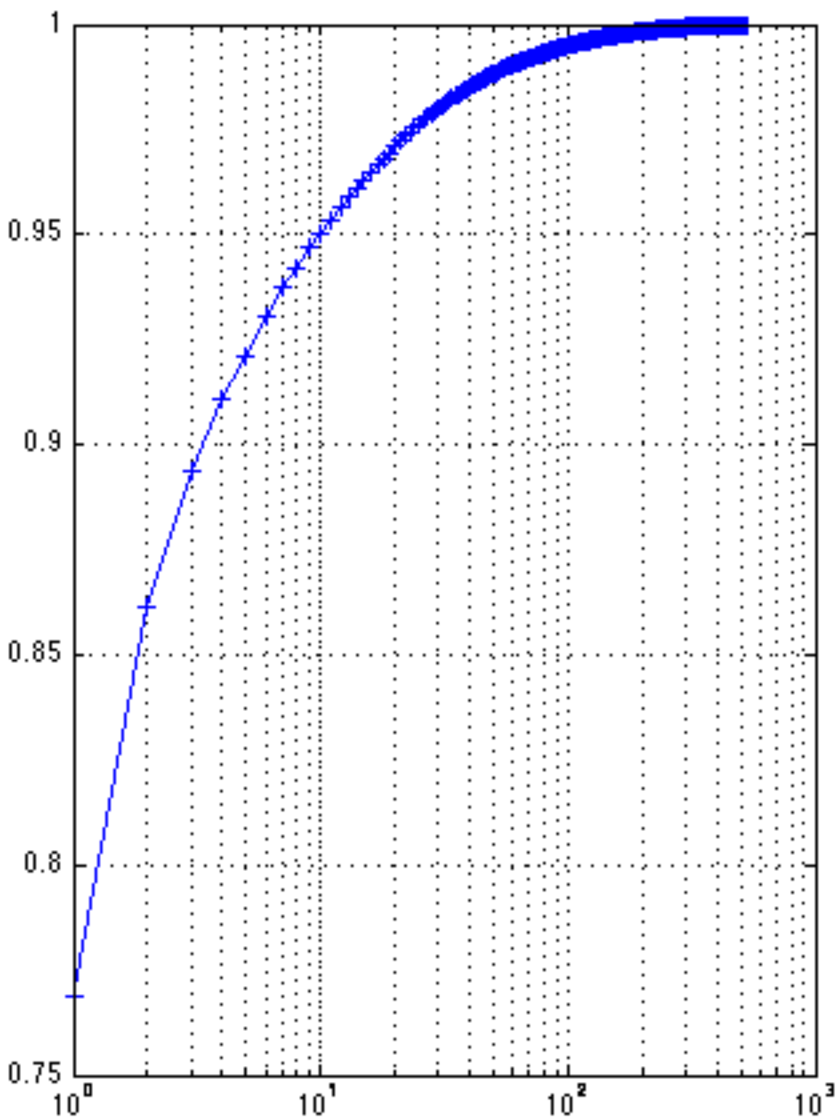
"fraction" of variance explained (<1)

