# Research Methods in Climate Science

Lecturer:

Babatunde J. Abiodun

University of Cape Town

South Africa

# Research: Think, Analyse, and Write

The purpose of this course is to motivate you to:

- Think: a series of lecture/discussions on
    - conceptualization of creative and original research topic
    - proposal writing
    - research plan, design and implementation

- Analyse: the nuts and bolts of quantitative analysis
    - The quantitative analysis sessions
    - not a basic statistics course
    - The primary objective is data manipulations, decomposition and interpretations
    - The techniques investigated are (mostly) the fundamental methods found commonly in the literature.

- Write: the basic rules of academic writing and research
    - Conscious critical reading and writing skills
    - read academic texts with greater insight and apply these to your writing

# Course Contents: Analysis section

- The content will include:
  - Data Classification
    - Cluster analysis
    - Principal Component Analysis (PCA)
    - Self Organizing Maps
  - Time series
    - Trend Analysis and Detrending
    - Auto-correlation Analysis
    - Fourier Analysis
    - Wavelet Analysis
  - The Grand Challenge

# Data Classification and Cluster Analysis

Babatunde J. Abiodun
babiodun@csag.uct.ac.za

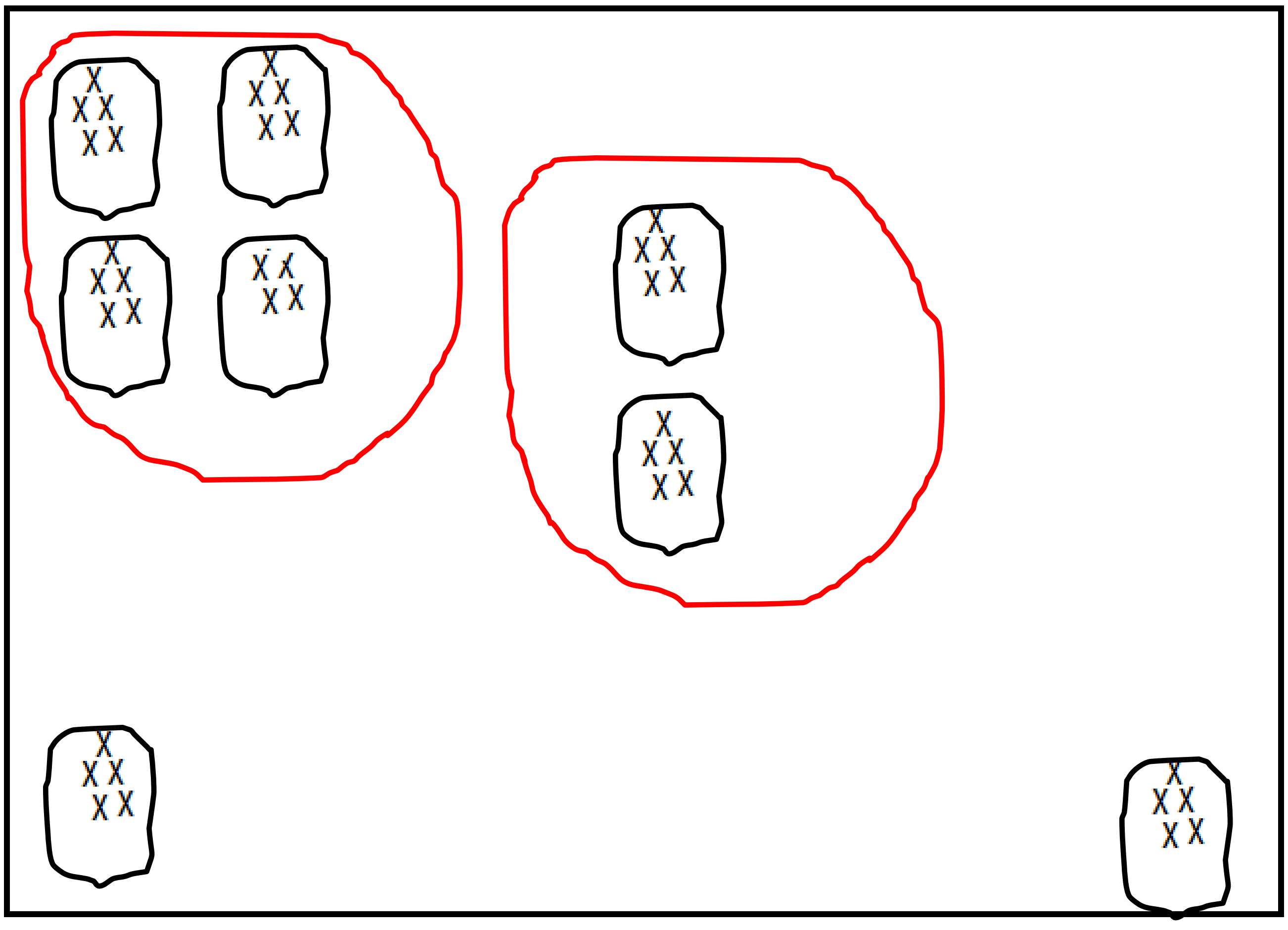
# Data Classification

# Data Classification

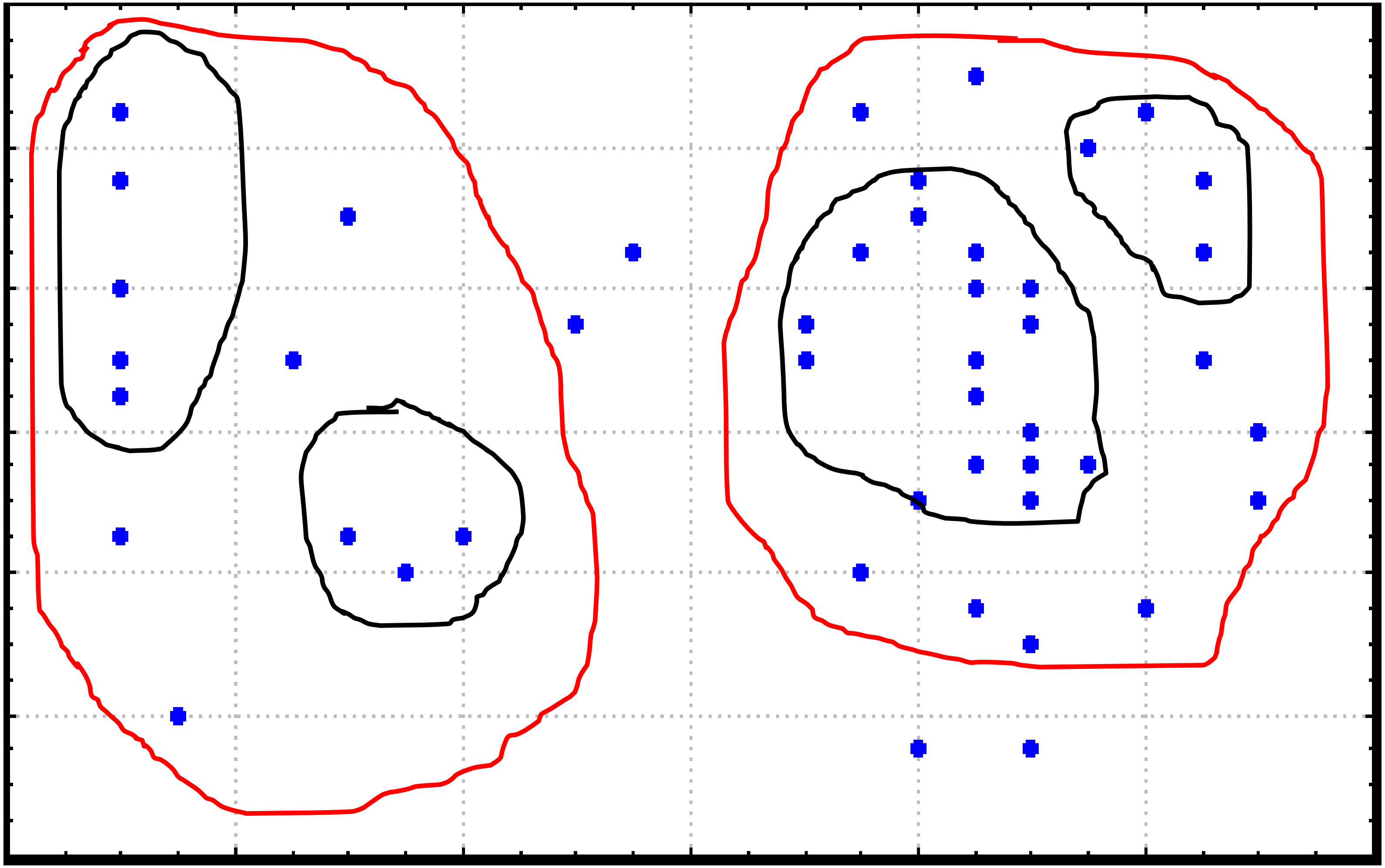
- Generalize data into groups of similar type, to reduce the data size.

- Objective:
  - To identify groups in parameter space
  - User specifiable measure of similarity
  - Place boundaries through the space
  - Ideal solution: minimize within group variance and maximize between group variance.

# Example (1)

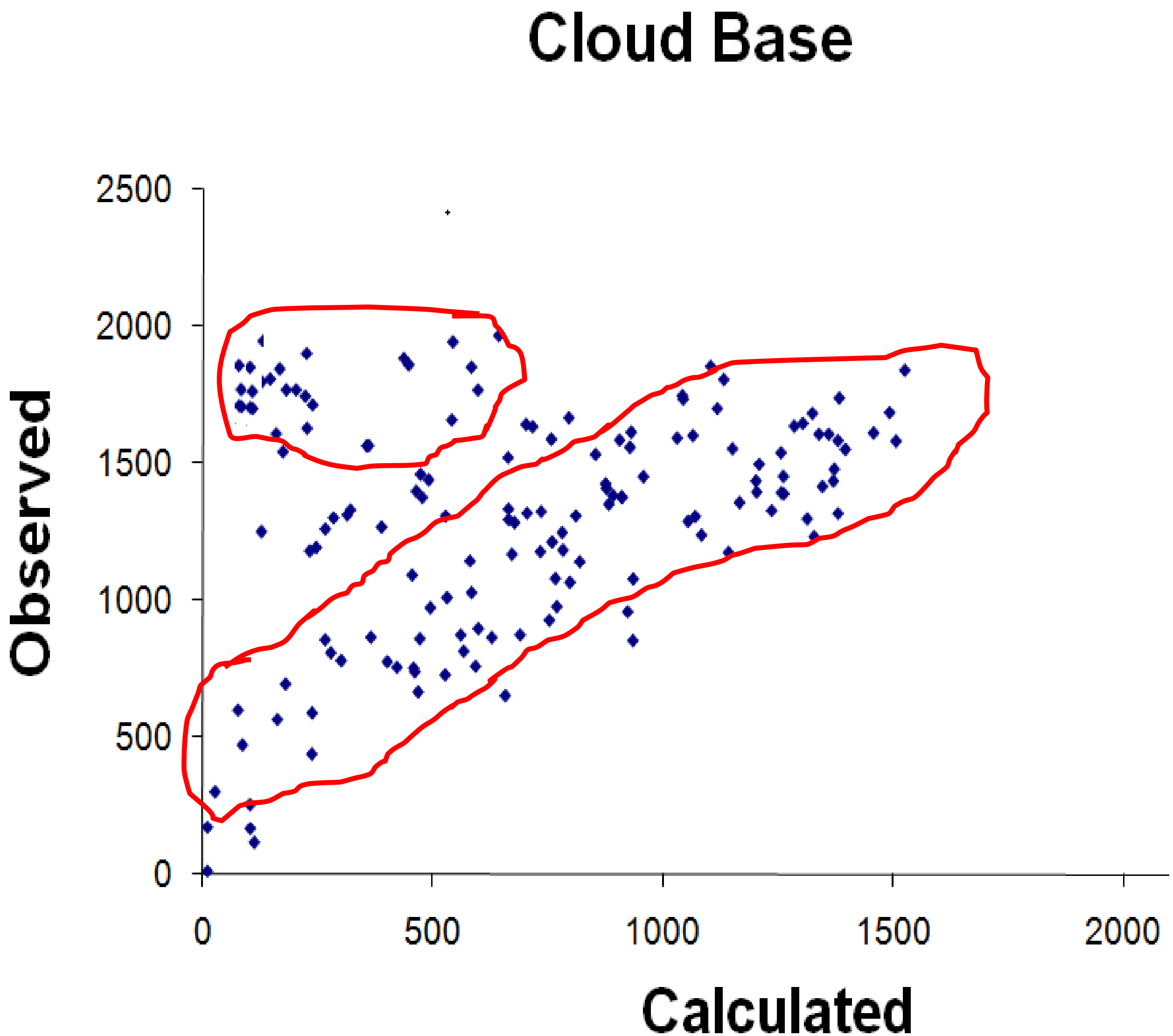


# Example(2)



- The data may be grouped by drawing boundaries between clusters...
- some data points may be poor fits...
- some may be outliers...
- the positioning of the boundaries is subjective...

