Multivariate Data Analysis (PCA)

Neeti

neeti@terisas.ac.in

Department of Natural Resources

TERI School of Advanced Studies

TERI-NORCE research school on Towards data science in climate research: perspectives on Climate Extremes
Oct 16, 2019

Multivariate Data Analysis

Data Analysis

- **What is data analysis?**

To extract relevant information contained in the data which can then be used to solve a given problem.

How do we extract information from data?

- Using some statistical technique
- Number of schemes exist for classifying techniques
- Most of them are based on
  - Measurement Scale
  - Variables (types)

What is Measurement?

Measurement is a process by which numbers or symbols are attached to given characteristics or properties of stimuli according to predetermined rules or procedures

Examples: any region can be described with respect to a number of characteristics: temperature, rainfall, agroclimatic zone, humidity etc.

Characteristics of ocean: Ocean colour, temperature etc…

Measurement Scale

Steven (1946) postulated four types of measurement scale:

- Nominal
- Ordinal
- Interval
- Ratio

Variables

- A variable is a property or characteristics of a thing or people that varies in quality and quantity

- Variables can be classified as:
  - Metric: measured using interval and ratio scale
  - Non-metric: measured using ordinal and nominal scale

Classification of Data Analytic methods

- Dependence Methods
  - One (or more) variables are dependent variables, to be explained or predicted by others (independent variables)

- Interdependence methods
  - No variables thought of as “dependent”

Dependence Methods

- Classification of dependence methods based on:
  - Number of variables (one or more) for independent and dependent variables
  - Measurement scale (metric/non-metric) for independent and dependent variables

Dependence Methods

Independent Variable(s)		Dependent variables			
		One		More than one	
		Metric	Non metric	Metric	Nonmetric
One	Metric	Regression	Discriminant Analysis, logistic regression	Canonical Correlation	Multiple group discriminant analysis
	Non Metric	T-test	Discrete discriminant analysis	MANOVA	Discrete MDA
More than one	Metric	Multiple regression	Discriminant Analysis; Logistic regression	Canonical Correlation	MDA
	Non metric	ANOVA	Discrete Discriminant analysis Conjoint Analysis	MANOVA	Discrete MDA

Sharma: Applied multivariate techniques

Interdependence Methods

	Type of Data	
Number of variables	Metric	Non metric
Two	Simple Correlation	Two-way contingency table Loglinear models
More than two	$\boxed{\text{Principal Components}}$ Factor Analysis, Cluster Analysis	Multiway Contigency tables Loglinear models Correspondence Analysis

Sharma: Applied multivariate techniques

Fundamentals on Data for PCA

- Mean and mean-centered data
- Degree of Freedom
- Variance, Sum of squares, and cross products
- Standardization

Mean and Mean centering

- Mean: measure of central tendency
- Another way to represent data is by centering with respect to mean

Degree of Freedom

- Number of independent pieces of information contained in a dataset that are used for computing a given summary measure.

- Degree of freedom for mean centered data is n-1 where n is the number of observations

Variance

- Variance: amount of dispersion in the dataset

- Variance is average of square of difference from the mean

$$s^2 = \frac{\sum_{i=1}^{N} (x_i - \bar{x})}{n - 1} = SS / df$$

- Where s^2 stands for the sample variance
- $\bar{x}$ is the sample mean
- n is the total number of values in the sample
- x_i is the value of the i-th observation.
- $\sum$ represents a summation

SS = Sum of Squares
= Sum of squared deviation from mean
df = Degree of freedom

Measures of Association

- Scatter diagram plot provides a graphical description of positive/negative, linear/non-linear relationship
- Some numerical description of the positive/negative, linear/non-linear relationship are obtained by:
  - Covariance
  - Coefficient of correlation

Covariance

- A measure of covariation between variables
- Variance is average of square of difference from the mean

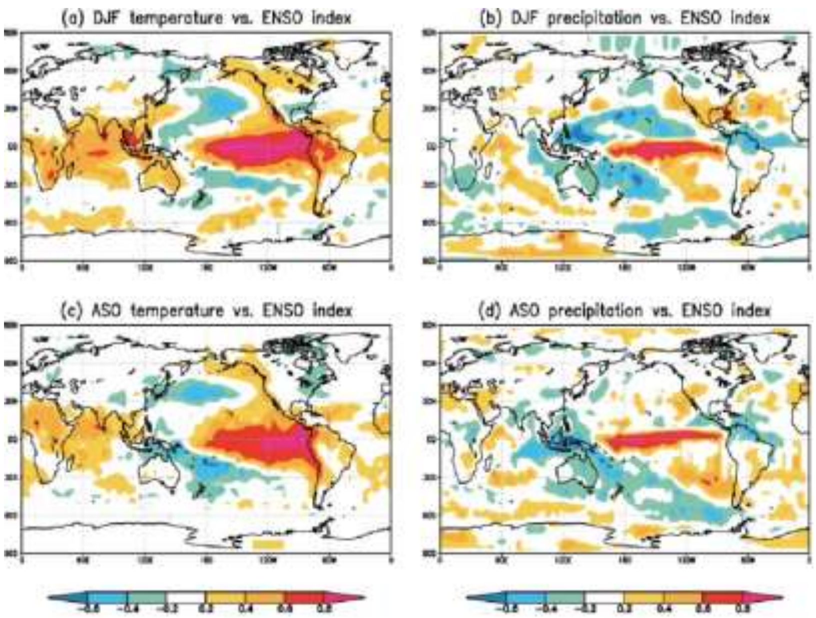
$$s^2 = \frac{\sum_{i=1}^{N} (x_i - \bar{x})^2}{n - 1} = SS / df$$

- Where s^2 stands for the sample variance
- $\bar{x}$ is the sample mean
- n is the total number of values in the sample
- x_i is the value of the i-th observation.
- $\sum$ represents a summation

SS = Sum of Squares
df = Degree of freedom

Measures of Association: Example

- A sample of monthly rainfall data and ENSO index are collected and shown below:



- How is the relationship between Rainfall and MEI index? Is the relationship linear/non-linear, positive/negative, etc.

Covariance

- Measure of the covariation between variables
- Mean of products of deviations from the variable mean:

$$\operatorname{cov}(X,Y) = \frac{\sum_{i=1}^{n-1} \left(x_i - \bar{x}\right)\left(y_i - \bar{y}\right)}{n-1} = SCP / df$$

- Where *cov(X,Y)* is the covariance
- $\bar{x}, \bar{y}$ are the means of *X* and *Y* respectively
- *n* is the total number of values in the population
- x_i, y_i are the values of the *i*-th observations of *X* and *Y* respectively.
- $\sum$ represents a summation
- SCP Sum of cross product
- df Degree of freedom

Sum of Squared and cross product matrix(SSCP)

SS and SCP are summarized in a matrix called sum of squared and cross products matrix SSCP

SSCP matrix

	Variable 1	**Variable 2**
Variable 1	SS_1	SCP_{12}
Variable 2	SCP_{21}	SS_2

2 variables: 2 variance, 1 covariance
p variables: p variance, p(p-1)/2 covariance

Sum of Squared and cross product matrix(SSCP)

SS and SCP are summarized in a matrix called sum of squared and cross products matrix SSCP

SSCP matrix

	Variable 1	**Variable 2**
Variable 1	SS_1	SCP_{12}
Variable 2	SCP_{21}	SS_2

2 variables: 2 variance, 1 covariance
p variables: p variance, p(p-1)/2 covariance

Variance Covariance Matrix (S)

$$S = SSCP / df$$

	Variable 1	**Variable 2**
Variable 1	$(SS_1)/df$	$(SCP_{12})\ /df$
Variable 2	$(SCP_{21})\ /df$	$(SS_2)/df$

Covariance

- If two variables increase/decrease together→ large positive covariance →Positive Relationship
- If with increase in one variable, the other decreases and vice versa → Large negative covariance → Negative relationship
- If two variables are unrelated, the covariance may be a small number.
- How large is large? How small is small?

Covariance

- How large is large?
- How small is small?
- A drawback of covariance is that it is usually difficult to provide any guideline how large covariance shows a strong relationship and how small covariance shows no relationship.
- Coefficient of correlation can overcome this drawback to a certain extent.

Coefficient of Correlation

- The coefficient of correlation is the covariance divided by the standard deviations of *X* and *Y*:

$$\rho = \frac{COV(X,Y)}{\sigma_x \sigma_y}$$

- Where *r* is the sample coefficient of correlation
- *cov(X,Y)* is the covariance
- s_x, s_y are the means of *X* and *Y* respectively

Standardization

$$\rho = \frac{COV(X,Y)}{\sigma_x \sigma_y}$$

Division by Standard Deviation

Standardized data are obtained by **dividing mean corrected data** by respective **Standard deviation**