γ_m



Mode number

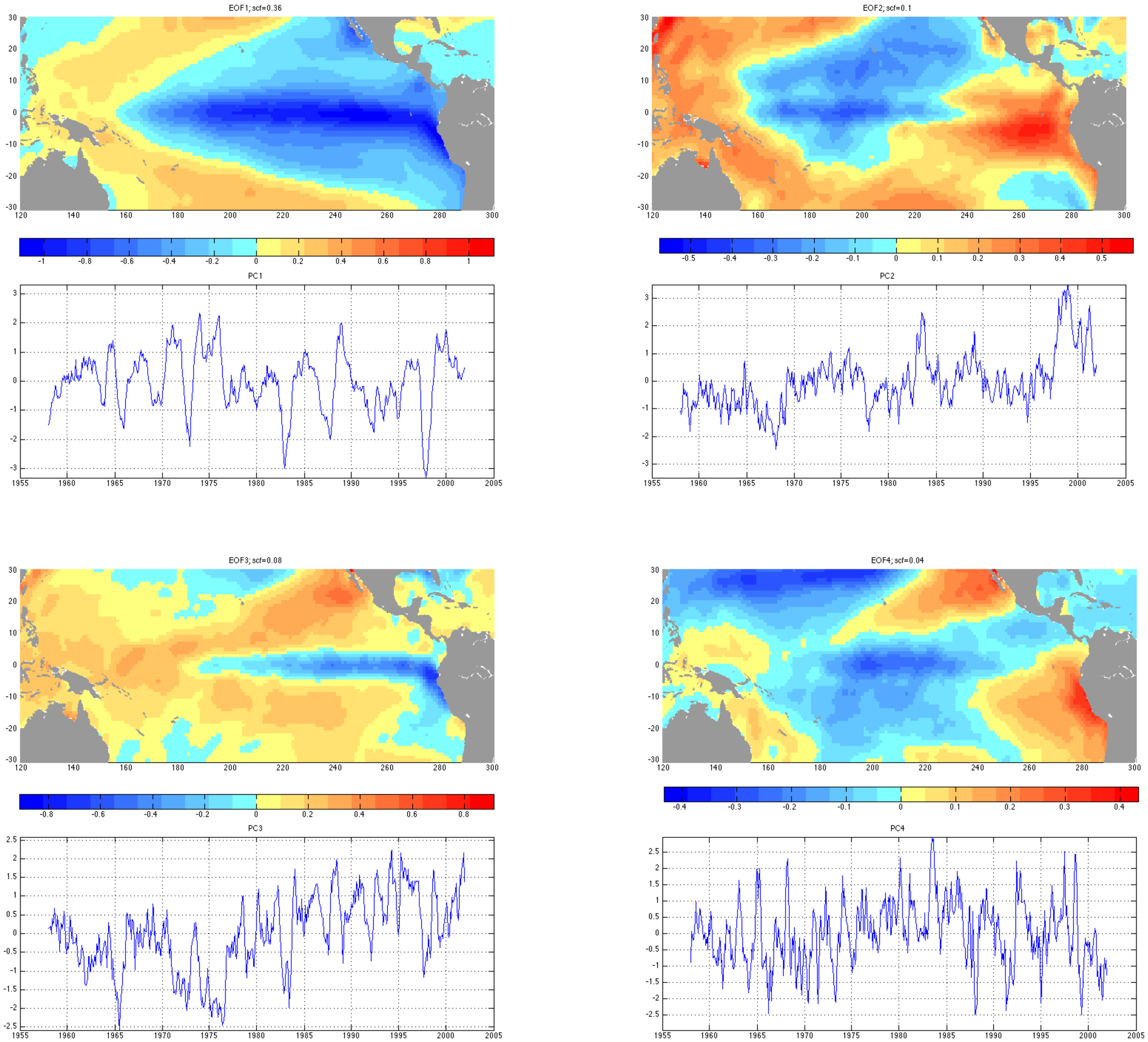
cumulative sum of scf_m



Mode number

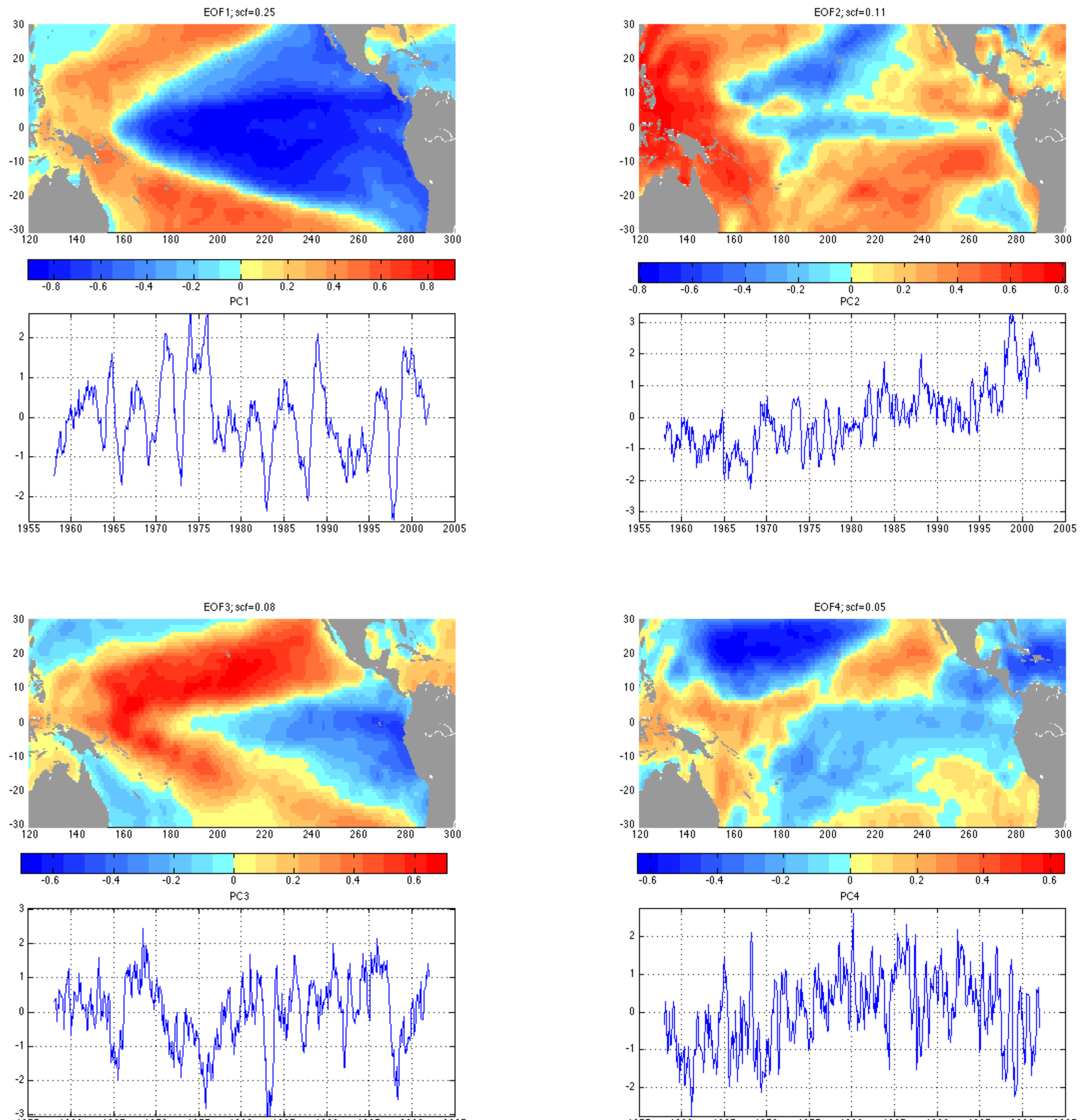
There is not only one way of computing EOFs …

EOF analysis of the same dataset after subtracting monthly averages (seasonal cycle)



There is not only one way of computing EOFs …

EOF analysis of the same dataset after subtracting monthly averages (seasonal cycle) AND normalizing, which means we decomposed correlation matrix instead of covariance matrix



$$\tilde{X}_k(n) = \frac{X_k(n)}{\sqrt{S_{X_k X_k}}}$$

How to do it in Matlab?

Do not use unknown "blackbox" code. Write your own, it takes a few lines. Really.

The not-so-efficient-way:

```
D;
X = detrend(D,0);
C = X'*X/(N-1);
[U,G] = eig(C);
A = X*U;
```

D is your data matrix ordered in columns

Remove the mean of each column

Calculate the covariance matrix

Calculate the eigen decomposition

Calculate the Principal Components

Slightly better: calculate only the top 6 eigen values and vectors:

```
D;
X = detrend(X,0);
C = X'*X/(N-1);
[U,G] = eigs(C);
A = X*U;
```

Calculate only the 6 largest eigen values and vectors

How to do it in Matlab?