# Example (3)



Cloud Base
2500
2000
1500
1000
500
0
Observed
0
500
1000
1500
2000
Calculated

# Main Issues...

- looking for general groups
- identifying outliers
- accuracy <--> number of clusters

# Multiple procedures available... Which is right?

- The "right" procedure is the one that aids your interpretation i.e. give insight into what is in the data.
- Any classification method is subjective
- Ideal solution depends on user specification of algorithm, measure of similarity, etc.
- Know your data!
- Manual classification (given manageable data volumes) is often as good as anything else. [See journal paper by Key and Crane (1986)].

Key J, Crane RG. (1986): A comparison of synoptic classification schemes based on 'objective' procedures. *Journal of Climatology* ***6:*** 375–388.

# Objective procedures

- Automated clustering is rule based. Data points are "joined" based on some measure of distance (e.g.: Euclidian distance).

- ***Note:  Assumption of orthogonality: What are the implications.... !!!!!!!!!!!!!***

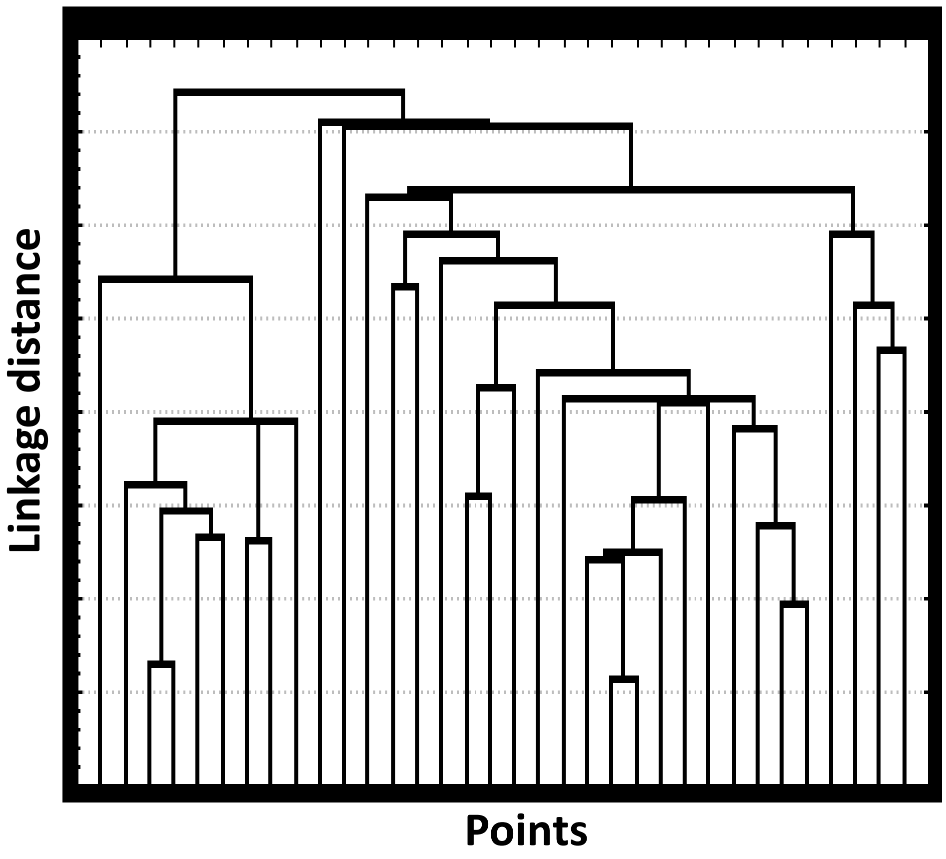
# Cluster Analysis

# Cluster Analysis

- There are numerous algorithms to undertake clustering, but fall into two basic categories: -
  – Clustering where the number of groups are defined by the user after the procedure
  – Clustering where the user pre-defines the number of groups
- We will focus on the first option
  – single linkage,
  – average linkage
  – Wards algorithm

# Single linkage

- <u>Single linkage:</u>
  - start with the two closest points and link
  - find next two closest and link
  - find next two .....
  .
  - until all data points are linked
- On completion one can generate a cluster tree, whereby the linkage between data points is graphically shown as a function of the distance between the points:
- **<u>Note single linkage problem:</u>** leads to chaining ..... the *"next closest point"* is determined by measuring unclustered points against all points in a cluster.....
- With a cluster tree, first decision: Where to cut the tree.....
- look for jumps in the distances between joins....
- **<u>Thus:</u>** single linkage tends to identify outliers....



# Average linkage

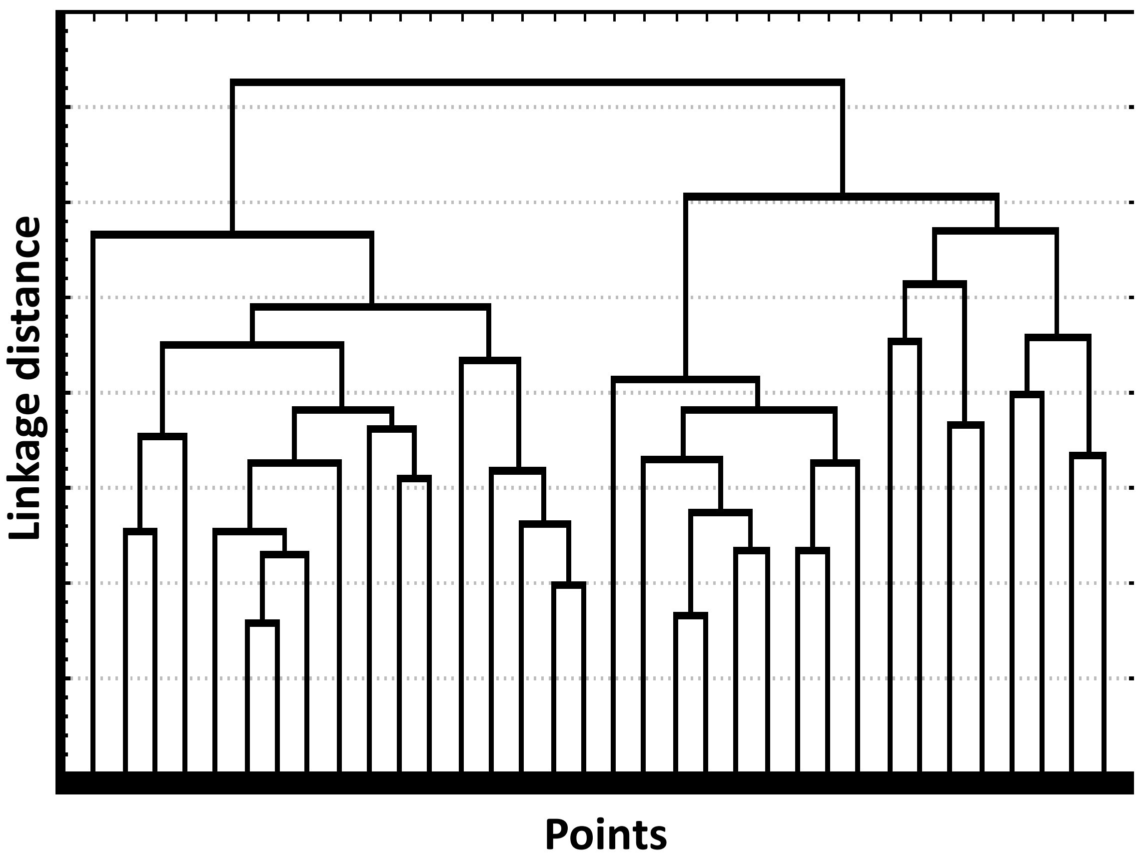
A better routine to find groups is to use the centroid or mean of a cluster when measuring new points.

Start clustering as before, but for any group of data points, represent these by a single point with a value calculated as the centroid of all points in the group.

**Thus:**

- Outliers still identified

- Groups here are better defined, but clear breaks in groups still not very clear.



# Wards algorithm

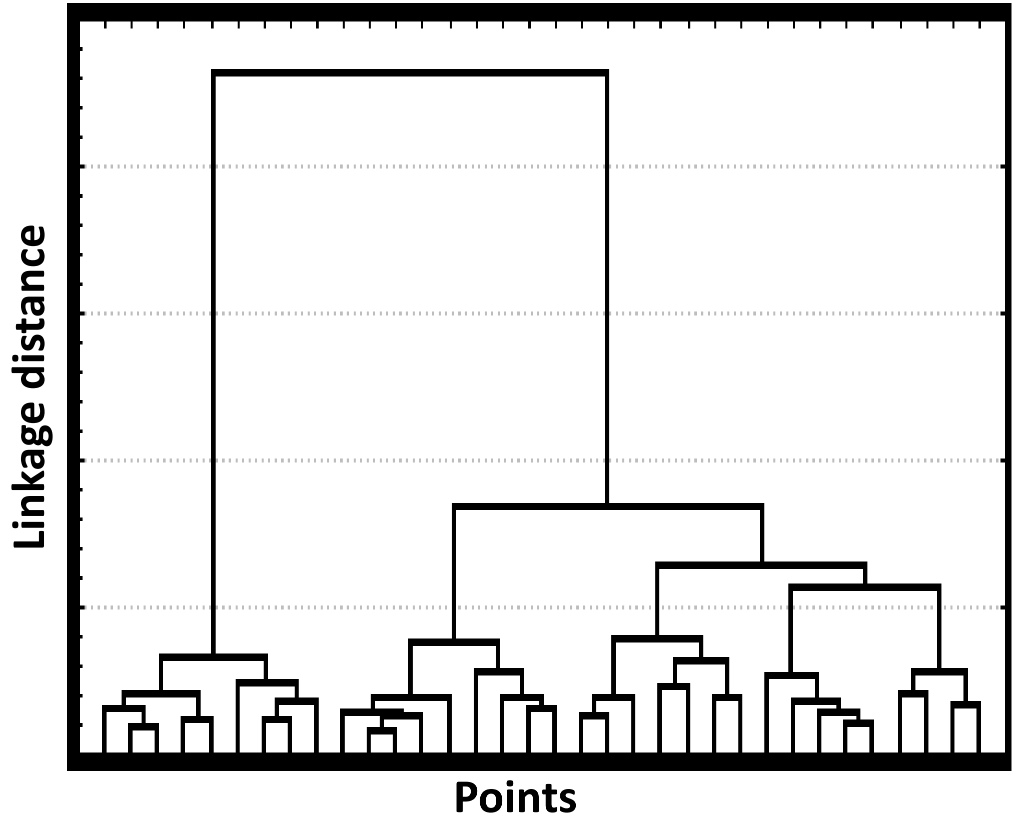
Another procedure is Wards, based on analysis of variance.

This seeks to minimise within group variance and maximise between group variance.