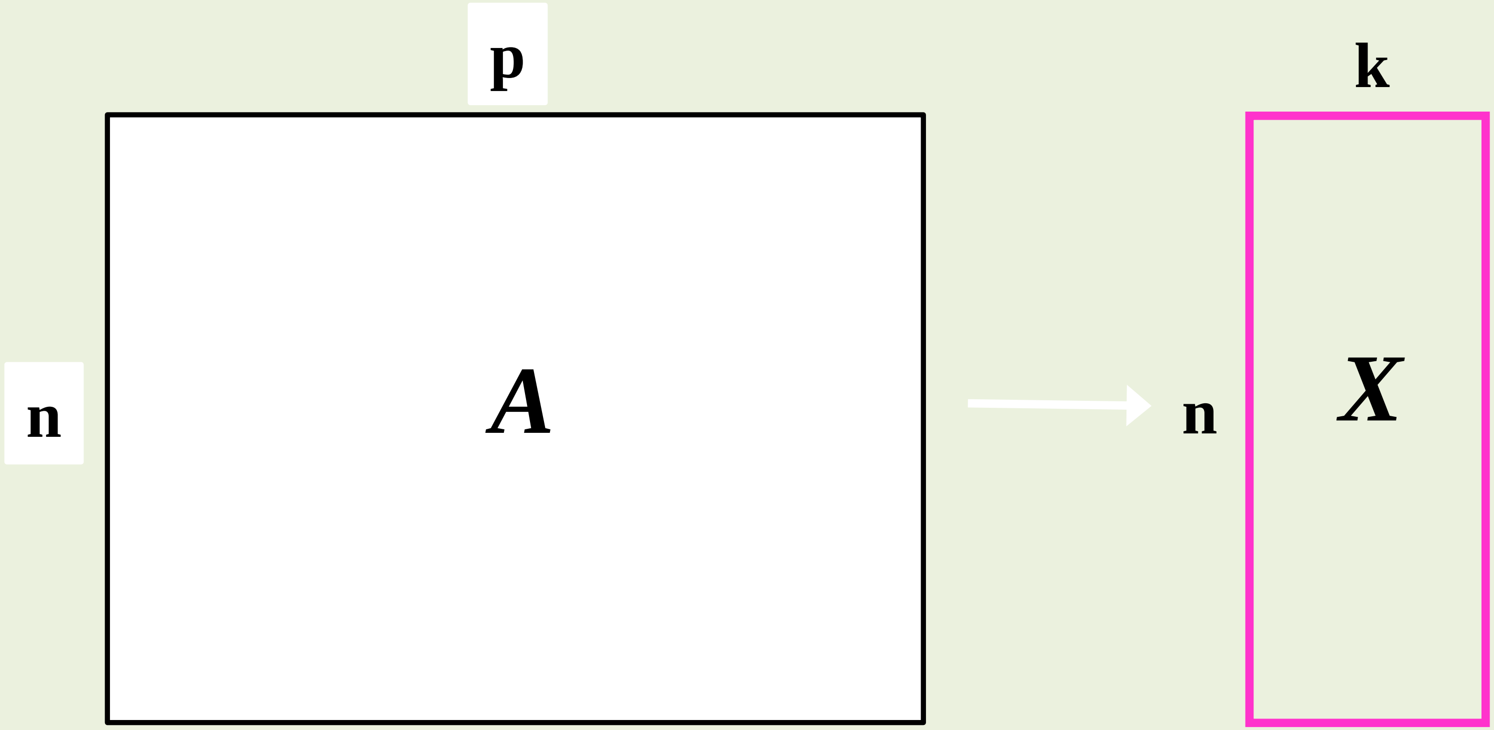
Covariance of two **standardized variables** is called **correlation coefficient** or Pearson Product moment correlation.

Principal Component Analysis (PCA)

- What is PCA?
- When do we use PCA?
- What is the geometric interpretation of PCA?
- What is the mathematical structure of PCA?
- How do we interpret the results from PCA?

Data Reduction

Summarization of data with many (p) variables by a smaller set of (k) derived variables.



PCA

- Invented by Pearson (1901) and Hotelling (1933)

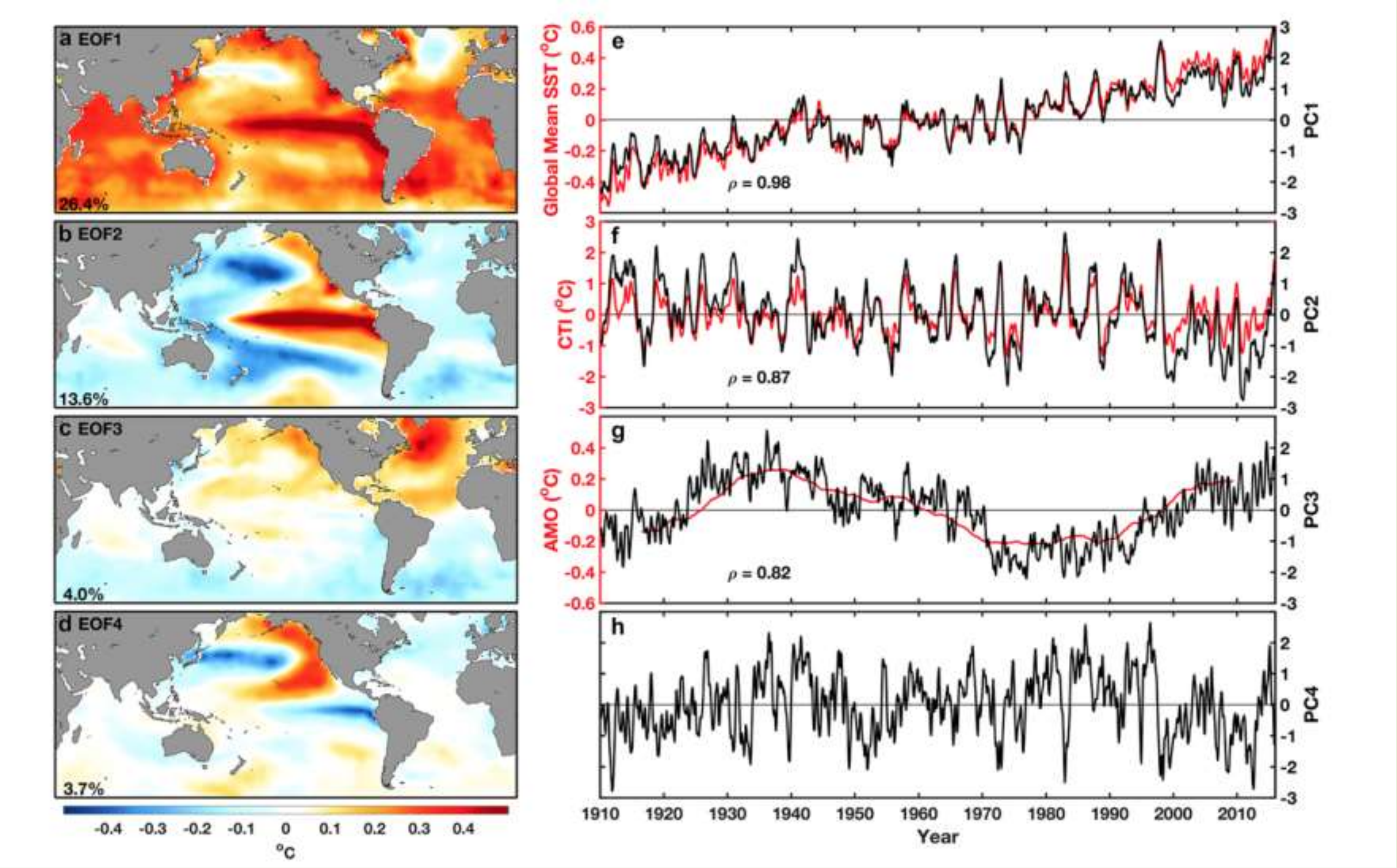
- Summarizes a data matrix of *n* objects by *p* variables, which may be correlated through uncorrelated axes (principal components) that are linear combinations of the original *p* variables

- The maximum number of new variables (uncorrelated) that can be formed is equal to number of original variables

When can we use PCA?

- For dimension reduction without much loss of information: To remove redundancy
- To extract important features/dominant patterns from large dataset
- To identify similarity and dissimilarity among variables

PCA on SST Anomaly (1910-2015)



Source: Chen and Tung 2018, Global-mean surface temperature variability: space–time perspective from rotated EOFs

SCIENTIFIC REPORTS

OPEN

Increased influence of ENSO on Antarctic temperature since the Industrial Era

Received: 19 July 2018
Accepted: 2 April 2019
Published online: 12 April 2019

Waliur Rahama

Under the influe
Southern Oscilla
community. Anta
(SH) teleconnecti
past five centurie
East and West An
frequency bands
increasing trend
CE). This observa
based on tropical
characterized by
Our investigatio
in PSA activity a
temperature are

ENSO is a major
studies have sugg
PDO[1,3], however,
changes due to So
changes in tropica
time particularly
To reconstruc
on model simula
Intercomparison P
made based on in
and PDO. These r
records[11]. In additi
in sample depth,
misrepresents clin
ics/subtropics and
Though ENSO sig
to non-stationary
changes in ENSO
surface air tempe
investigated temp
on Antarctic SAT
post-industrial gr

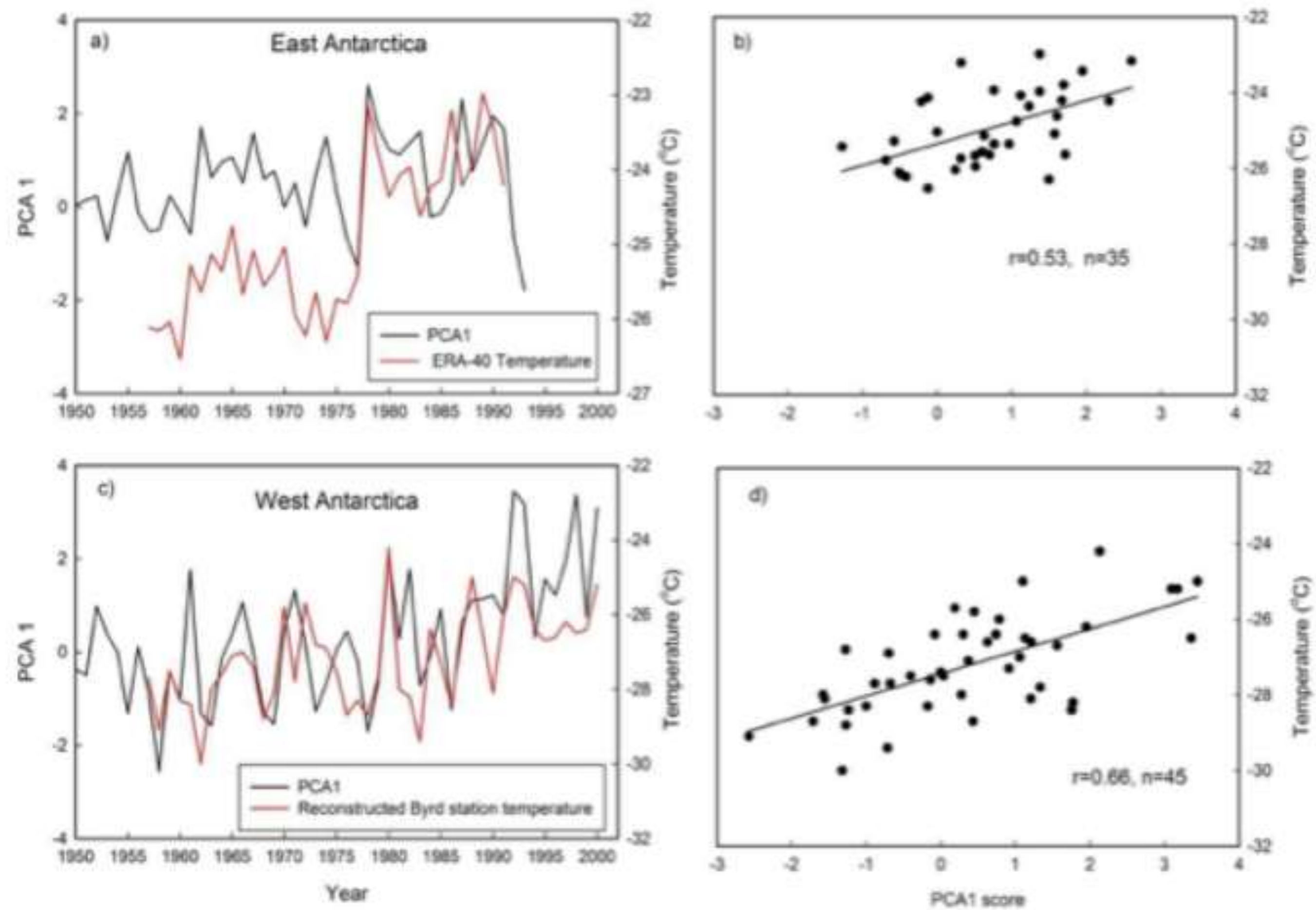


Fig. S2 Extraction of temperature signals from the oxygen isotopes records of multiple ice cores: a, b)