<u>The smart way</u>: actually no need to compute the covariance matrix!

D is your data matrix ordered in columns

Remove the mean of each column

```
D;
X = detrend(D,0);
[P,L,Q] =svd(X);
A = P*L;
G = ctranspose(L)*L;
C = Q*ctranspose(L)*L*ctranspose(Q)/(N-1);
```

Done! No kidding (almost). Q contains the eigen vectors in its column.

That's the PCs in the columns

That's the diagonal matrix with eigen values

That's the covariance matrix, only if you want it.

Proof:

SVD stands for Singular Value Decomposition. It is a mathematical decomposition of a matrix which is possible for any N x M matrix, real or complex :

$$\mathbf{X} = \mathbf{PLQ}^H$$

Proof

SVD stands for Singular Value Decomposition. It is a mathematical decomposition of a matrix which is possible for any N x M matrix, real or complex :

$$\mathbf{X} = \mathbf{PLQ}^H$$

$$\begin{aligned}\mathbf{C}_{XX} &= \mathbf{X}^{\mathrm{H}}\mathbf{X}/(N-1)\\ &= (\mathbf{PLQ}^{\mathrm{H}})^{\mathrm{H}}(\mathbf{PLQ}^{\mathrm{H}})/(N-1)\\ &= \mathbf{QL}^{\mathrm{H}}\mathbf{LQ}^{\mathrm{H}}/(N-1)\end{aligned}$$

$$\mathbf{C}_{XX}\mathbf{U} = \mathbf{U\Gamma} \quad \text{or} \quad \mathbf{C}_{XX} = \mathbf{U\Gamma U}^{\mathrm{H}}$$