Data points are joined to a particular group on a basis of whether their addition to this group improves the within group variance more than addition to some other group.

**Thus:**

- Groups are well defined, of more equal size.
- Outliers are not well identified



# Practical
# Using Statistical or other software

- Notes:

  - In your data matrix, you may cluster rows <u>or</u> columns -- this influences the dimensionality of your data space
  - Choice of algorithm is subjective and influences the result
  - Euclidian distances which assume orthogonality are generally used
  - Other more sophisticated non-linear techniques will be explored later.

# Principal Component Analysis

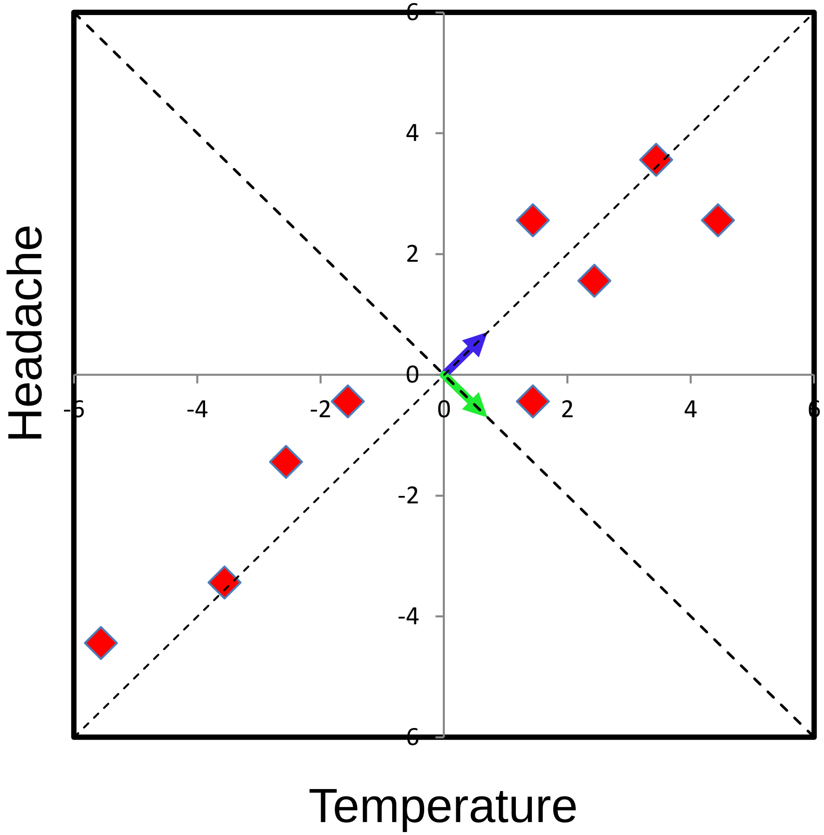
# PCA: Aliases

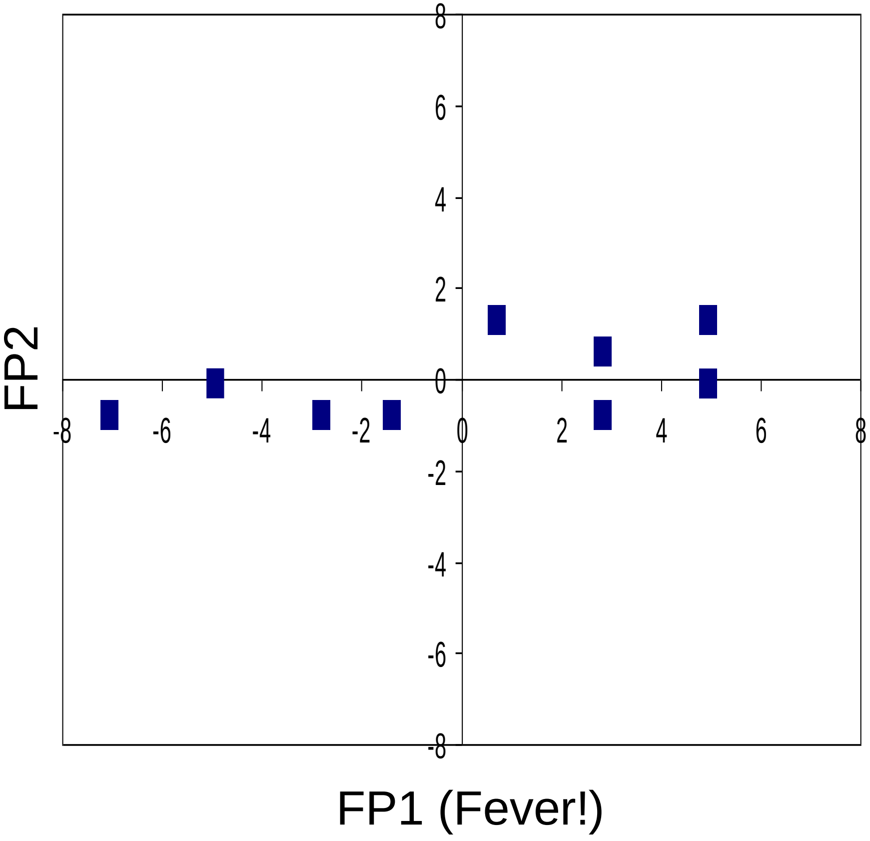
- Eigenvector Analysis
- Factor Analysis
- Empirical Orthogonal Function (EOF)

# PCA: What is it?

- A procedure to ....
  - extract fewer and independent underlying dimensions around which the data variance is organised -- *the "real" or pseudo-variables (components)*
  - express the original variables in their relation to these components
  - re-express the original data as if measured in terms of these components.

# Example (1)PCA (Old data vs New data)



# PCA (2): Example

| Variable | OBSERVED | | | CLM | | | GEMLAM | | | MRCC | | | RCA3 | | | RSM | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | PF1 | PF 2 | PF 3 | PF1 | PF 2 | PF 3 | PF1 | PF 2 | PF 3 | PF1 | PF 2 | PF 3 | PF1 | PF 2 | PF 3 | PF1 | PF 2 | PF 3 |
| NTR | 0.96 | 0.12 | 0.04 | 0.96 | 0.11 | 0.00 | 0.96 | 0.16 | -0.01 | 0.92 | 0.25 | -0.03 | 0.92 | 0.23 | 0.07 | 0.95 | 0.18 | -0.02 |
| SHF | 0.88 | 0.08 | -0.16 | 0.84 | 0.06 | 0.31 | 0.83 | 0.26 | -0.27 | 0.89 | 0.12 | -0.24 | 0.83 | 0.17 | -0.34 | 0.87 | 0.10 | -0.30 |
| LHF | 0.94 | 0.15 | 0.17 | 0.90 | 0.11 | -0.24 | 0.90 | 0.03 | 0.21 | 0.88 | 0.29 | -0.05 | 0.77 | 0.28 | 0.38 | 0.86 | 0.30 | 0.00 |
| CBH | 0.78 | 0.17 | 0.05 | 0.85 | 0.26 | 0.06 | 0.51 | 0.62 | -0.16 | 0.80 | 0.25 | -0.16 | 0.52 | -0.16 | -0.10 | 0.81 | 0.27 | 0.02 |
| T2M | 0.48 | 0.82 | 0.09 | 0.44 | 0.85 | 0.04 | 0.41 | 0.87 | 0.06 | 0.46 | 0.86 | -0.06 | 0.39 | 0.86 | 0.07 | 0.42 | 0.87 | 0.06 |
| QV | 0.15 | 0.92 | 0.05 | 0.21 | 0.90 | 0.03 | 0.19 | 0.78 | 0.22 | 0.28 | 0.93 | -0.04 | 0.07 | 0.92 | 0.10 | 0.16 | 0.93 | 0.04 |
| SMI | -0.01 | -0.81 | 0.04 | 0.04 | -0.76 | 0.05 | 0.12 | -0.80 | 0.10 | -0.28 | -0.05 | 0.62 | 0.05 | -0.74 | 0.13 | -0.22 | -0.80 | 0.16 |
| PS | 0.15 | -0.27 | -0.78 | 0.12 | -0.18 | 0.75 | 0.10 | -0.15 | -0.80 | 0.11 | -0.17 | -0.76 | 0.18 | -0.22 | -0.71 | 0.15 | -0.22 | -0.77 |
| VABS | 0.20 | -0.21 | 0.83 | 0.09 | -0.20 | -0.86 | 0.07 | -0.10 | 0.81 | 0.10 | -0.33 | 0.74 | 0.09 | -0.23 | 0.83 | 0.02 | -0.31 | 0.82 |
| Tot. Var (%) | 38.84 | 26.17 | 15.04 | 38.18 | 25.35 | 16.29 | 32.46 | 27.81 | 16.61 | 38.17 | 21.94 | 17.81 | 28.76 | 26.92 | 16.59 | 36.92 | 29.06 | 15.31 |

# Talk about....

- Background mathematical concepts
  - Statistics
  - Matrix algebra

- Principal Component Analysis (PCA)

# Basic mathematical concepts (1): Statistics

- Mean: $$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$
- Standard Deviation: $$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^{2}}{(n-1)}}$$
- Variance: $$s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^{2}}{(n-1)}$$

# Basic mathematical concepts (1): Statistics

- Covariance: $$cov(x,y) = \frac{\sum_{i=1}^{n}(x_i - \bar{x})\,(y_i - \bar{y})}{(n-1)}$$

- Covariance Matrix:

C(x,y,z) = $$\begin{pmatrix} cov(x,x) & cov(x,y) & cov(x,z) \\ cov(y,x) & cov(y,y) & cov(y,z) \\ cov(z,x) & cov(z,y) & cov(z,z) \end{pmatrix}$$

# Basic Mathematical concept (2): Matrix Algebra

- **Eigenvector:** Given a matrix, an eigenvector of that matrix is a nonzero vector which, when the matrix is multiplied with, may change in length, but not direction.

- **Eigenvalue:** For each eigenvector, there is a corresponding eigenvalue, which is a scalar value that determines the amount the eigenvector is scaled when multiplied with the matrix.

# Basic Mathematical concept (2): Example of Eigenvector and eigenvalue

$$\begin{pmatrix} 2 & 3 \\ 2 & 1 \end{pmatrix} \times \begin{pmatrix} 1 \\ 3 \end{pmatrix} = \begin{pmatrix} 11 \\ 5 \end{pmatrix}$$

$$\begin{pmatrix} 2 & 3 \\ 2 & 1 \end{pmatrix} \times \begin{pmatrix} 3 \\ 2 \end{pmatrix} = \begin{pmatrix} 12 \\ 8 \end{pmatrix} = 4 \times \begin{pmatrix} 3 \\ 2 \end{pmatrix}$$

Hence:

$\begin{pmatrix} 1 \\ 3 \end{pmatrix}$ is not an eigenvector of $\begin{pmatrix} 2 & 3 \\ 2 & 1 \end{pmatrix}$

$\begin{pmatrix} 3 \\ 2 \end{pmatrix}$ is an eigenvector of $\begin{pmatrix} 2 & 3 \\ 2 & 1 \end{pmatrix}$ with is an eigenvalue of 4

# Properties of eigenvectors

- Can only be found for a square matrix
- *n x n* marix, that has eigenvectors, will have *n* of them.
- Eigenvectors are perpendicular (i.e. orthogonal)
- Use eigenvectors of that have a unit length.