Caribbean probably exceeds the U.S. Environmental Protection Agency's 24-hour standard. Although there is no evidence that exposure to dust across this region presents a health problem, it does demonstrate how climate processes can bring about changes in our environment that could have a wide range of consequences on intercontinental scales.

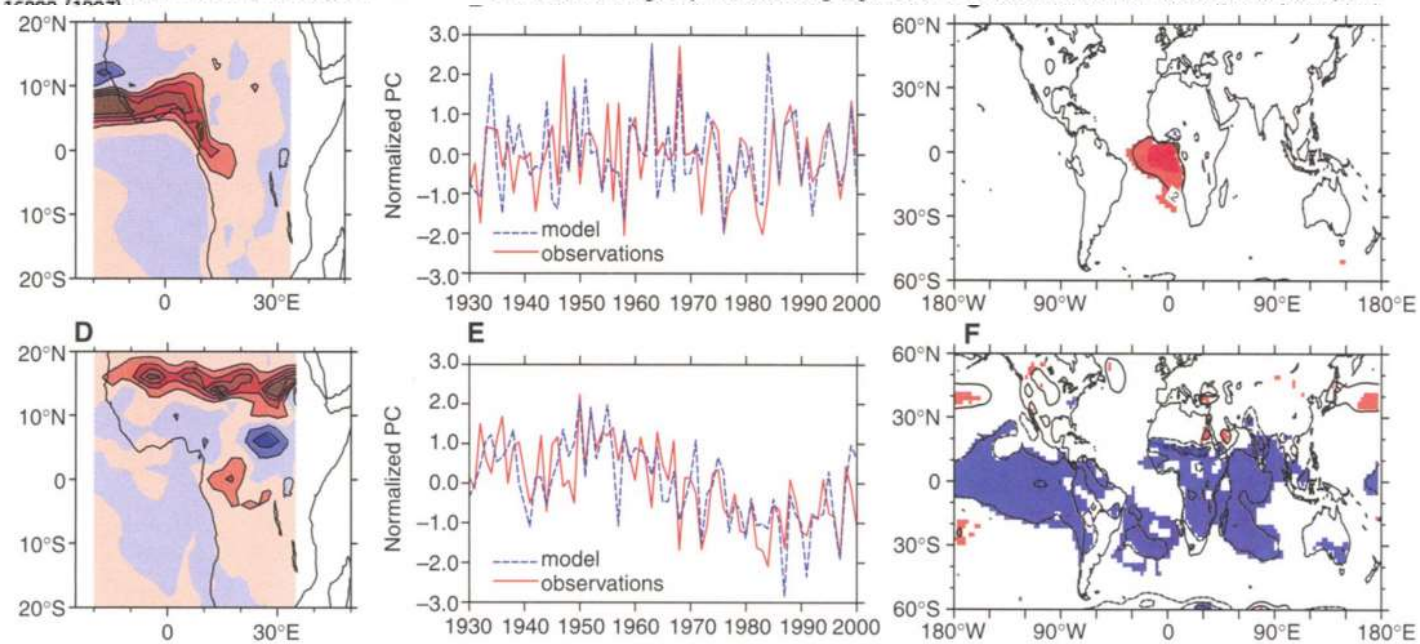
References and Notes

1. Y. K. Kaufman, D. Tanré, O. Boucher, *Nature* **419**, 215 (2002).
2. Satellite images are available at http://visibleearth.nasa.gov/Atmosphere/Aerosols.
3. R. B. Husar, J. M. Prospero, L. L. Stowe, *J. Geophys.*

Oceanic Forcing of Sahel Rainfall on Interannual to Interdecadal Time Scales

A. Giannini,[1*†] R. Saravanan,[1] P. Chang[2]

We present evidence, based on an ensemble of integrations with NSIPP1 (version 1 of the atmospheric general circulation model developed at NASA's Goddard Space Flight Center in the framework of the Seasonal-to-Interannual Prediction Project) forced only by the observed record of sea surface temper-



PCA of northern summer rainfall over tropical Africa during 1930–2000. The two leading patterns of observed precipitation explain 25% and 15% of the total variance, their modeled counterparts 32% and 21% of the ensemble-mean variance, respectively. (**A** and **D**) Leading spatial patterns (EOFs) in the model. Red, positive precipitation anomalies; blue, negative anomalies. (**B** and **E**) Leading PCs; (B) is interannual in nature, whereas (E) captures the well-known trend in Sahel rainfall. The correlation between observed (red, solid line) and modeled (blue, dashed line) Gulf of Guinea PCs is 0.62; that between Sahel PCs is 0.73. (**C** and **F**) Regression maps of the leading model PCs with ensemble-mean surface temperature. Contour interval is every 0.4 K, starting at 0.2 K, and shading represents statistical significance of the anomalies at the 99.9% level or higher. Solid lines and red color, positive anomalies; dashed lines and blue color, negative anomalies.

This content downloaded from 182.64.153.68 on Sun, 13 Oct 2019 23:46:46 UTC

Contents lists available at ScienceDirect

Environment International

journal homepage: www.elsevier.com/locate/envint

Intra-urban biomonitoring: Source apportionment using tree barks to identify air pollution sources

Tiana Carla Lopes Moreira [a,c,*], Regiani Carvalho de Oliveira [a,c], Luís Fernando Lourenço Amato [a,c], Choong-Min Kang [d], Paulo Hilário Nascimento Saldiva [a,c], Mitiko Saiki [b,c]

[a] Medical School of São Paulo University (FMUSP), São Paulo, SP, Brazil
[b] Nuclear and Energy Research Institute (IPEN-CNEN/SP), São Paulo, SP, Brazil
[c] National Institute for Integrated Analysis of Environmental Risk (INAIRA), São Paulo, SP, Brazil
[d] Harvard School of Public Health (HSPH), Boston, MA, USA

ARTICLE INFO

Article history:
Received 1 December 2015
Received in revised form 4 March 2016
Accepted 6 March 2016
Available online 17 March 2016

ABSTRACT

It is of great interest to evaluate if there is a relationship between possible sources and trace elements using biomonitoring techniques. In this study, tree bark samples of 171 trees were collected using a biomonitoring technique in the inner city of São Paulo. The trace elements (Al, Ba, Ca, Cl, Cu, Fe, K, Mg, Mn, Na, P, Rb, S, Sr and Zn) were determined by the energy dispersive X-ray fluorescence (EDXRF) spectrometry. The Principal Component Analysis (PCA) was applied to identify the plausible sources associated with tree bark measurements. The greatest source was vehicle-induced non-tailpipe emissions derived mainly from brakes and tires wear-out and road dust resuspension (characterized with Al, Ba, Cu, Fe, Mn and Zn), which was explained by 27.1% of the variance, followed by cement (14.8%), sea salt (11.6%) and biomass burning (10%), and fossil fuel combustion (9.8%). We also verified that the elements related to vehicular emission showed different concentrations at different sites of the same street, which might be helpful for a new street classification according to the emission source. The spatial distribution maps of element concentrations were obtained to evaluate the different levels of pollution in streets and avenues. Results indicated that biomonitoring techniques using tree bark can be applied to evaluate dispersion of air pollution and provide reliable data for the further epidemiological studies.
© 2016 Elsevier Ltd. All rights reserved.