By identification:

$$\begin{aligned}\mathbf{U} &\equiv \mathbf{Q}\\ \mathbf{\Gamma} &\equiv \mathbf{L}^{\mathrm{H}}\mathbf{L}/(N-1)\end{aligned}$$

Principal components are:

$$\mathbf{A} = \mathbf{XU} = (\mathbf{PLQ}^{\mathrm{H}})\mathbf{Q} = \mathbf{PL}$$

How to do it in Matlab?

The smartest way: In our example, N = 528 and P = 9560. With svd.m, Matlab will calculate 9560 eigen vectors ...

```
D;
```
D is your data matrix ordered in columns with N<P

```
X = detrend(D,0);
```
Remove the mean of each column

```
[Q,L,P] =svd(ctranspose(X),0);
```
Calculate the “economy” SVD, i.e. N eigen vectors instead of P eigen vectors

Note! I exchanged the names!

Note! I calculate the SVD of the transpose

```
A = P*L;
```
That’s the PCs in the columns

```
G = ctranspose(L)*L;
```
That’s the diagonal matrix with eigen values

```
C = Q*ctranspose(L)*L*ctranspose(Q)/(N-1);
```
That’s the covariance matrix, only if you want it.

Proof: for you to do as a homework!

What about "SVD analysis"?

"reft" or "first" data matrix (e.g. SST)

$$\mathbf{X} = [\mathbf{X}_1\ \mathbf{X}_2\ \cdots \mathbf{X}_\mathbf{P}] = \begin{array}{c} \xrightarrow{\text{P locations}} \\ \begin{bmatrix} X_1(1) & X_2(1) & \cdots & X_P(1) \\ X_1(2) & X_2(2) & \cdots & X_P(2) \\ \vdots & \vdots & \cdots & \vdots \\ X_1(N) & X_2(N) & \cdots & X_P(N) \end{bmatrix} \end{array} \Big\downarrow \text{N times}$$

"right" or "second" data matrix (e.g.zonal wind stress)

$$\mathbf{Y} = [\mathbf{Y}_1\ \mathbf{Y}_2\ \cdots \mathbf{Y}_\mathbf{M}] = \begin{array}{c} \xrightarrow{\text{M locations}} \\ \begin{bmatrix} Y_1(1) & Y_2(1) & \cdots & Y_M(1) \\ Y_1(2) & Y_2(2) & \cdots & Y_M(2) \\ \vdots & \vdots & \cdots & \vdots \\ Y_1(N) & Y_2(N) & \cdots & Y_M(N) \end{bmatrix} \end{array} \Big\downarrow \text{N times}$$

Build P ×M cross-covariance matrix:

$$\mathbf{C}_{XY} = \frac{1}{N-1}\mathbf{X}^T\mathbf{Y} \qquad c_{kn} = \frac{1}{N-1}\sum_{k=1}^{N} X_k Y_n$$

SVD analysis (maximum covariance analysis)

$$\mathbf{C}_{XY} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^H$$

SVD decomposition of cross covariance matrix; assume M<P

$$\mathbf{U} = \begin{array}{c|c} & \xrightarrow{\text{M vectors}} \\ \downarrow \text{P locations} & \begin{bmatrix} U_{11} & U_{12} & \cdots & U_{1M} \\ U_{21} & U_{22} & \cdots & U_{2M} \\ \vdots & \vdots & \cdots & \vdots \\ \vdots & \vdots & \cdots & \vdots \\ U_{P1} & U_{M2} & \cdots & U_{PM} \end{bmatrix} \end{array}$$

$$\mathbf{V} = \begin{array}{c|c} & \xrightarrow{\text{M vectors}} \\ \downarrow \text{M locations} & \begin{bmatrix} V_{11} & V_{12} & \cdots & V_{1M} \\ V_{21} & V_{22} & \cdots & V_{2M} \\ \vdots & \vdots & \cdots & \vdots \\ V_{M1} & V_{M2} & \cdots & V_{MM} \end{bmatrix} \end{array}$$