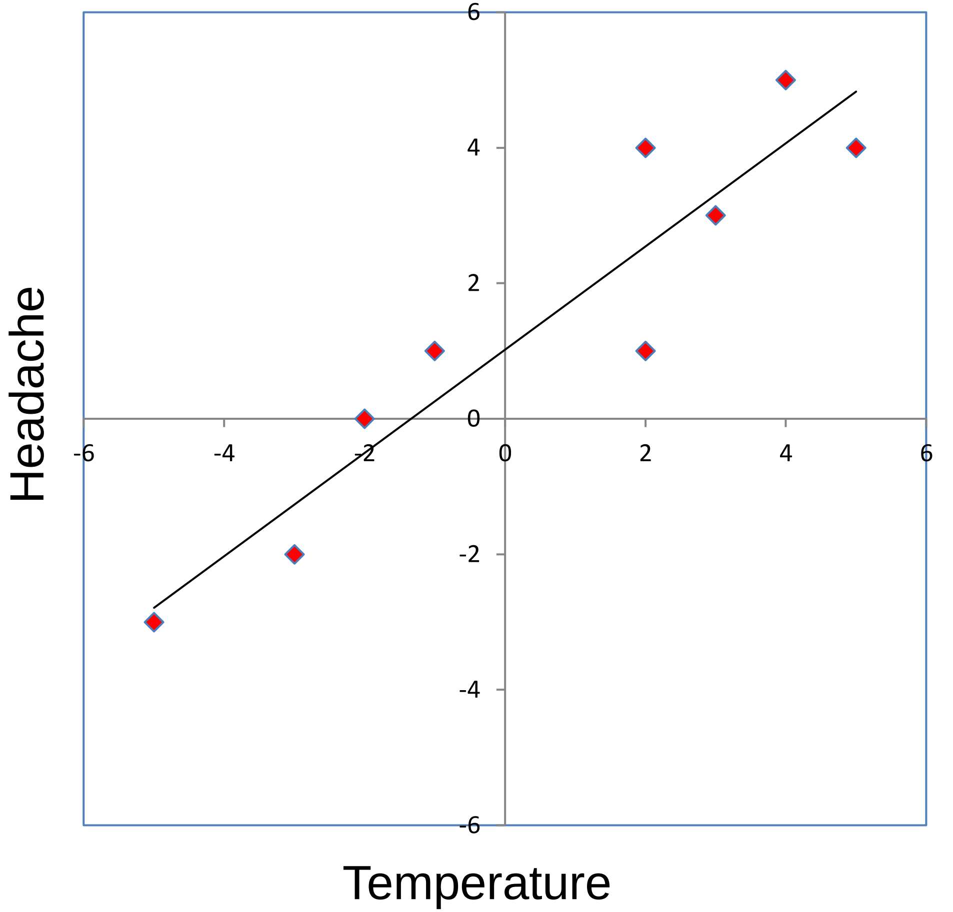
# Principal Component Analysis

- Method (6 steps)
- PCA with Statistical

# PCA Method: Step 1

- Get some data

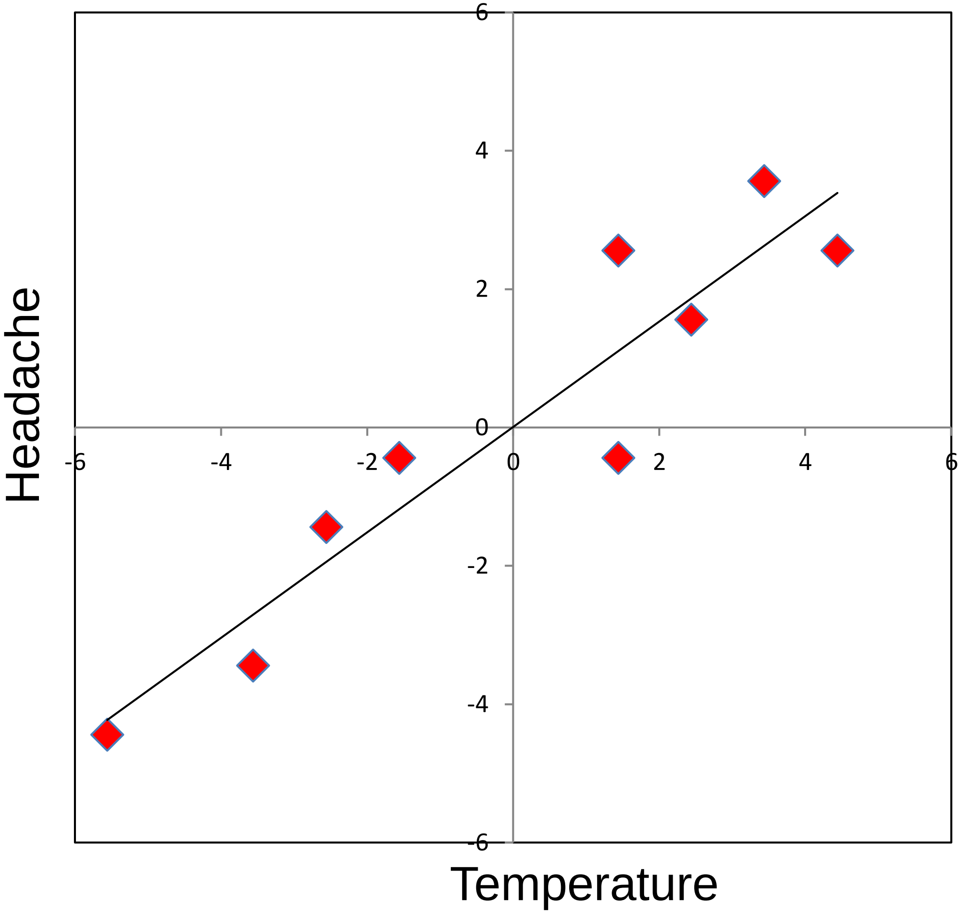
| Temperature | Headache |
|---|---|
| 4 | 5 |
| 2 | 4 |
| -3 | -2 |
| -5 | -3 |
| 2 | 1 |
| -1 | 1 |
| 3 | 3 |
| -2 | 0 |
| 5 | 4 |



# PCA Method: Step 2

- Standardize the data (i.e. subtract the mean)

| Temperature | Headache |
|---|---|
| 3.44 | 3.56 |
| 1.44 | 2.56 |
| -3.56 | -3.44 |
| -5.56 | -4.44 |
| 1.44 | -0.44 |
| -1.56 | -0.44 |
| 2.44 | 1.56 |
| -2.56 | -1.44 |
| 4.44 | 2.56 |



# PCA Method: Step 3

- Calculate the covariance matrix

| | Temperature | Headache |
|---|---|---|
| Temperature | 11.7778 | 8.9722 |
| Headache | 8.9722 | 7.7778 |

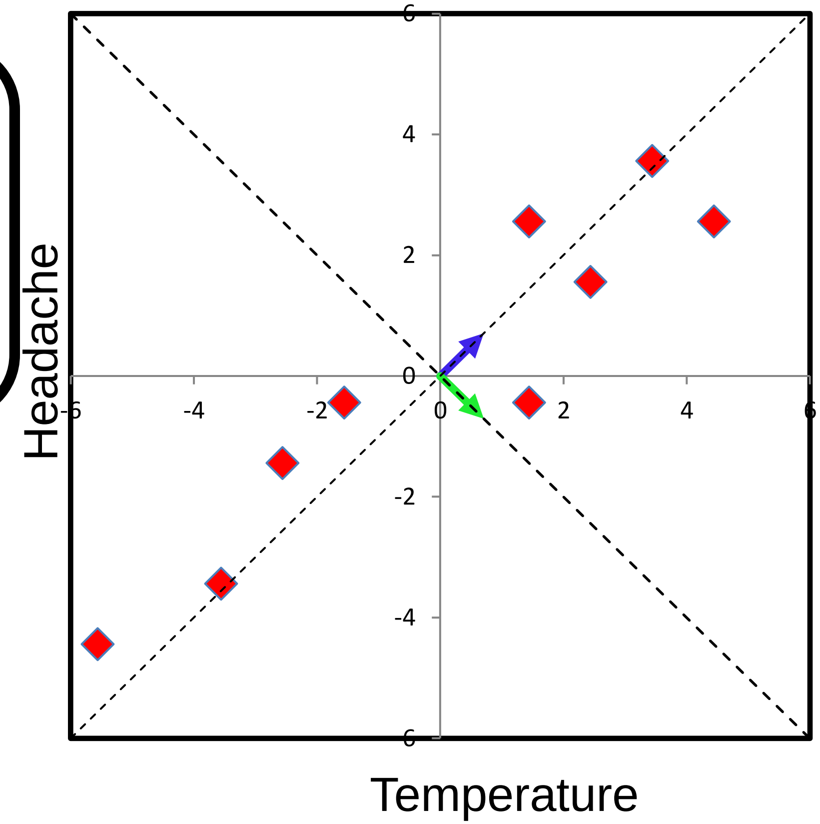
$$C = \begin{pmatrix} 11.7778 & 8.9722 \\ 8.9722 & 7.7778 \end{pmatrix}$$

# PCA Method: Step 4

- Calculate the eigenvector and eigenvalues of the covariance matrix

$$\text{Eigenvectors} = \begin{pmatrix} \boxed{\begin{matrix} 0.7071 \\ 0.7071 \end{matrix}} & \boxed{\begin{matrix} 0.7071 \\ -0.7071 \end{matrix}} \end{pmatrix}$$

$$\text{Eigenvalues} = \begin{pmatrix} \boxed{1.9374} \\ \boxed{0.0626} \end{pmatrix}$$



# PCA Method: Step 5

- Choose components and forming a feature vector.

$$\text{Eigenvalues} = \begin{pmatrix} 1.9374 \\ 0.0626 \end{pmatrix} \quad \text{OR} \quad \begin{pmatrix} 1.9374 \end{pmatrix}$$

$$\Downarrow \qquad\qquad\qquad \Downarrow$$

$$\text{Feature vector} = \begin{pmatrix} 0.7071 & 0.7071 \\ 0.7071 & -0.7071 \end{pmatrix} \qquad \begin{pmatrix} 0.7071 \\ 0.7071 \end{pmatrix}$$

# Loadings

| | PF1 | PF2 |
|---|---|---|
| Temperature | 0.7071 | 0.7071 |
| Headache | 0.7071 | -0.7071 |

Eigenvalues = 1.9374
0.0626

# PCA Method: Step 6

- Deriving the new data set

FinalDataSet = FeatureVector (transposed) x AdjustedDataSet (transposed)

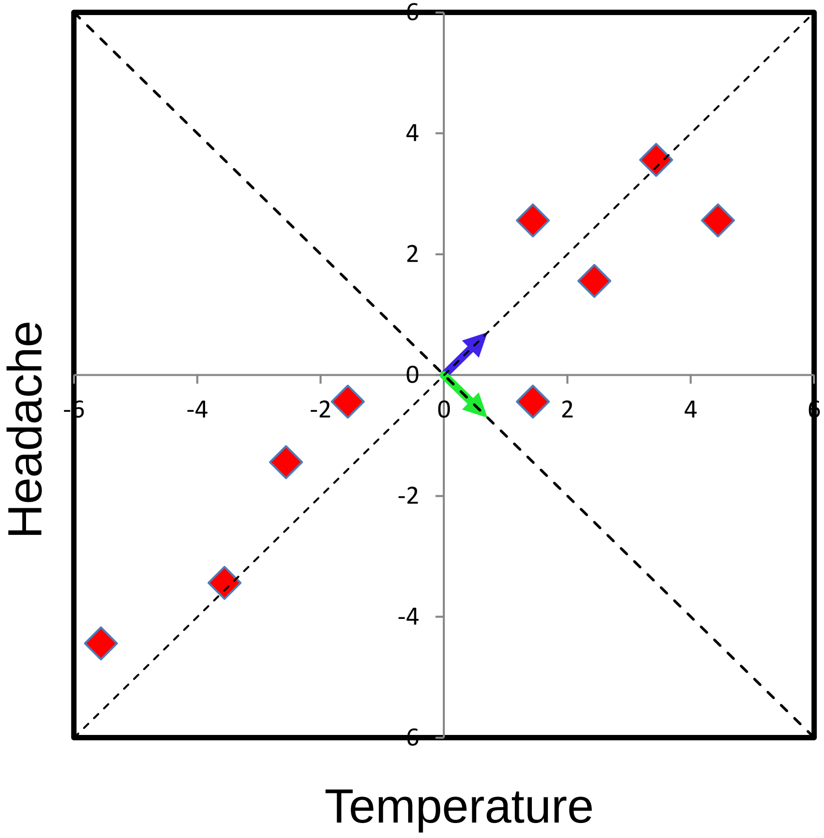
e.g.:

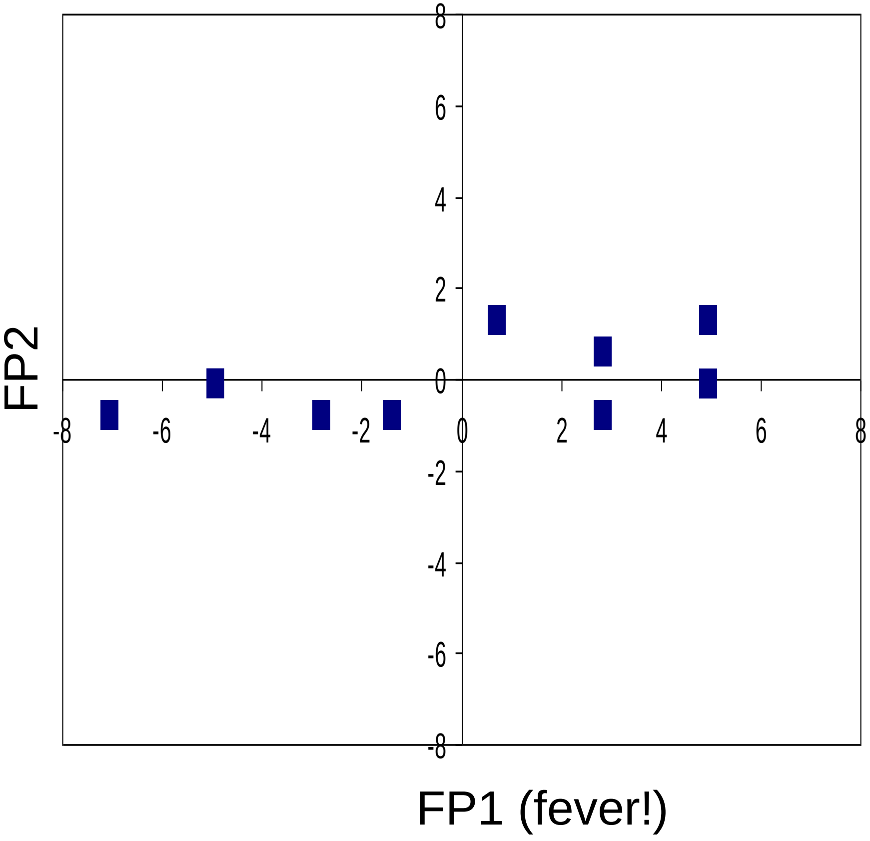
| FP1 | FP2 |
|---|---|
| 4.95 | -0.08 |
| 2.83 | -0.79 |
| -4.95 | -0.08 |
| -7.07 | -0.79 |
| 0.71 | 1.33 |
| -1.41 | -0.79 |
| 2.83 | 0.62 |
| -2.83 | -0.79 |
| 4.95 | 1.33 |

$$= \begin{pmatrix} 0.7071 & 0.7071 \\ 0.7071 & -0.7071 \end{pmatrix} \times$$

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Temp | 3.44 | 1.44 | -3.56 | -5.56 | 1.44 | -1.56 | 2.44 | -2.56 | 4.44 |
| Headach e | 3.56 | 2.56 | -3.44 | -4.44 | -0.44 | -0.44 | 1.56 | -1.44 | 2.56 |

# PCA (Old data vs New data)



# Some Applications

# PCA Part 2:
# Some Applications

Theor Appl Climatol
DOI 10.1007/s00704-014-1336-3

ORIGINAL PAPER

# Impacts of drought on grape yields in Western Cape, South Africa

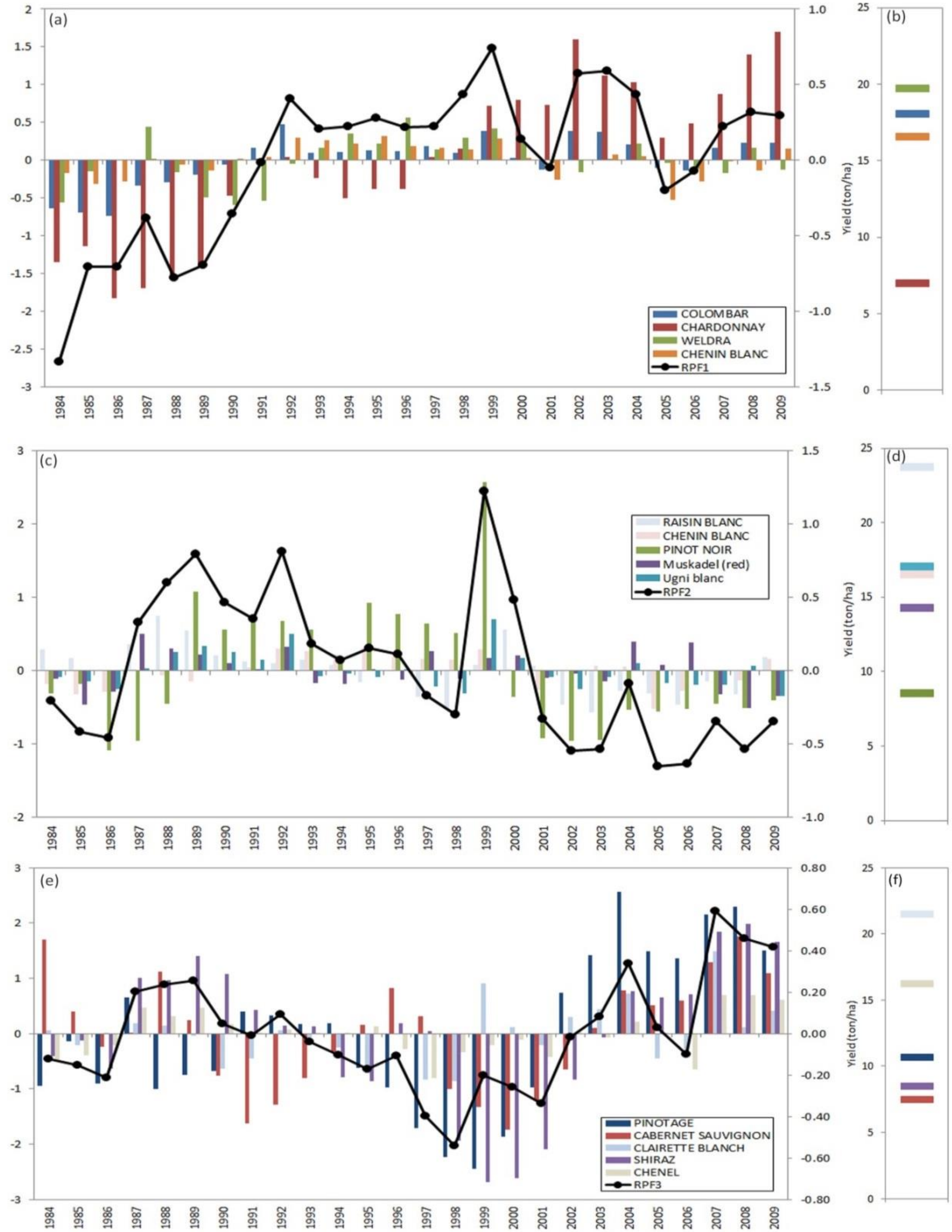
Julio A. Araujo • Babatunde J. Abiodun • Olivier Crespo