Earth Sci Inform (2011) 4:117–124
DOI 10.1007/s12145-011-0082-7

RESEARCH ARTICLE

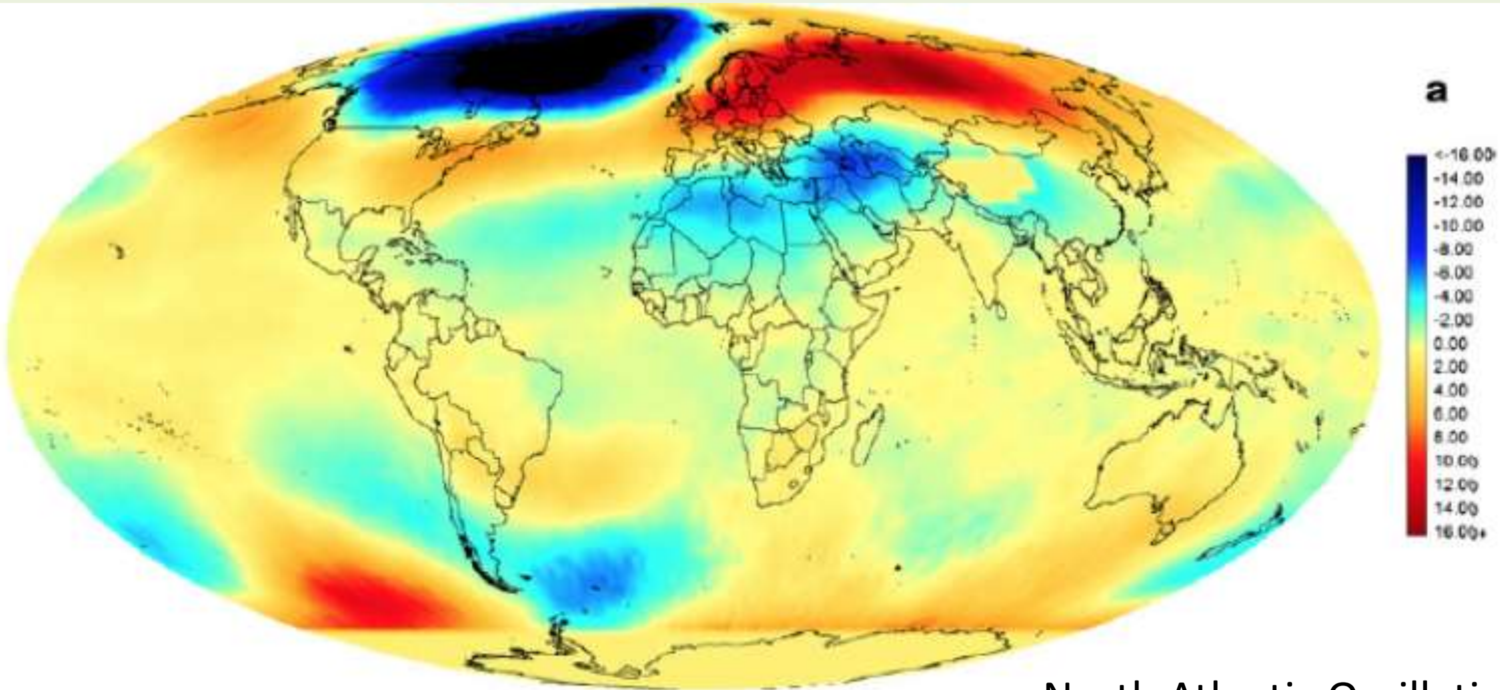
Implications of space-time orientation for Principal Components Analysis of Earth observation image time series

Elia Axinia Machado-Machado · Neeti Neeti · J. Ronald Eastman · Hao Chen

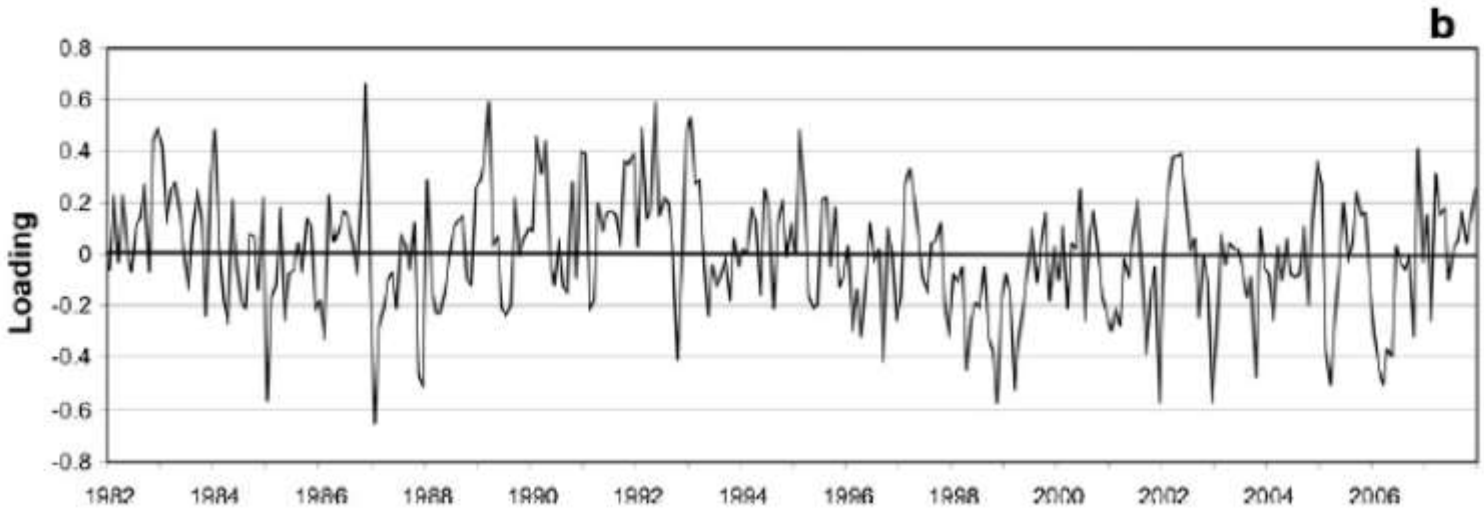
Received: 24 August 2010 / Accepted: 17 May 2011 / Published online: 2 June 2011
© Springer-Verlag 2011

Abstract A time series of geographic images can be viewed from two perspectives: as a set of images, each image representing a slice of time, or as a grid of temporal profiles (one at each pixel location). In the context of Principal Components Analysis (PCA), these different orientations are known as T-mode and S-mode analysis respectively. In the detrends over space. Further, in the formation of components, S-mode PCA preferences patterns that are prevalent over space while T-mode PCA preferences patterns that are prevalent over time. The two orientations thus provide complementary insights into the nature of variability within the series.

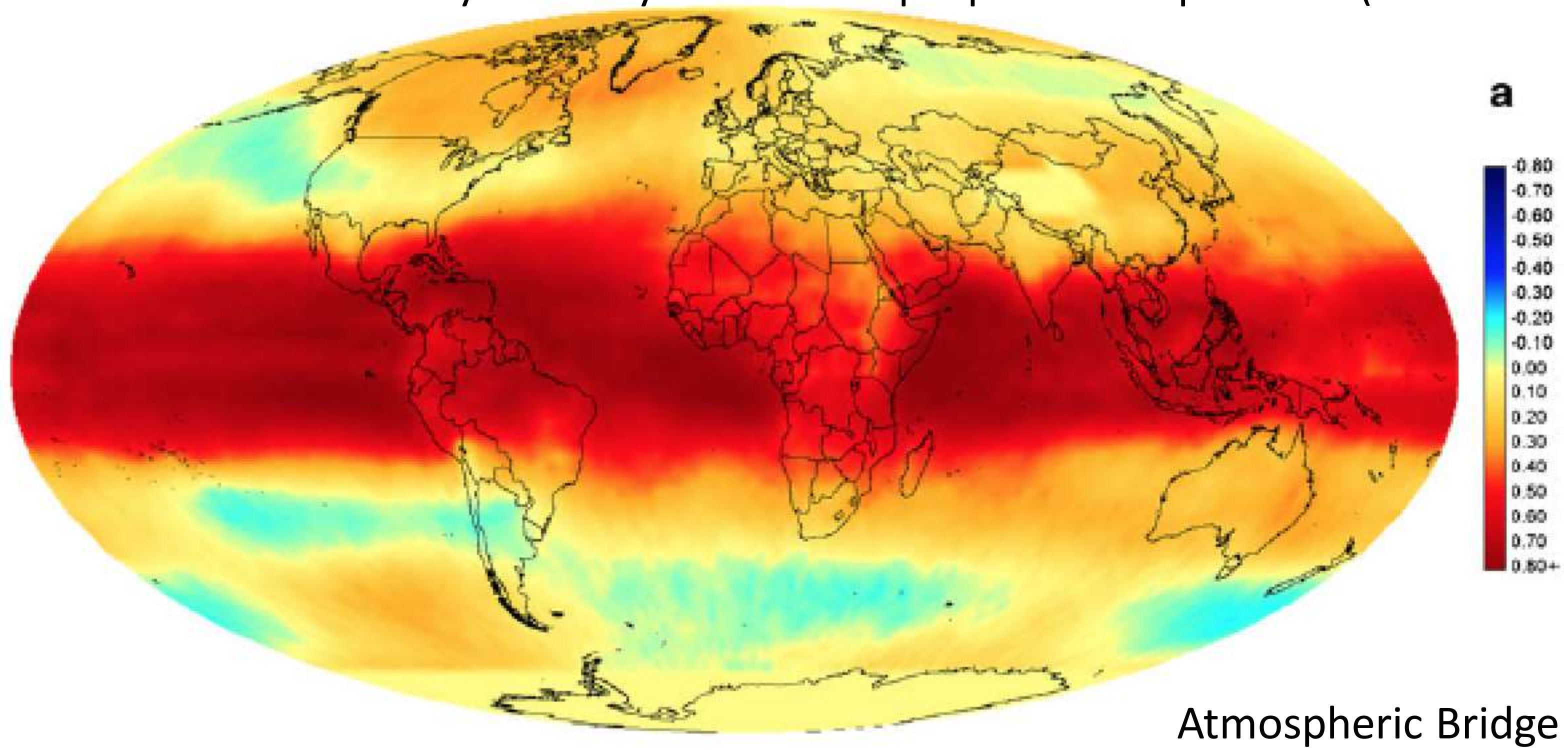
PCA on time series of monthly anomaly of lower tropospheric temperature (1982-2007)