M "left singular vectors"

M "right singular vectors"

$$\mathbf{\Lambda} = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_M \end{bmatrix}$$

M "singular values"

SVD analysis (maximum covariance analysis)

$\mathbf{A} = \mathbf{X}\mathbf{U}$ **A contains the principal components for the left field**

$\mathbf{B} = \mathbf{Y}\mathbf{V}$ **B contains the principal components for the right field**

Decompose your data matrices into P coupled modes:

$$\mathbf{X} = \mathbf{A}\mathbf{U}^H = \sum_{k=1}^{P} \mathbf{A}_k \mathbf{U}_k^H$$

$$\mathbf{Y} = \mathbf{B}\mathbf{V}^H = \sum_{k=1}^{P} \mathbf{B}_k \mathbf{V}_k^H$$

$$r_k = \frac{\mathbf{A}_k^H \mathbf{B}_k}{\left(\mathbf{A}_k^H \mathbf{A}_k \mathbf{B}_k^H \mathbf{B}_k\right)^{1/2}}$$

"coupling coefficient" : correlation between PC of left field and PC of right field

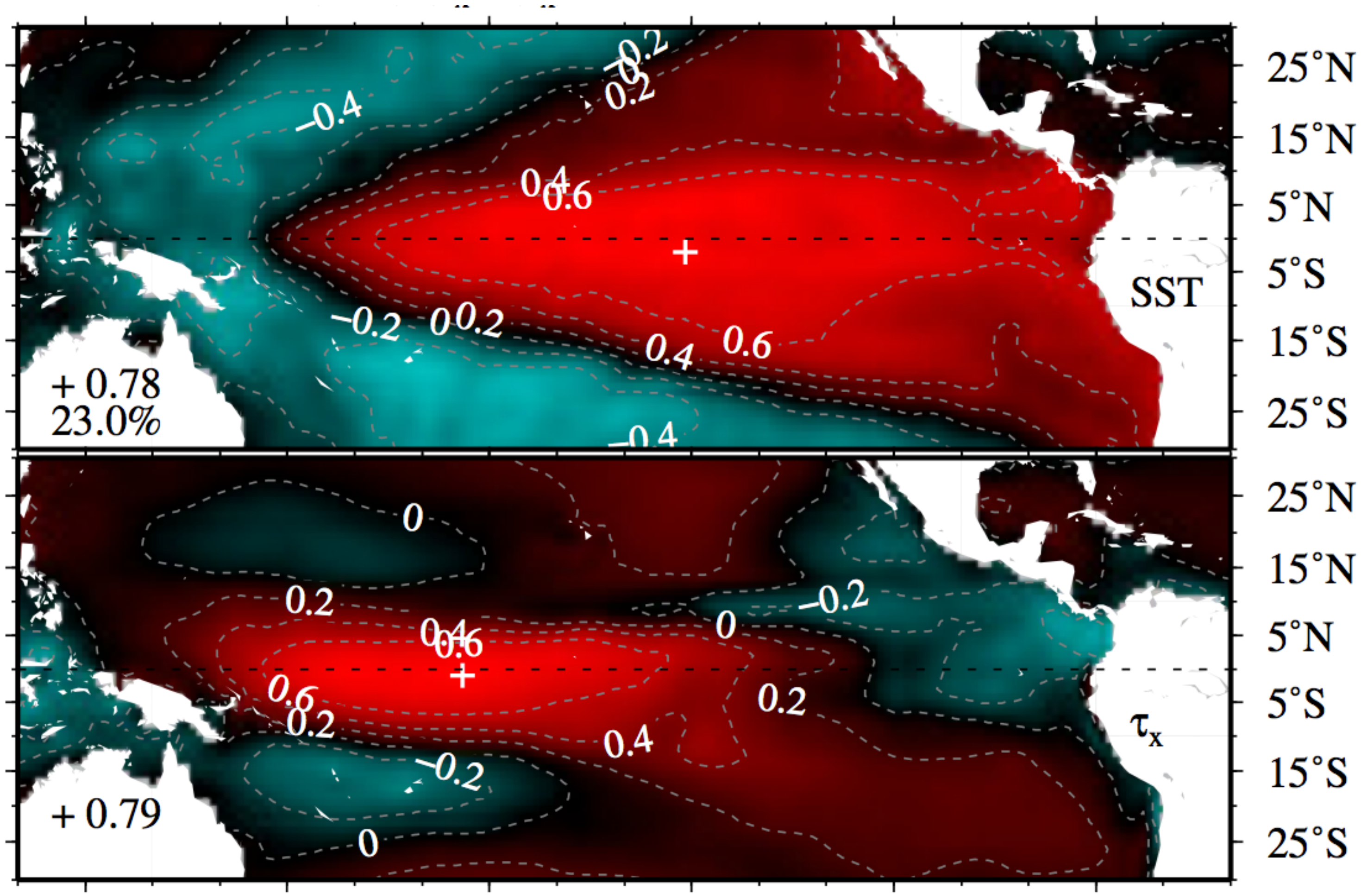
SVD analysis (maximum covariance analysis)