Received: 7 May 2014 / Accepted: 8 December 2014
© Springer-Verlag Wien 2014

# Yield Groups

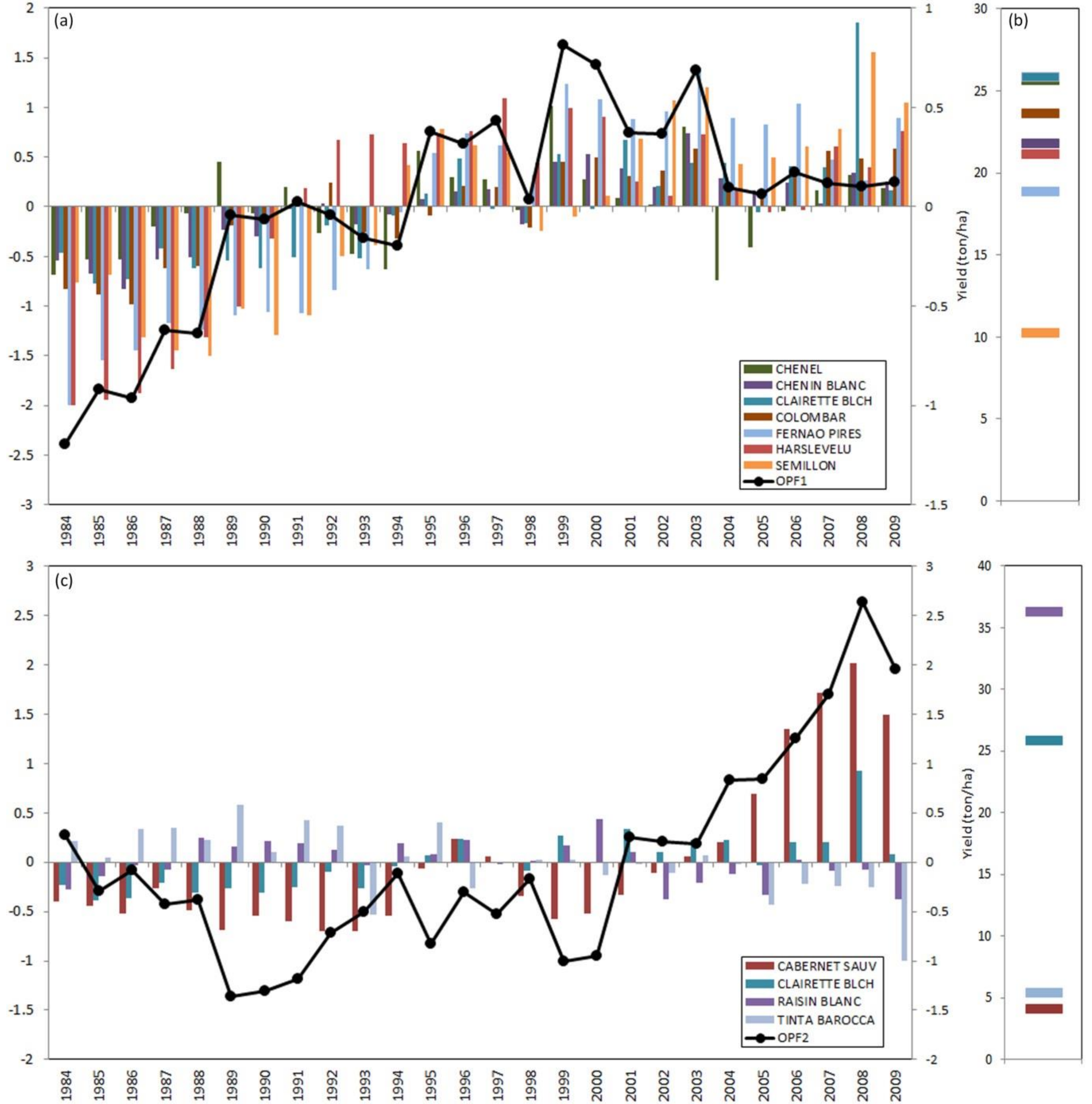
Robertson

- PCA jointly explains 62% of variance in the data



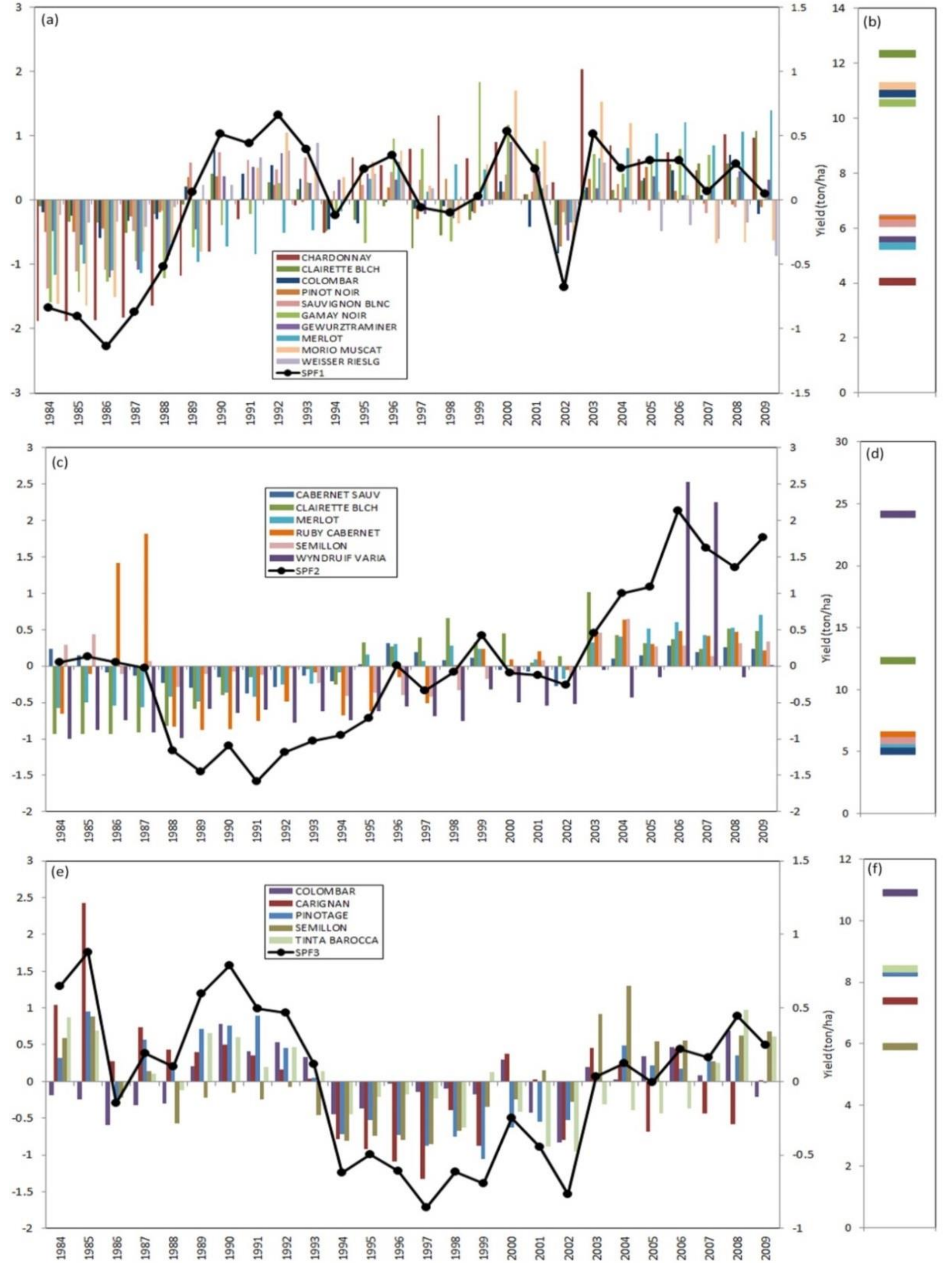
Olifants River

- PCA jointly explains 83% of variance in the data

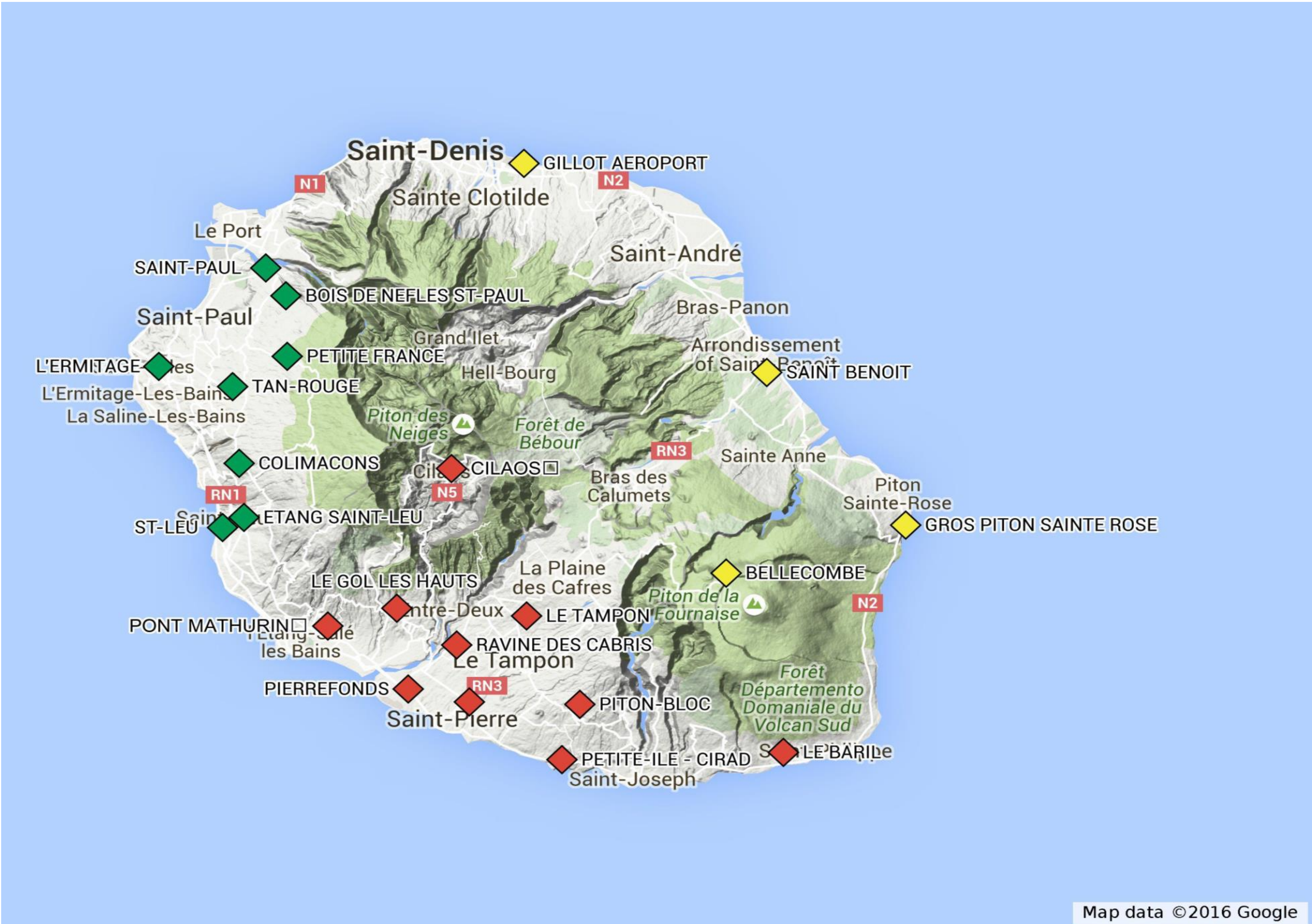


Stellenbosch

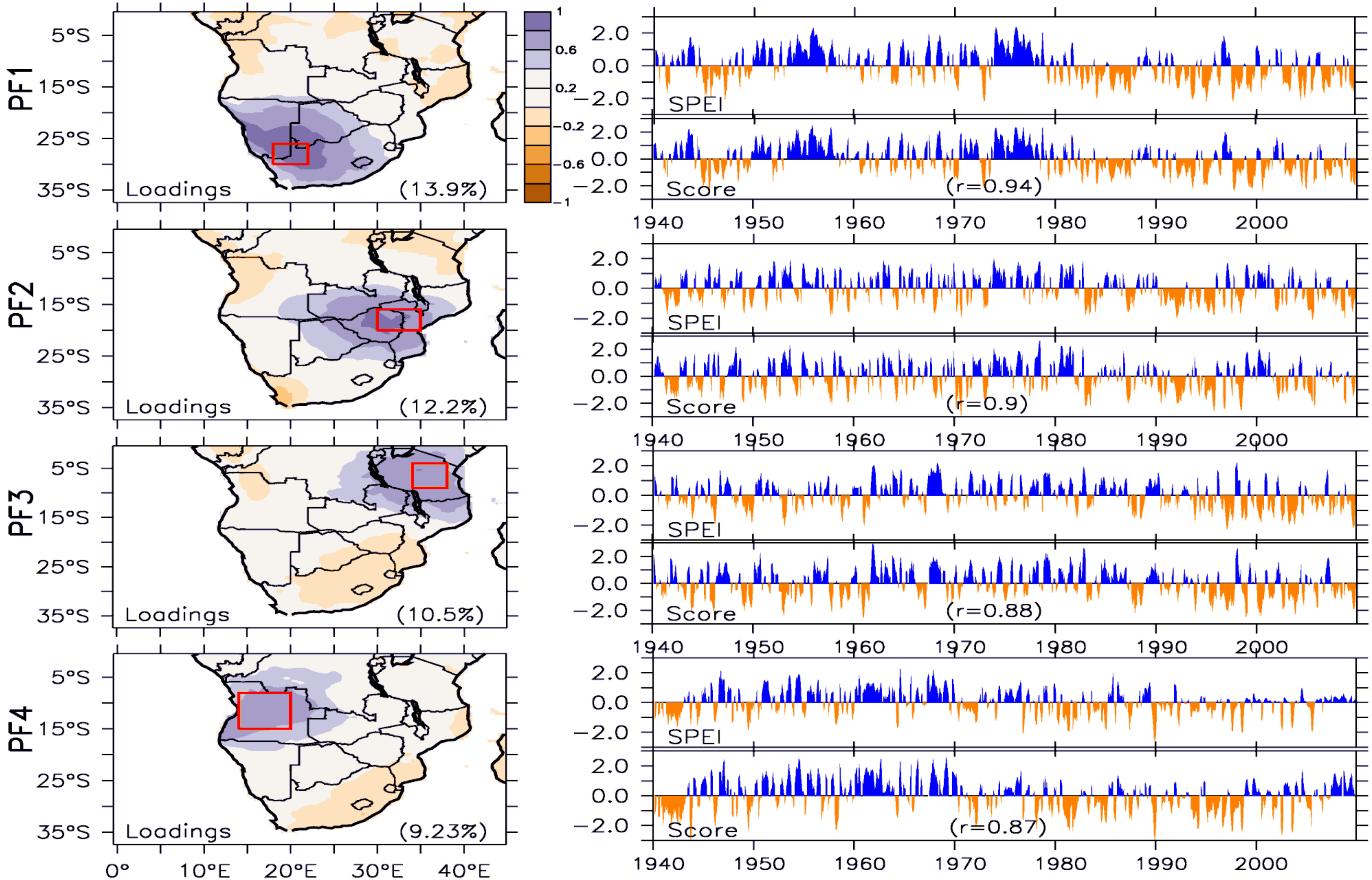
- PCA jointly explains 72% of variance in the data



# Re-union Island (Solar Radiation)



# Drought Modes




**Ujeneza and Abiodun (2015)**

# Practical
# Using Stastistical or other software

- Notes:

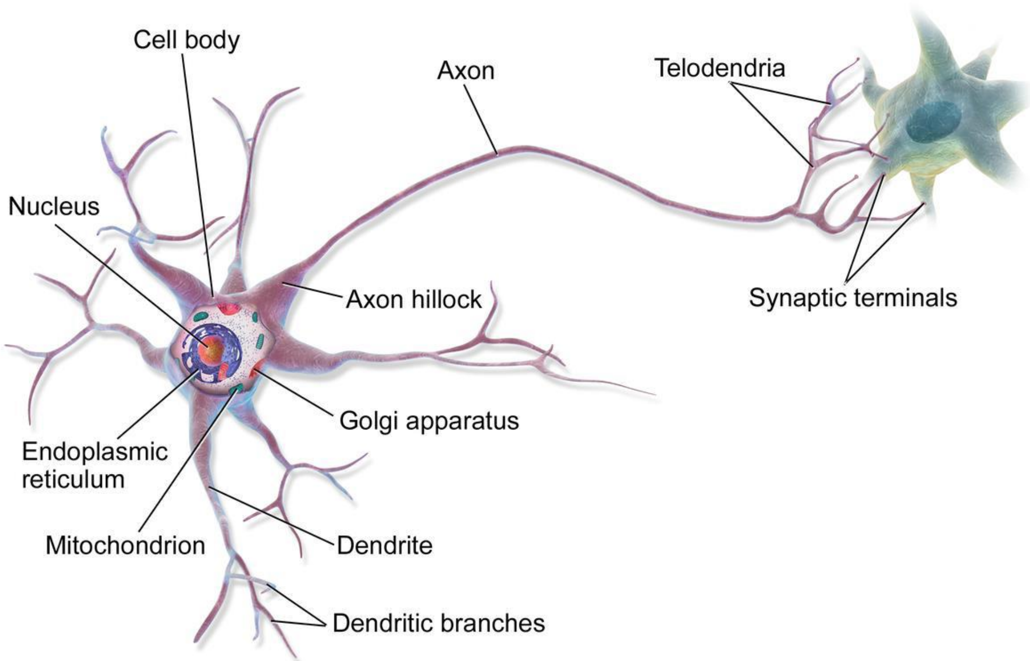
  - In your data matrix, you may cluster rows <u>or</u> columns -- this influences the dimensionality of your data space
  - Choice of algorithm is subjective and influences the result
  - Euclidian distances which assume orthogonality are generally used
  - Other more sophisticated non-linear techniques will be explored later.

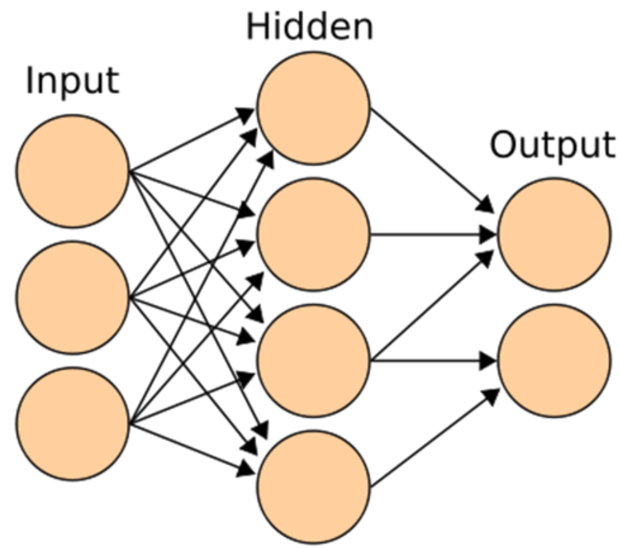
# Self-Organizing Maps

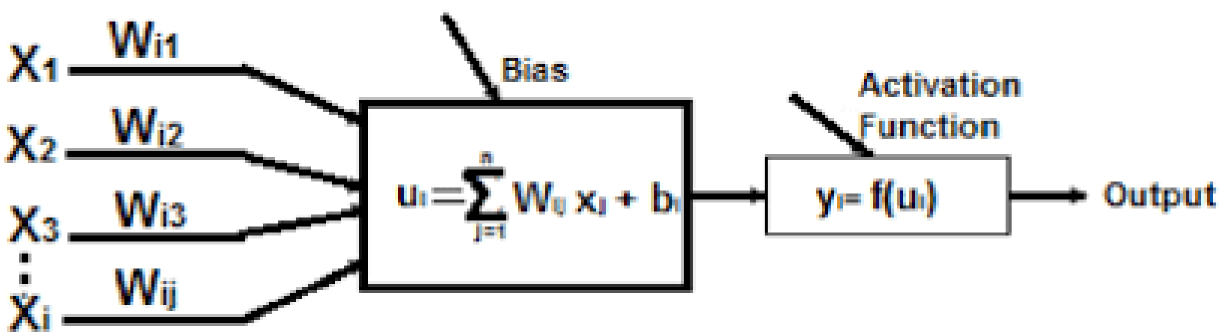
# What is a Self-Organizing Map?

- A Self-Organizing Maps (SOM) is a form of Artificial Neural Network(ANN)
- It uses unsupervised learning
- It used for data clustering, visualization and dimensionality reduction
- It takes a high dimensional data as input and reduces it to a low-dimensional (typically two-dimensional) input space of training samples called maps.

# What is an Artificial Neural Network



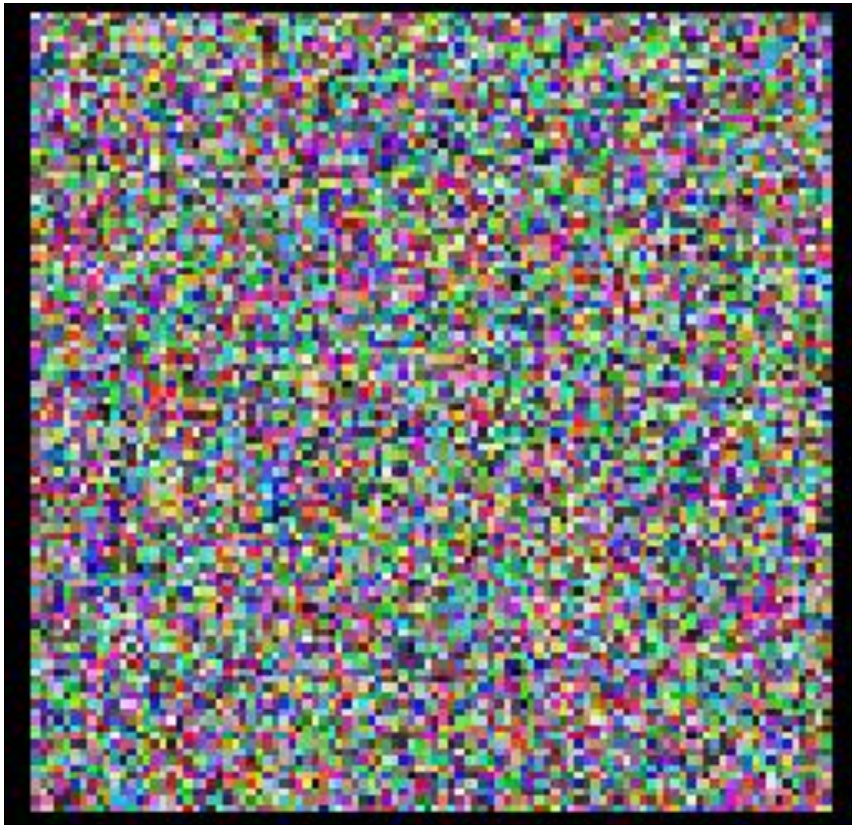



# Self-Organising Maps

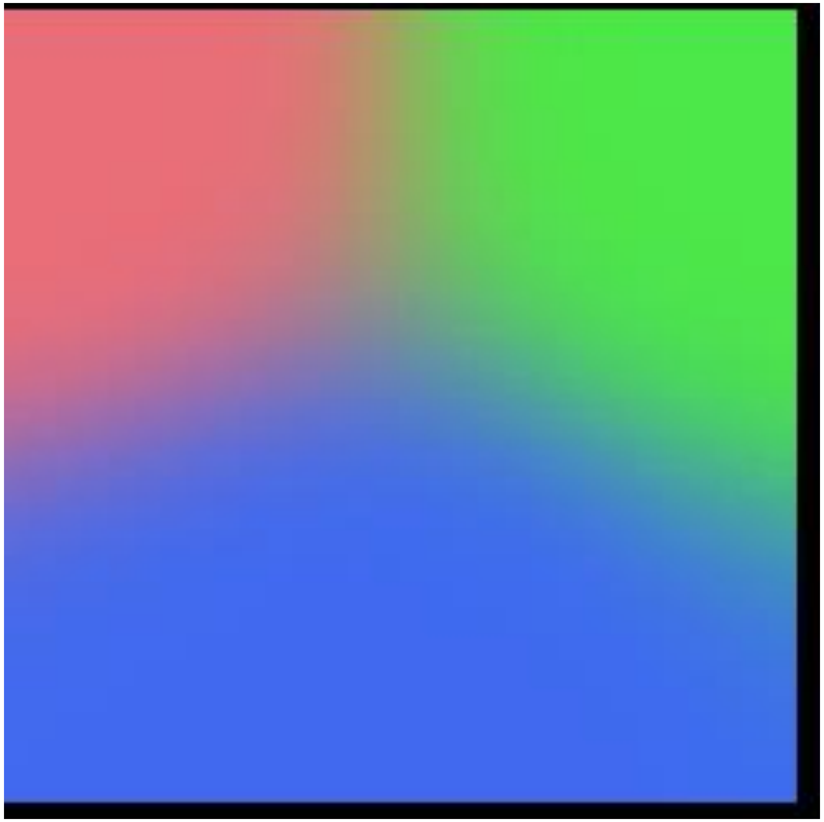
- Invented by Teuvo Kohonen.
- Unsupervised ANN using competitive learning.
- Provide mechanism for visualising complex relationships in multi-dimensional data sets.
- A tool used for clustering, visualisation and dimension reduction.

"Given an *N*-dimensional cloud of data points, the SOM will seek to place an arbitrary number of nodes within the data space such that the distribution of nodes is representative of the multi-dimensional distribution function, with the nodes being more closely spaced in regions of high data densities"