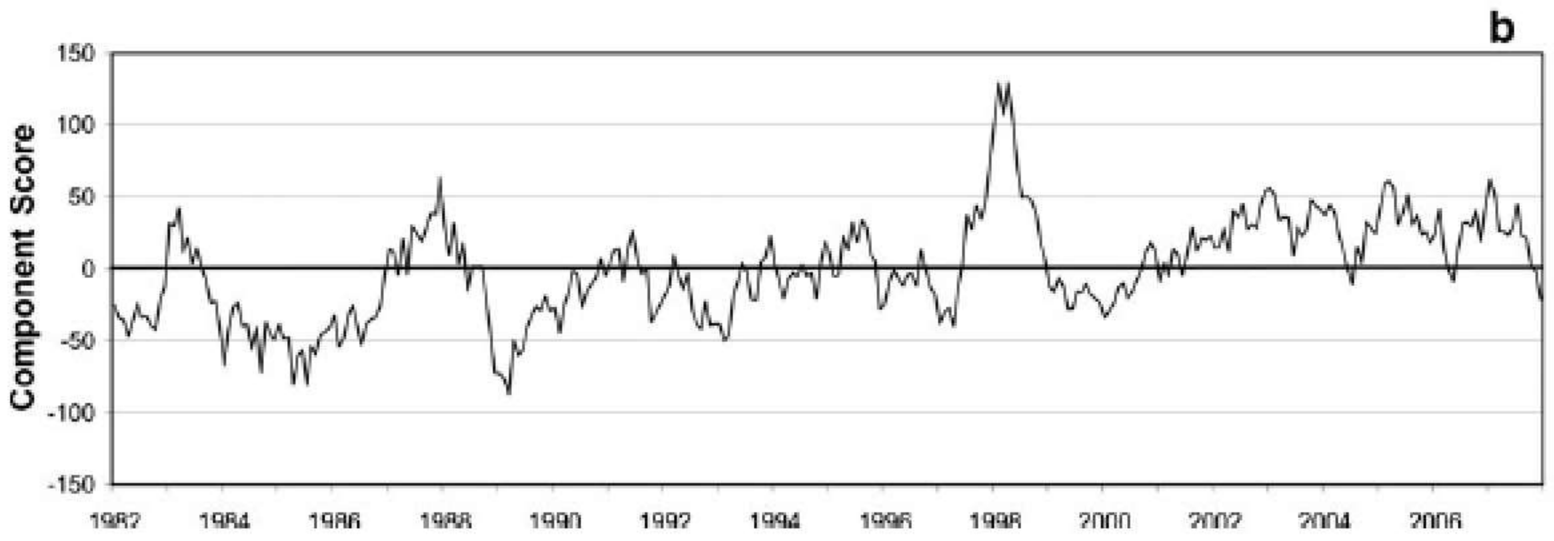
North Atlantic Oscillation



PCA on time series of monthly anomaly of lower tropospheric temperature (1982-2007)

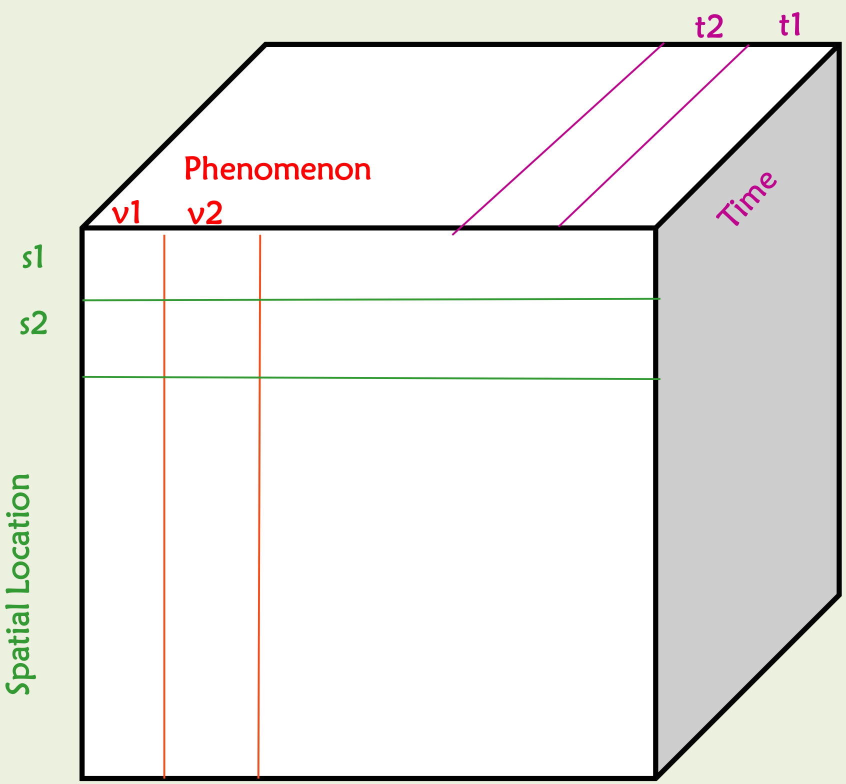


Atmospheric Bridge



Modes of Analysis in PCA

In PCA /Factor Analysis literature, there are different modes of analysis for analyzing multi-dimensional data which depends on how the data is organized called *orientation.*



(Source: Berry (1964))

Common Ways of Analyzing the Cube

(Cattell's Modes of Analysis)

Mode	Data Matrix
O	Phenomenon / Time
P	Time / Phenomenon
Q	Phenomenon / Location
R	Location / Phenomenon
S	Time / Location
T	Location / Time

Geometric Interpretation of PCA

- Objects are represented as a cloud of *n* points in a multidimensional space with an axis for each of the *p* variables
- The centroid of the points is defined by the mean of each variable
- The variance of each variable is the average squared deviation of its *n* values around the mean of that variable.

$$V_i = \frac{1}{n-1} \sum_{m=1}^{n} \left(X_{im} - \bar{X}_i \right)^2$$

Geometric Interpretation of PCA

- degree to which the variables are linearly correlated is represented by their covariances.

$$S_{ij} = \frac{1}{n-1}\sum_{m=1}^{n}\left(X_{im} - \bar{X}_i\right)\left(X_{jm} - \bar{X}_j\right)$$

Covariance of variables *i* and *j*

Sum over all *n* observations

Value of variable *I* for observation *m*

Mean of variable *i*

Value of variable *j* For observation *m*

Mean of variable *j*

Geometric Interpretation of PCA

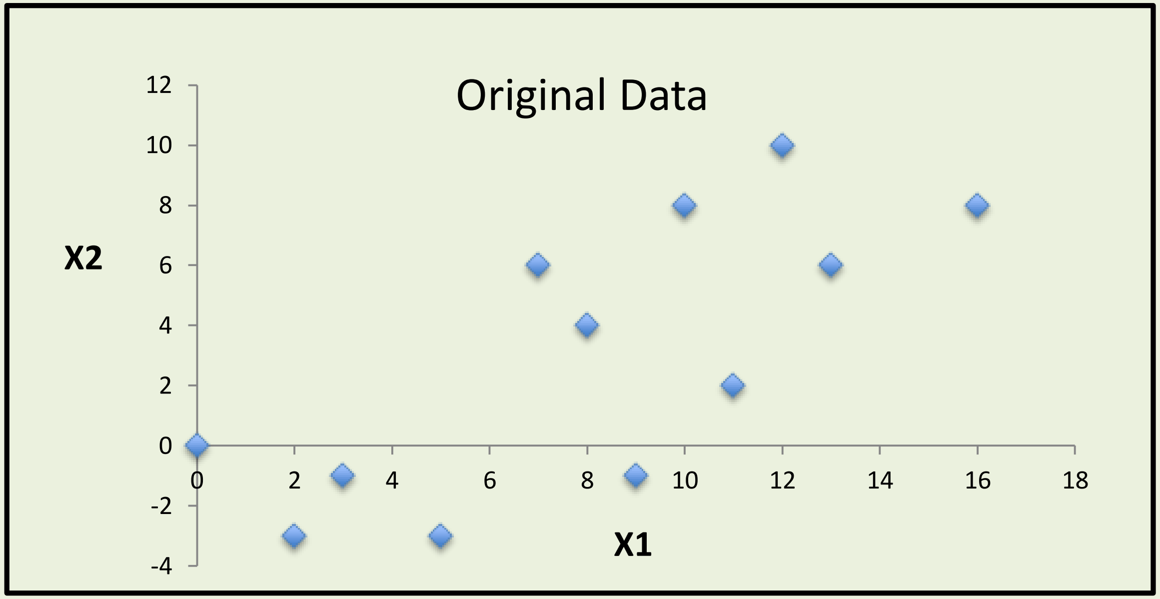
- Objective of PCA is to **rotate** the axes of this *p*-dimensional space to new positions (**principal axes**) that have the following properties:
  - ordered such that **principal axis 1 has the highest variance**, axis 2 has the next highest variance, ... , and axis *p* has the lowest variance
  - covariance among each pair of the principal axes is zero (**the principal axes are uncorrelated**).

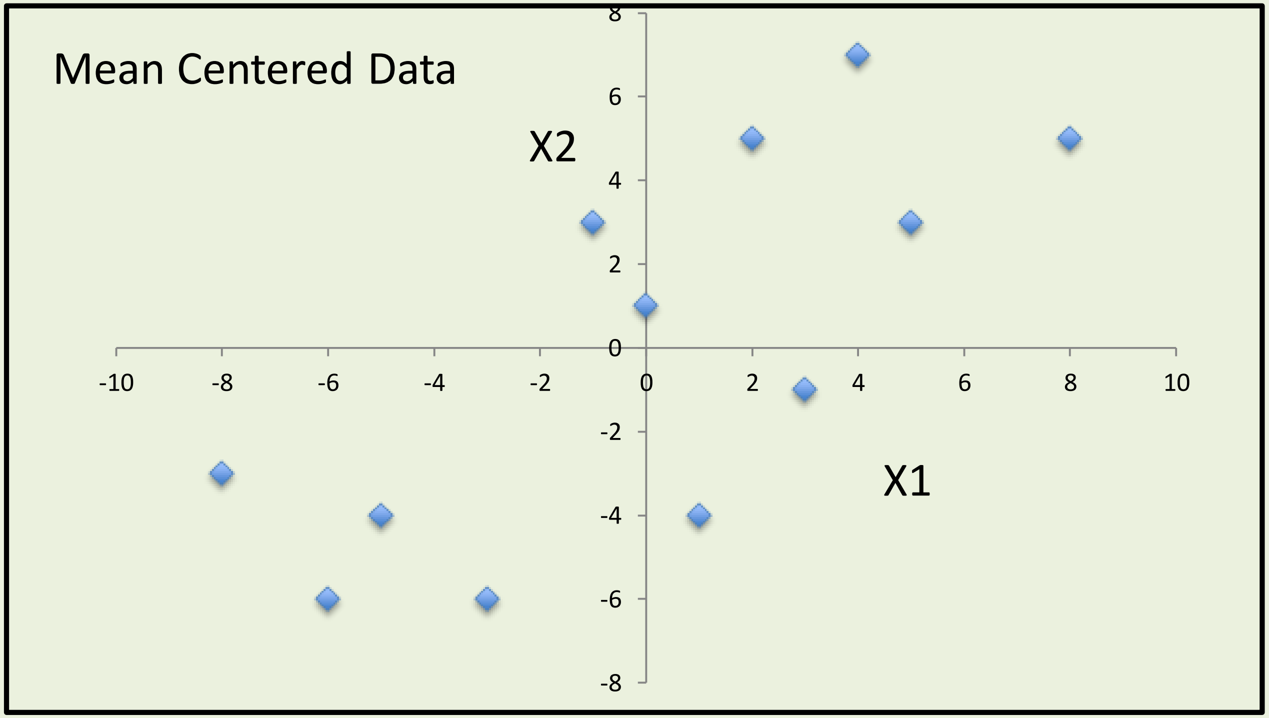
Identifying alternative axes which can explain maximum variance in data and forming new variables with respect to new set of axes