Example: SST (left field) and zonal wind stress (right field)

singular vectors

$\mathbf{U}_1$

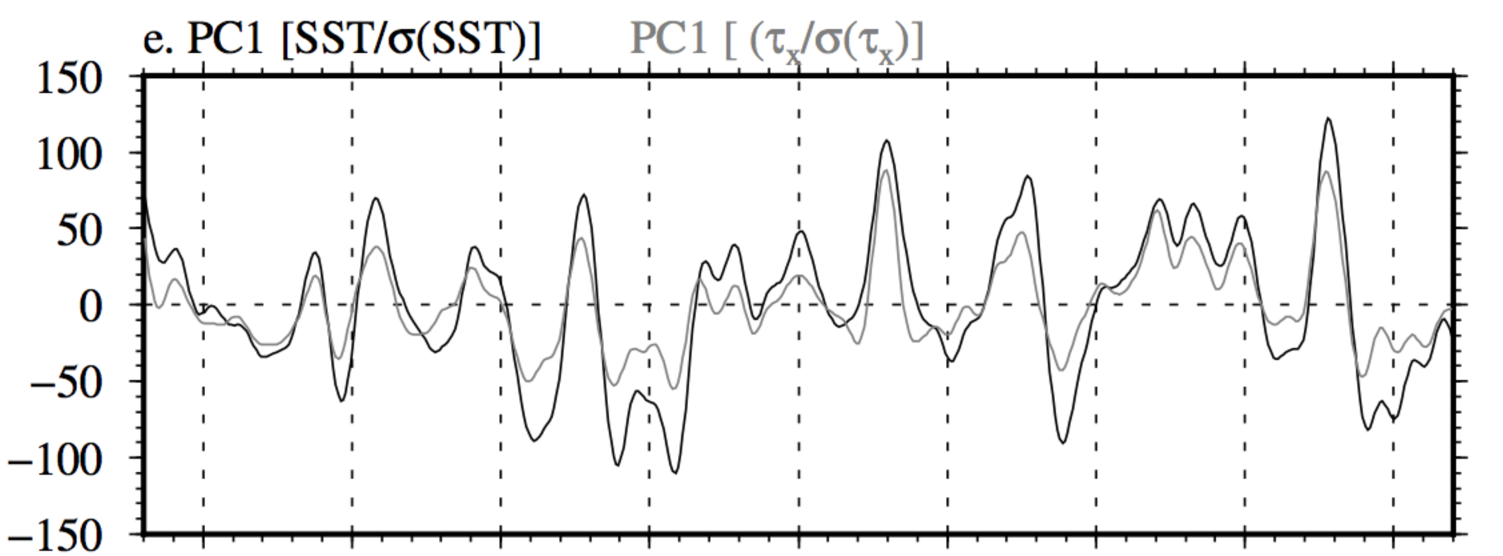
$\mathbf{V}_1$



PC time series

$\mathbf{A}_1$

$\mathbf{B}_1$



$r_k = 0.85$

What I have not talked about:

Significance!

- Eigen decomposition and singular value decomposition are mathematical operations on matrices: you will ALWAYS get something