Example of PCA

Observation	X1	X2	Mean_corX1	Mean_corX2
1	16	8	8	5
2	12	10	4	7
3	13	6	5	3
4	11	2	3	-1
5	10	8	2	5
6	9	-1	1	-4
7	8	4	0	1
8	7	6	-1	3
9	5	-3	-3	-6
10	3	-1	-5	-4
11	2	-3	-6	-6
12	0	0	-8	-3
Mean	**8**	**3**	**0**	**0**
Variance	**23.09**	**21.09**	**23.09**	**21.09**

Total Variance: 44.182
Variance of X1 is 23.09
Variance of X2 is 21.09

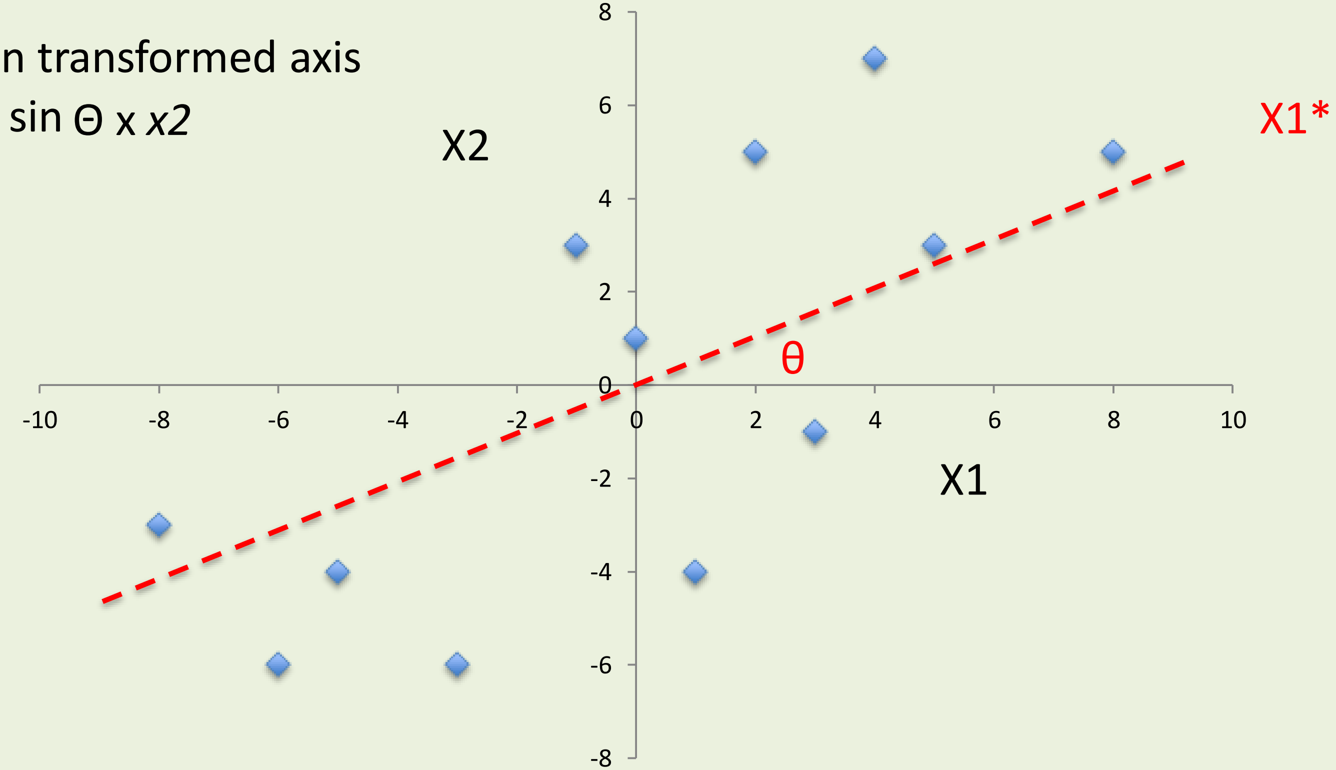

Original Data
X2
X1
12
10
8
6
4
2
0
-2
-4
0
2
4
6
8
10
12
14
16
18


Mean Centered Data
X2
X1
8
6
4
2
0
-2
-4
-6
-8
-10
-8
-6
-4
-2
0
2
4
6
8
10

Example of PCA

Total Variance: 44.182
Variance of X1 is 23.09 How do we maximize it?
Variance of X2 is 21.09

Any observation on transformed axis
$x1^* = \cos \Theta \times x1 + \sin \Theta \times x2$



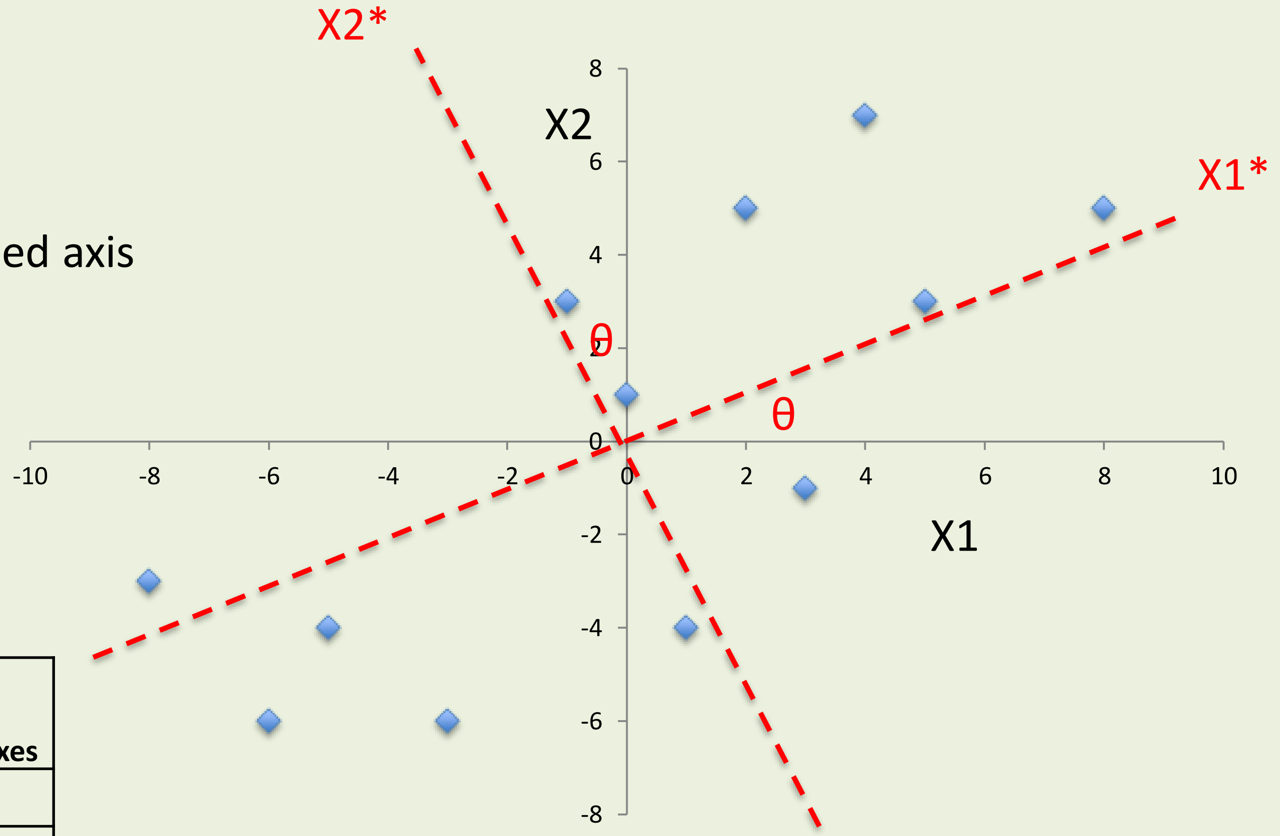
Rotation of Axis X1

Angle between X1 and X1*(theta)	Variance of X1*	% of variance explained
0	23.091	52.26336517
10	28.659	64.86578245
20	33.434	75.67335114
30	36.841	83.38463628
40	38.469	87.06939478
43.261	**38.576**	**87.31157485**
50	38.122	86.28400706
60	35.841	81.12127111
70	31.902	72.2058757
80	26.779	60.61065592
90	21.091	47.73663483

Example of PCA

Any observation on transformed axis
x1* = cos Θ x *x1* + sin Θ x *x2*

Any observation on transformed axis
x2* = -sin Θ x *x1* + cosΘ x *x2*



	Original	Rotated Axes
Variance X1	**23.091**	**38.576**
Variance X2	**21.09**	**5.61**