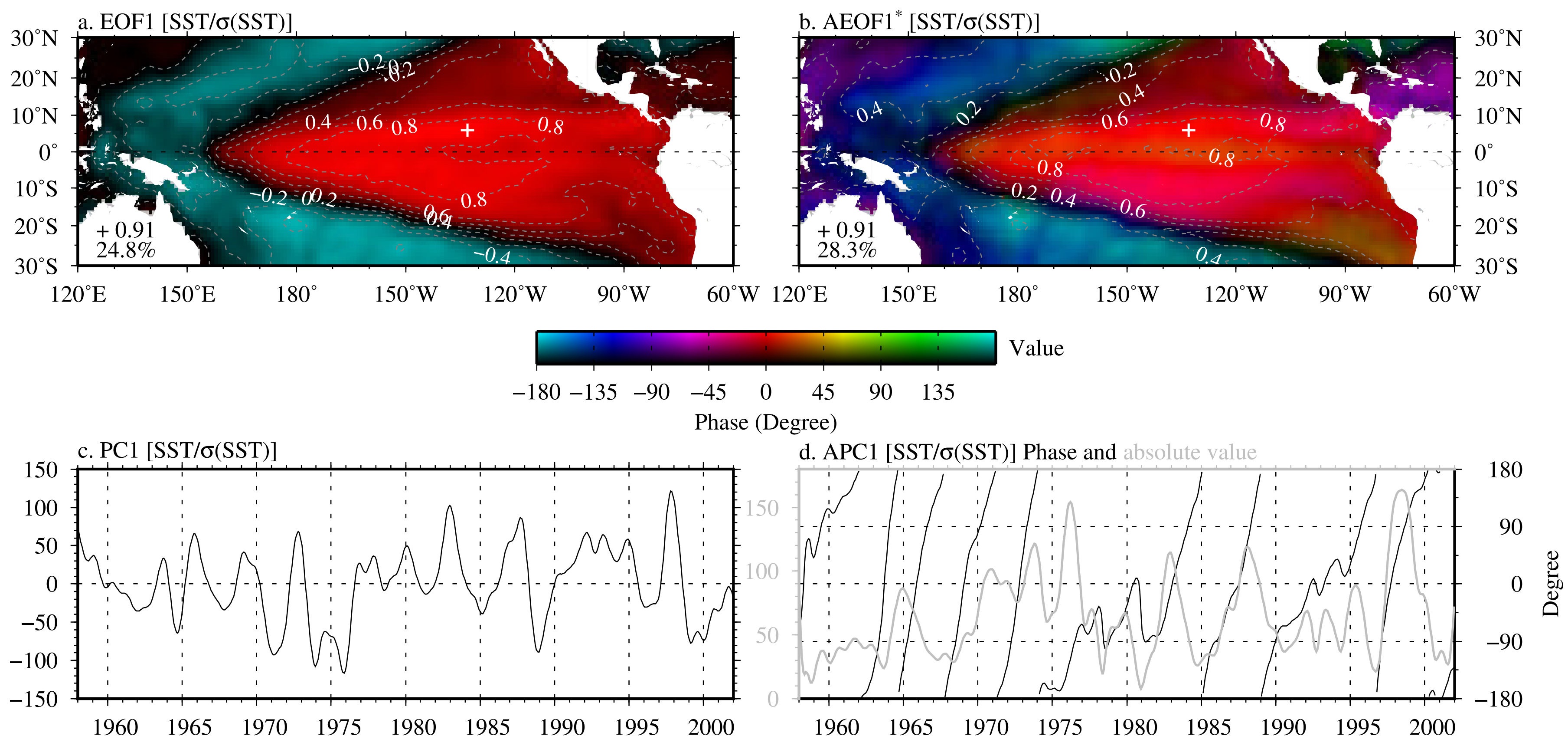
- You need to evaluate or test for the statistical significance of your results (how? good question)

- Finally, and ideally, you need to be able to provide a physically-based interpretation of the EOF or SVD modes (example: orthogonality of modes is not physically motivated)

What I have not talked about:

Complex EOF (and complex SVD)