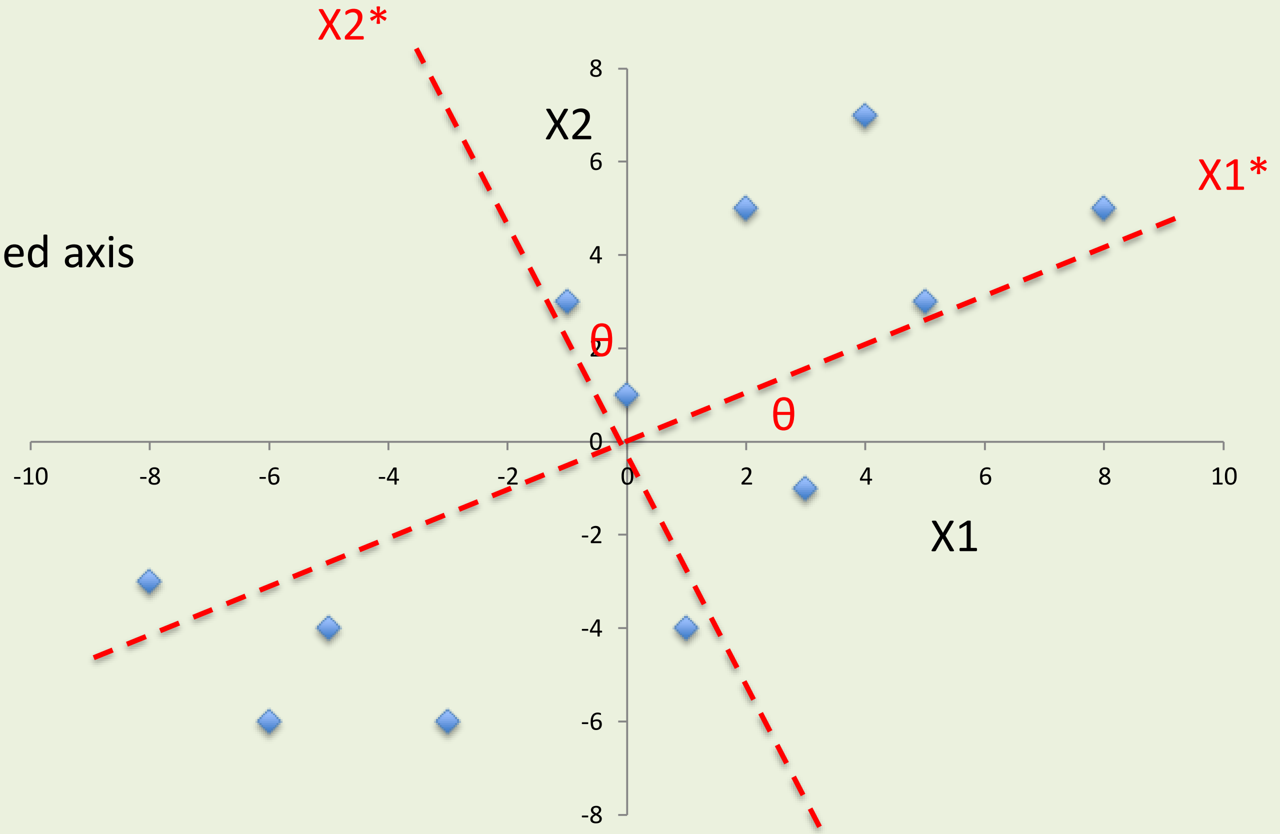
Total Variance: 44.182

Total variance did not change: Information content remains same

Example of PCA

Any observation on transformed axis
$x1^* = \cos \Theta \times x1 + \sin \Theta \times x2$

Any observation on transformed axis
$x2^* = -\sin \Theta \times x1 + \cos\Theta \times x2$



New axes
X1* and X2* are called **Principal Components**

Values x1* and x2* are called **principal component scores**

PCA-Dimension reduction

- In the example: For reducing dimension, lets just use only first transformed axes X1*
- Total variance explained = 38.576 (87.3%)
- We lose variance = 5.76 (12.68%)
- Representing data in lower dimensional space compared to original dimension is called **dimensional reduction**

Geometrically

- Objective of PCA is to identify *p* new sets of orthogonal axes for *p* variables such that:
  1. Each new variable is linear combination of original variables
  2. The first new variable explains maximum variance in the data
  3. Second new variable accounts for the maximum variance that is not explained by the first variable
  4. Third new variable: max variance not explained by first two variable and so on….
  5. The *p*th new variable accounts for the variance that has not been explained by the p-1 variables

PCA

- Principal Components
- Principal component score
- Loadings: Correlation between Principal component (new variable e.g., X1* and X2*) and original variable (X1, X2)
- Loading provides information on how influential was an original variable in forming new variable
- Higher loading, more influential the variable is in forming the new variable

Analytical Approach

- If there are p variables, the interest is to form p linear combinations:

$$
\begin{aligned}
y_1 &= w_{11}x_1 + w_{12}x_2 + \ldots\ldots\ldots + w_{1p}x_p \\
y_2 &= w_{21}x_1 + w_{22}x_2 + \ldots\ldots\ldots + w_{2p}x_p \\
&\vdots \\
y_p &= w_{p1}x_1 + w_{p2}x_2 + \ldots\ldots\ldots\ldots w_{pp}x_p
\end{aligned}
$$

where y_1, y_2 …..y_p are the p principal components and w_{ij} is the weight of the jth variable for the ith principal component.

Analytical approach

- Weights w_{ij} are estimated such that:
  1. The first principal component y1 accounts for maximum variance in the data, the second component accounts for maximum variance in data not explained by first component and so on…
  2. $w_{11}^2+w_{12}^2+\ldots.w_{ip}^2 = 1 \quad i=1\ldots\ldots.p$
  3. $w_{i1}w_{j1} + w_{i2}w_{j2}+\ldots..\ w_{ip}w_{jp} = 0 \quad \text{for } i \neq j$

- #2 is necessary is used to fix the scale of new variables as it is possible to increase the variance of a linear combination by change the scale of weights
- # 3 is to ensure new axes are orthogonal to each other

The Algebra of PCA

- First step is to calculate variance-covariance matrix *S* (correlation *R*) using all the p variables
- Such matrix will be square and symmetric
- Diagonals are the variances and off-diagonals are the covariances
- In Matrix term:

$$S = X'X$$

Where ‘ represents transpose

X is the nxp data matrix with each variable (mean centered or standardized)

The Algebra of PCA

- The sum of diagonals of the covariance matrix is called the **trace**
- Trace represents the total variance in the data
- It is the mean squared Euclidean distance between each observation and the centroid in p-dimensional space
- Finding the principal axes (principal components) involve eigenanalysis of the S or R matrix.
- The eigenvalues of S or R matrix are solutions to characteristics equation

$$|S - \lambda I| = 0$$

$$(S - \lambda I)U = 0$$

Where λ is eigen value, U is eigen vector

The Algebra of PCA

- The eigen values $\lambda_1, \lambda_2 \ldots. \lambda_p$ are the variances of the coordinates in each principal component axis
- The sum of all p eigen values equal to trace of S (the sum of the variances of all of the original variables)
- Each eigen vector consists of p values which represent the "contribution" of each variable to the principal component axis
- Eigen vectors are uncorrelated (orthogonal)
- Principal Scores can be written as

$$y_{ki} = u_{1k}x_{1i} + u_{2k}x_{2i} + \ldots\ldots\ldots.. + u_{pk}x_{pi}$$

Where y is the n x k matrix of Principal component (PC) scores

X is the n x p centered data matrix U is the p x k matrix of eigen vector

The Algebra of PCA

- The variance of the scores on each PC axis is equal to the corresponding eigen value for that axis
- The eigen value for an axis k represents the amount of variance explained by the k^{th} axis
- The variance-covariance matrix of S for PCs will have all off-diagonal elements 0 and diagonal elements (variance) are the eigen values λ extracted using equation $|S - \lambda I| = 0$

Covariance vs Correlation matrix

- Covariance is used only if all the variables are in the same unit
- Drawback of using covariance is that principal components will be dominated with variables with high variances
- This issue can be overcome by standardizing each variables to unit variance and zero
- Covariance matrix calculation using standardized variables give correlation matrix

How do we perform PCA?

- In R there are two commands

1. prcomp (SVD approach for R/S matrix decomposition)
2. Princomp (eigen vector based R/S matrix decompositon)

Interpreting PCA outputs

Input Data

Observation	X1	X2	Mean_corX1	Mean_corX2
1	16	8	8	5
2	12	10	4	7
3	13	6	5	3
4	11	2	3	-1
5	10	8	2	5
6	9	-1	1	-4
7	8	4	0	1
8	7	6	-1	3
9	5	-3	-3	-6
10	3	-1	-5	-4
11	2	-3	-6	-6
12	0	0	-8	-3
Mean	**8**	**3**	**0**	**0**
Variance	**23.09**	**21.09**	**23.09**	**21.09**

Results

1. Simple Statistics: Mean and Standard Deviation for original variables

2. Variance Covariance Matrix

3a. Eigenvalues

3b. Eigen vectors

4. Table of original variables and PCs with simple statistics

5. Loadings

6. Principal Components

1 SIMPLE STATISTICS

	X1	X2
MEAN	8.00000	3.00000
ST DEV	4.80530	4.59248

2 COVARIANCES

	X1	X2
X1	23.09091	16.45455
X2	16.45455	21.09091

TOTAL VARIANCE=44.18182

3a	EIGENVALUE	DIFFERENCE	PROPORTION	CUMULATIVE
PRIN1	38.5758	32.9698	0.873115	0.87312
PRIN2	5.6060	.	0.126885	1.00000

3b EIGENVECTORS

	PRIN1	PRIN2
X1	0.728238	-.685324
X2	0.685324	0.728238

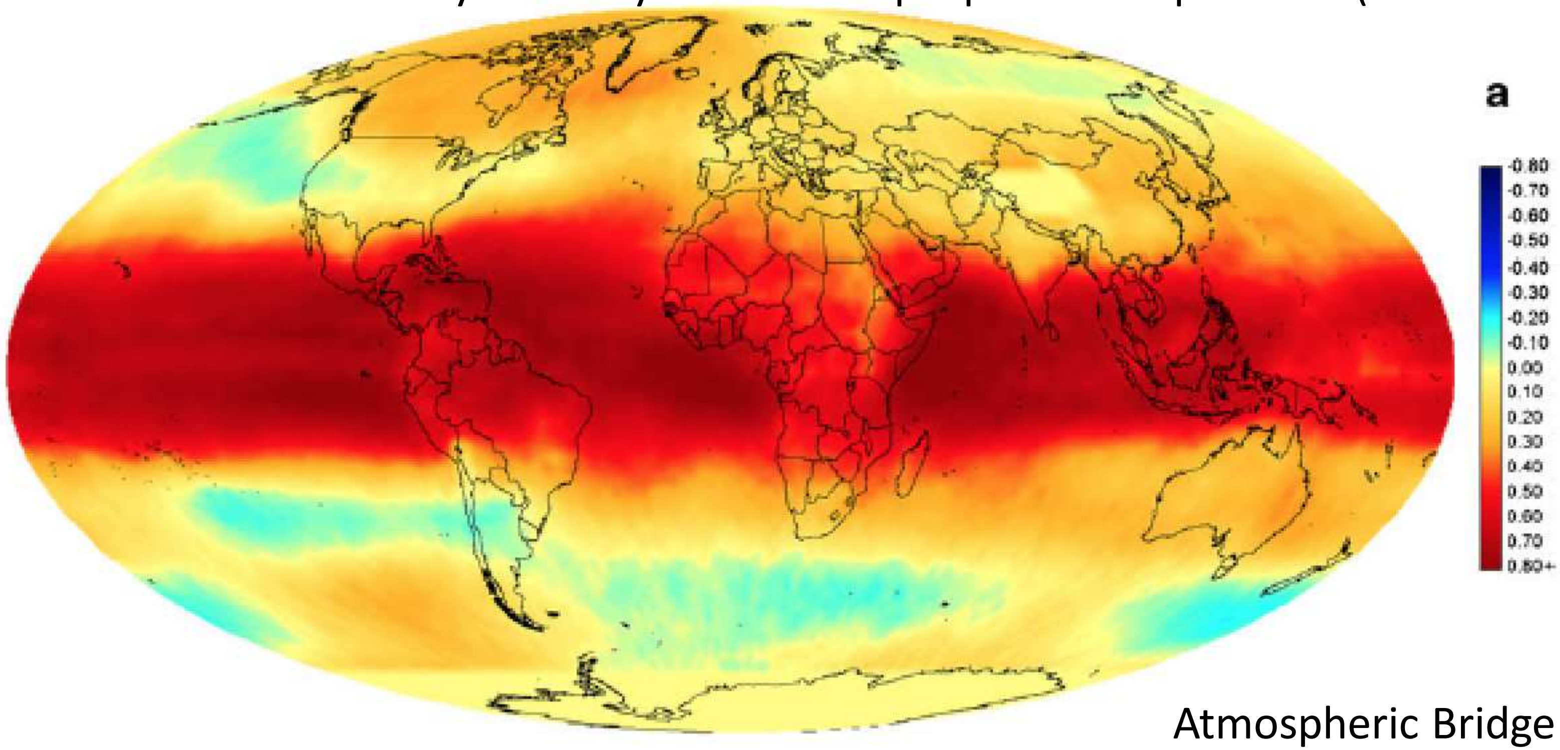
4 VARIABLE	N	MEAN	STD DEV
X1	12	8.000000	4.805300
X2	12	3.000000	4.592484
PRIN1	12	-8.697E-16	6.210943
PRIN2	12	0.000000	2.367700

5 PEARSON CORRELATION COEFFICIENTS

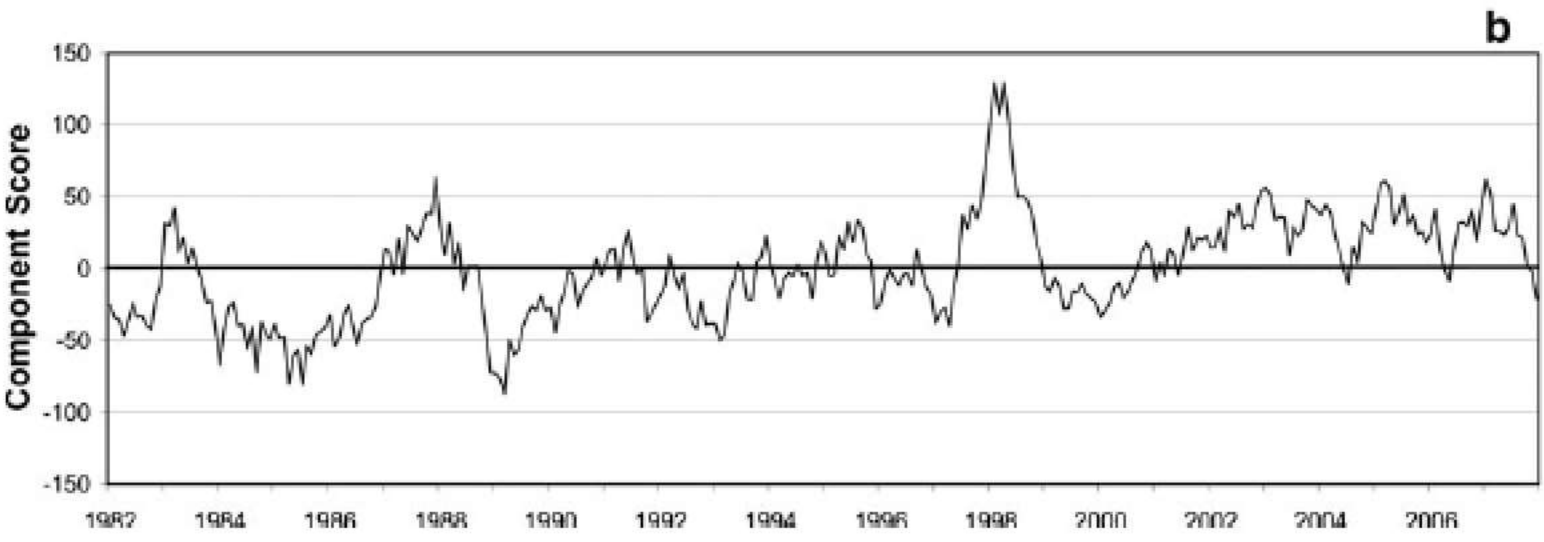
	X1	X2	PRIN1	PRIN2
X1	1.00000	0.74562	0.94126	-0.33768
X2	0.74562	1.00000	0.92684	0.37545
PRIN1	0.94126	0.92684	1.00000	0.00000
PRIN2	-0.33768	0.37545	0.00000	1.00000

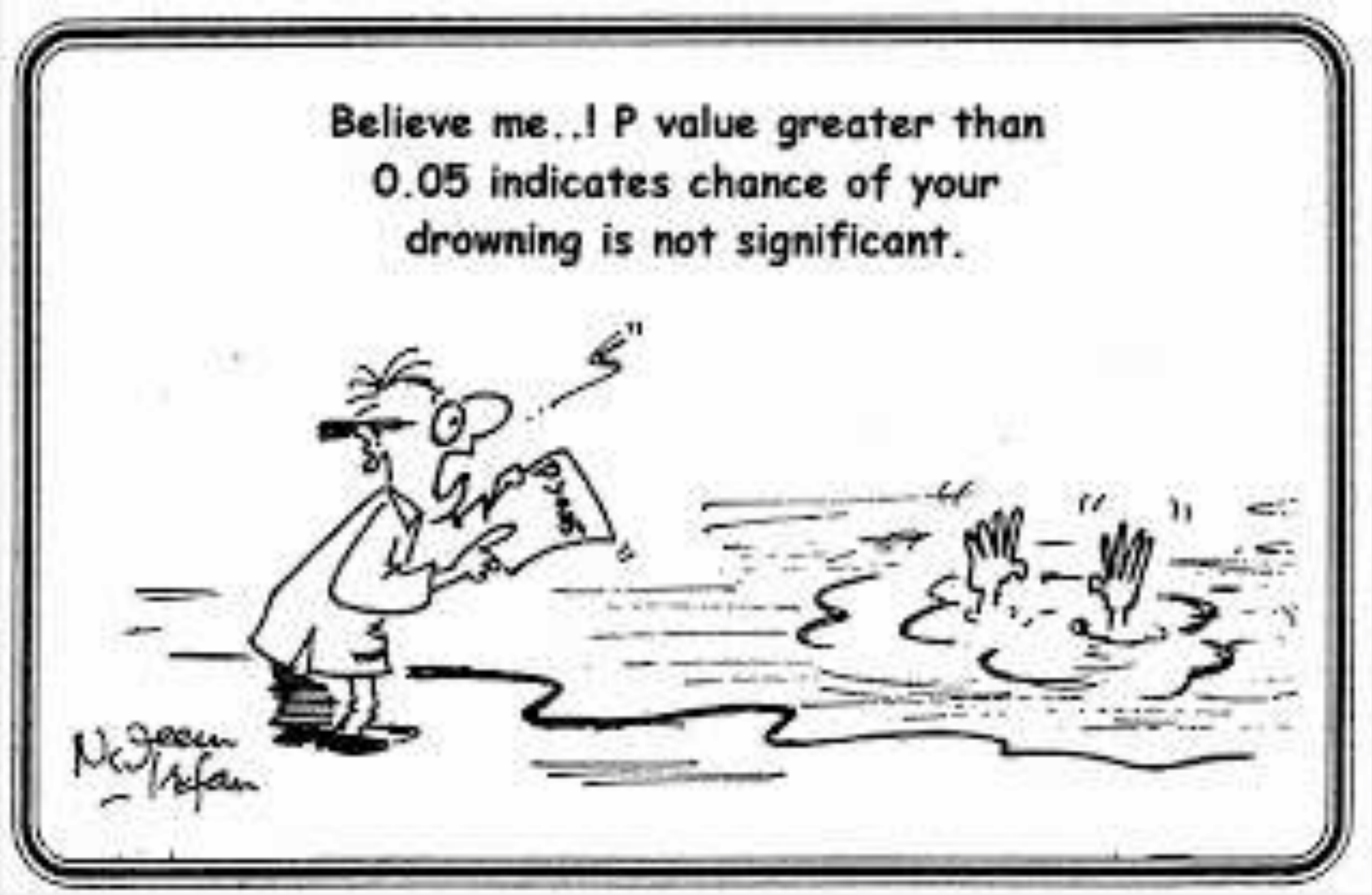
6 OBS	X1	X2	PRIN1	PRIN2
1	16	8	9.2525	-1.8414
2	12	10	7.7102	2.3564
3	13	6	5.6972	-1.2419
4	11	2	1.4994	-2.7842
5	10	8	4.8831	2.2705
6	9	-1	-2.0131	-3.5983
7	8	4	0.6853	0.7282
8	7	6	1.3277	2.8700
9	5	-3	-6.2967	-2.3135
10	3	-1	-6.3825	0.5137
11	2	-3	-8.4814	-0.2575
12	0	0	-7.8819	3.2979

PCA on time series of monthly anomaly of lower tropospheric temperature (1982-2007)



Atmospheric Bridge




Believe me..! P value greater than
0.05 indicates chance of your
drowning is not significant.