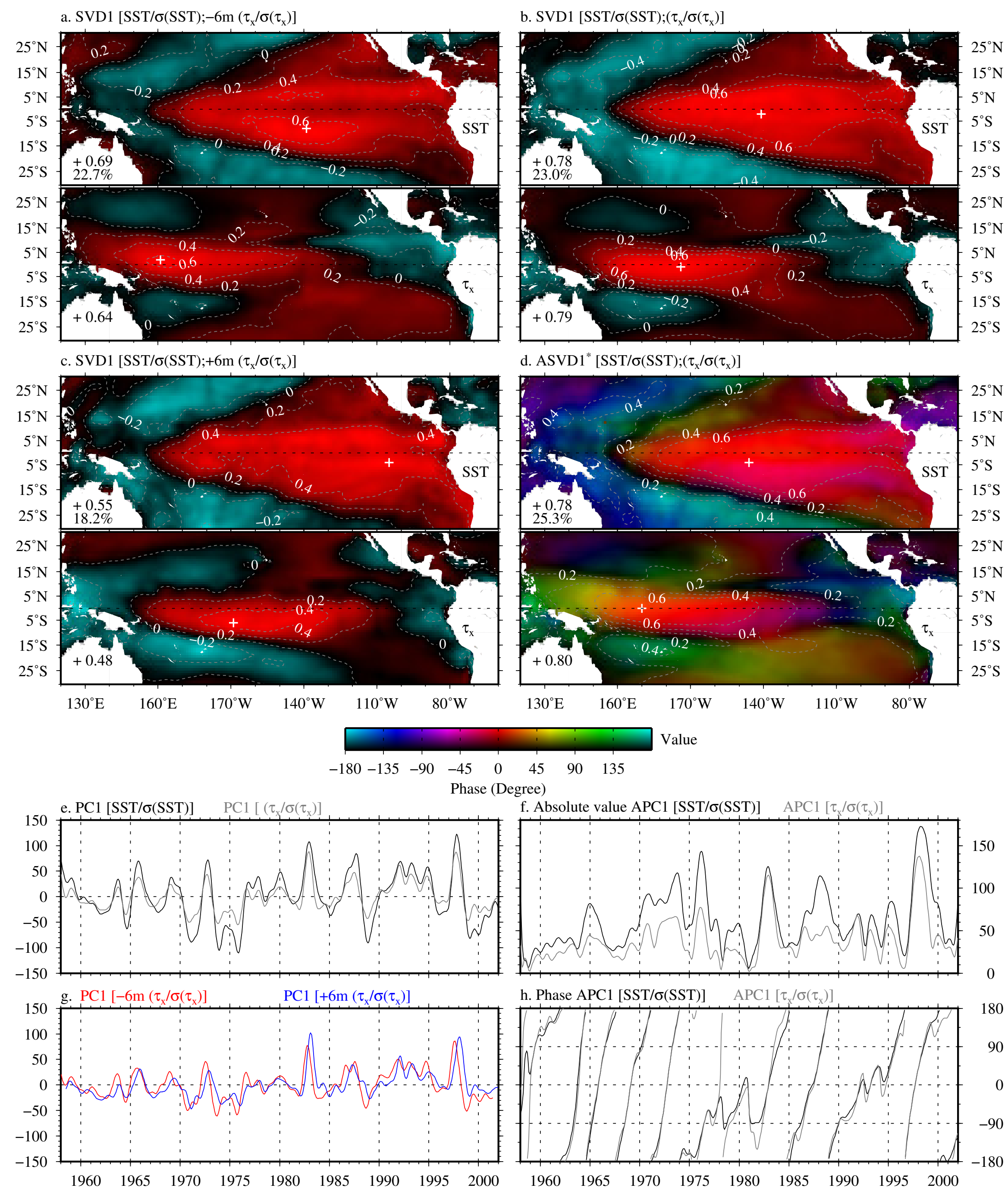
example: standard EOF vs “complex” EOF



What I have not talked about:

Lagged EOF (or lagged SVD)

example: lagged SVD vs “complex” SVD



Time for movies?

A few references, again!

Climate indices are often defined from EOFs:

http://www.esrl.noaa.gov/psd/data/climateindices/list/

Books:

Preisendorfer and Mobley (1988): *Principal Component Analysis in Meteorology and Oceanography*, Elsevier, New York. (RSMAS Library)

von Storch and Zwiers (2002): *Statistical Analysis in Climate Research*, Cambridge University Press, Cambridge (UM Library Online)

Emery and Thomson (2001): *Data analysis methods in physical oceanography*. Elsevier, second and revised edition (RSMAS Library)

Menke (2012), *Environmental data analysis with MatLab, (UM Library Online)*

Papers:

Bretherton C, Smith C, Wallace J, *et al.* (1992). *An intercomparison of methods for finding coupled patterns in climate data*. J. Clim. 5(6): 541–560.

Venegas S, Mysak L, Straub D. (1996). *Evidence for interannual and interdecadal climate variability in the South Atlantic*. Geophys. Res. Lett. 23(19): 2673–2676.

Barnett T. 1983. *Interaction of the monsoon and pacific trade wind system at interannual time scales. I- the equatorial zone*. Monthly Weather Review 111: 756–773.

Merrifield M, Guza R. 1990. *Detecting propagating signals with complex empirical orthogonal functions: A cautionary note*. J. Phys. Oceanogr. 20(10): 1628–1633.