# SOM Classification of Colour points



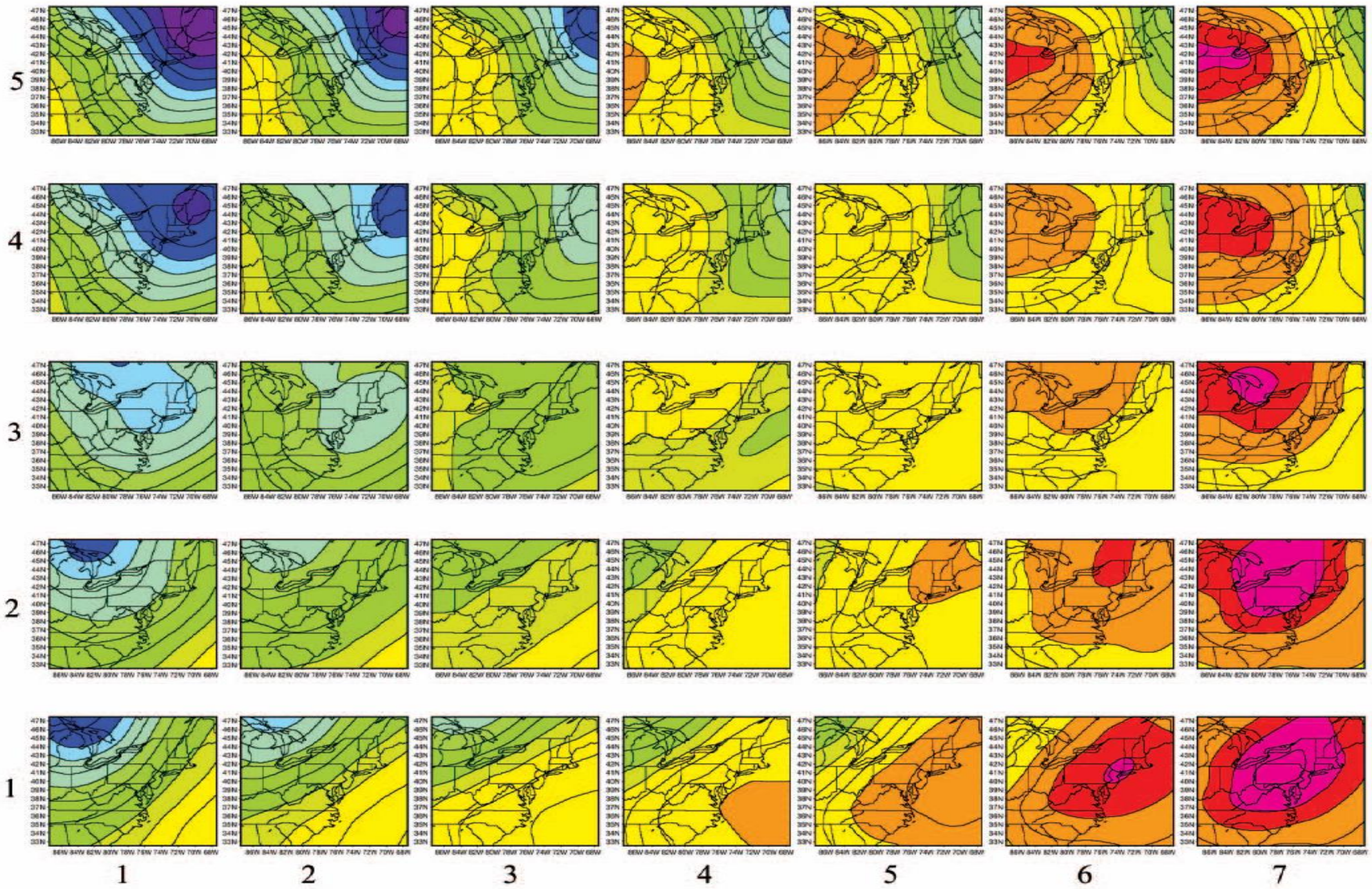
**Randomly initialized map**



**Map trained on shades of red, green, and blue**

# Example 1: 5x7 SOM array of Jan SLP for NE USA

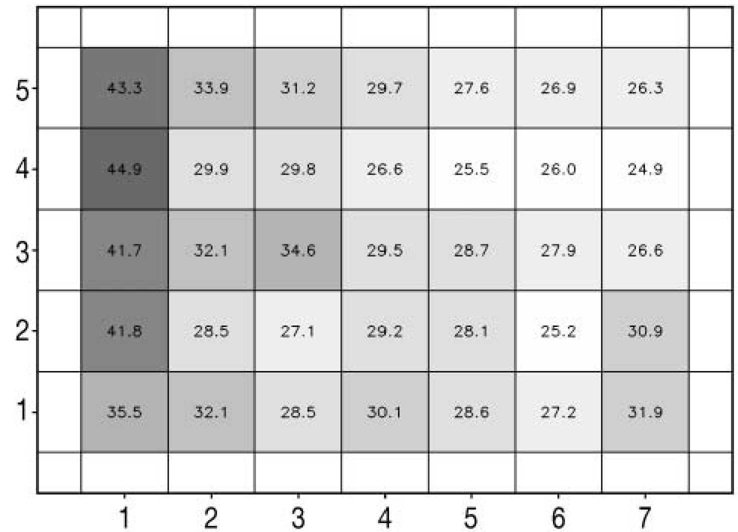
(Hewitson and Crane, 2002)



# Evaluation of the SOM Training

**Evaluation of the quantization error**

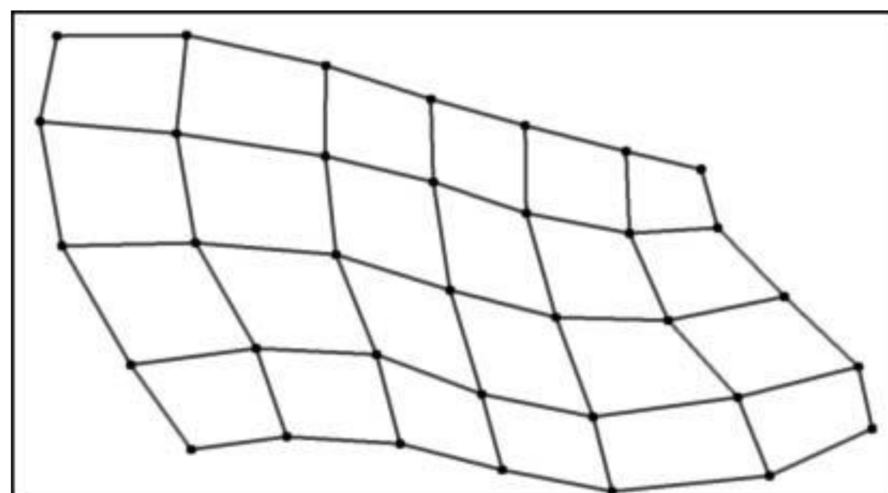
A list of the coordinates corresponding to the best-matched unit in the map for each data sample in the data file is produced. This list file also gives the individual quantization error for each sample.




(Hewitson and Crane, 2002)

**Sammon Map**

Generates the Sammon mapping from *n*-dimensional input vectors to 2-dimensional points on a plane whereby the distance between the image vectors tend to approximate the Euclidean distances of the input vectors.

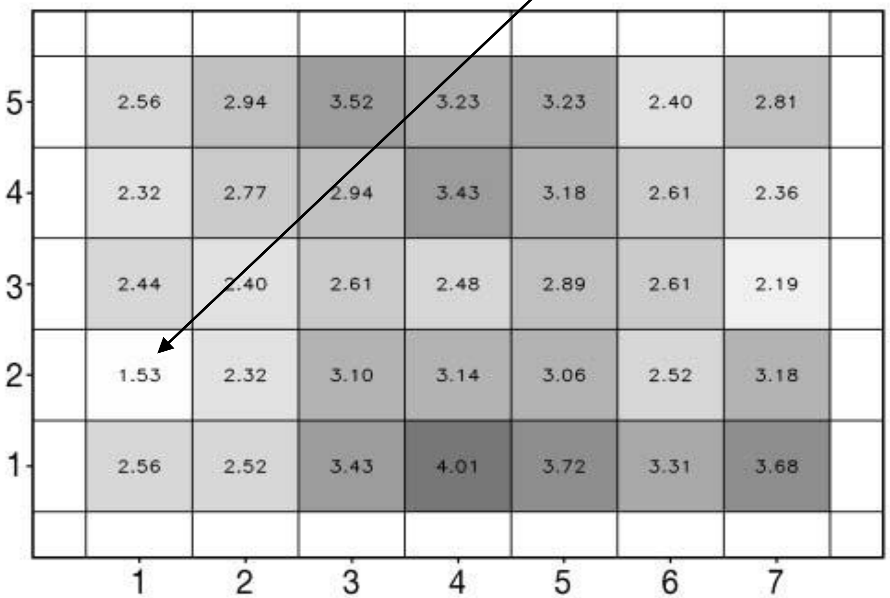


(Hewitson and Crane, 2002)

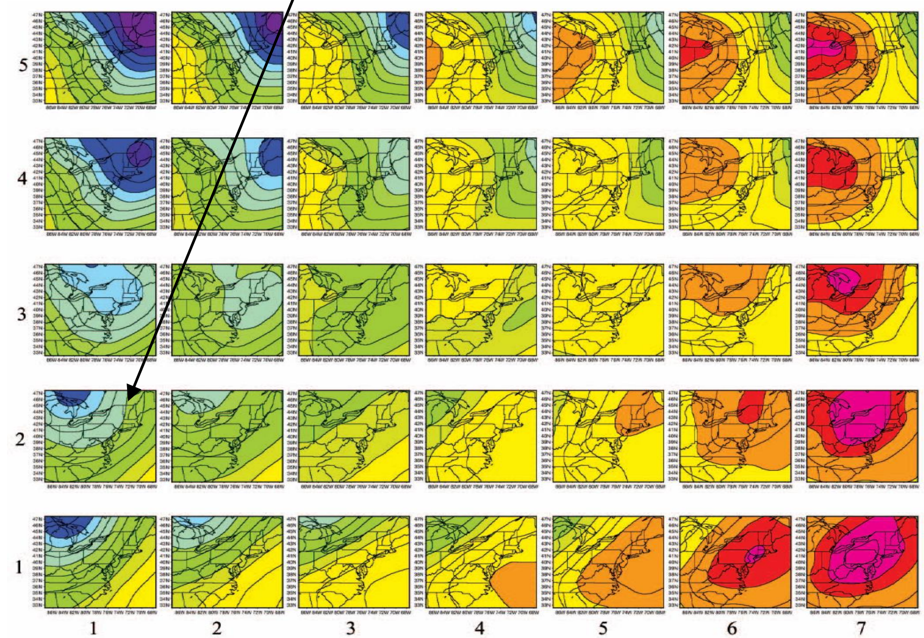
# Frequency of Occurrence

Climatological 40 yr mean frequency (%) of days in a month mapping to each SOM node.

eg. The frequency of Node 1:2 is 1.56 % and corresponds to this SLP pattern.




(Hewitson and Crane, 2002)




(Hewitson and Crane, 2002)

# What can a SOM be used for?

- Feature extraction
- Clustering
- Pattern recognition
- Dimensionality reduction
- Visualization of complex structures that would otherwise be hardly recognizable by humans.
- Supervised data classification ()
- Data interpolation()

# Advantages of SOM

- Attempt to maintain the probability density function
- Allow for a non-linear mapping
- Allow easy visualization and analysis
- Objective, "hands-free" classification process

# How does a SOM works?

- SOM building blocks,
  - a node set data structure representing the actual map content
  - Algorithms applied to that node set.

- Present the SOM network with dataset and network uses unsupervised learning to train the network
- Algorithm involves four phases:
  - Initialization
  - Competition
  - Adaptation
  - Cooperation

# SOM Algorithm

- STEP 1: Initialize the SOM (Using random values or values resulting from analysis of data)
- STEP 2: For each training vector X do steps 3 - 7
- STEP 3: For each SOM node j calculate the distance D(j) = SUM(Wj-X)
- STEP 4: Select the node J that has the lowest D(J) ie. the winning node
- STEP 5: For each node j within a topological radius R, update the weights according to Wj(new) = Wj(old) + a[X-Wj]
- STEP 6: Update learning rate a and learning radius R
- STEP 7: Test for stop condition

# Characteristics of SOM

- SOM is “topology preserving” in the sense that (as far as possible) neighborhoods are preserved through the mapping process.
- Generally, no matter how much we train the network, there will always be some difference between any given input pattern and the unit it is mapped to. This difference is called quantization error, and is used as a measure of how well map units represent the input patterns.

# Performance of the SOM algorithm

- Depends on two important parameters

1. Learning Rate
    - Determines how fast the weights move towards the data points
    - Needs to start off quite high so that the map moves rapidly towards the data distribution.
    - It must reduce to almost zero to allow the map to stabilize and settle on a final solution.
2. Learning Radius
    - Determines how many nodes surrounding the winning node are updated
    - This parameter should start quite high to ensure that the topology of the map is maintained strongly while it aligns itself with the data.
    - It is then slowly reduced to allow individual nodes to achieve a final stable position.

# Expected output

- Trained Map of SOM (Output node set)
- Node assignment statistics
  - Frequency of occurrence
  - Input data assignment ( which input sample was assigned to which node)

# SOM analysis interpretation

- Dominant pattern
- Causality

# Quality assurance of a SOM analysis

- Summon map
- Quantization error
- Visualization
- Frequency of Occurrence

# Conclusion

- A SOM is a powerful tool for data exploration, feature extraction, pattern recognition, clustering and dimensionality reduction.
- The special principle that SOMs aim to imitate is the mapping of high-dimensional input data to a low-dimensional output network in a topology preserving way (Christoph Brauer)

# Useful resources on SOM

- http://davis.wpi.edu/~matt/courses/soms/

# Climate Change and Regionally-Extensive Droughts in Southern Africa

**Babatunde J. Abiodun**

Department of Environmental and Geographical Science

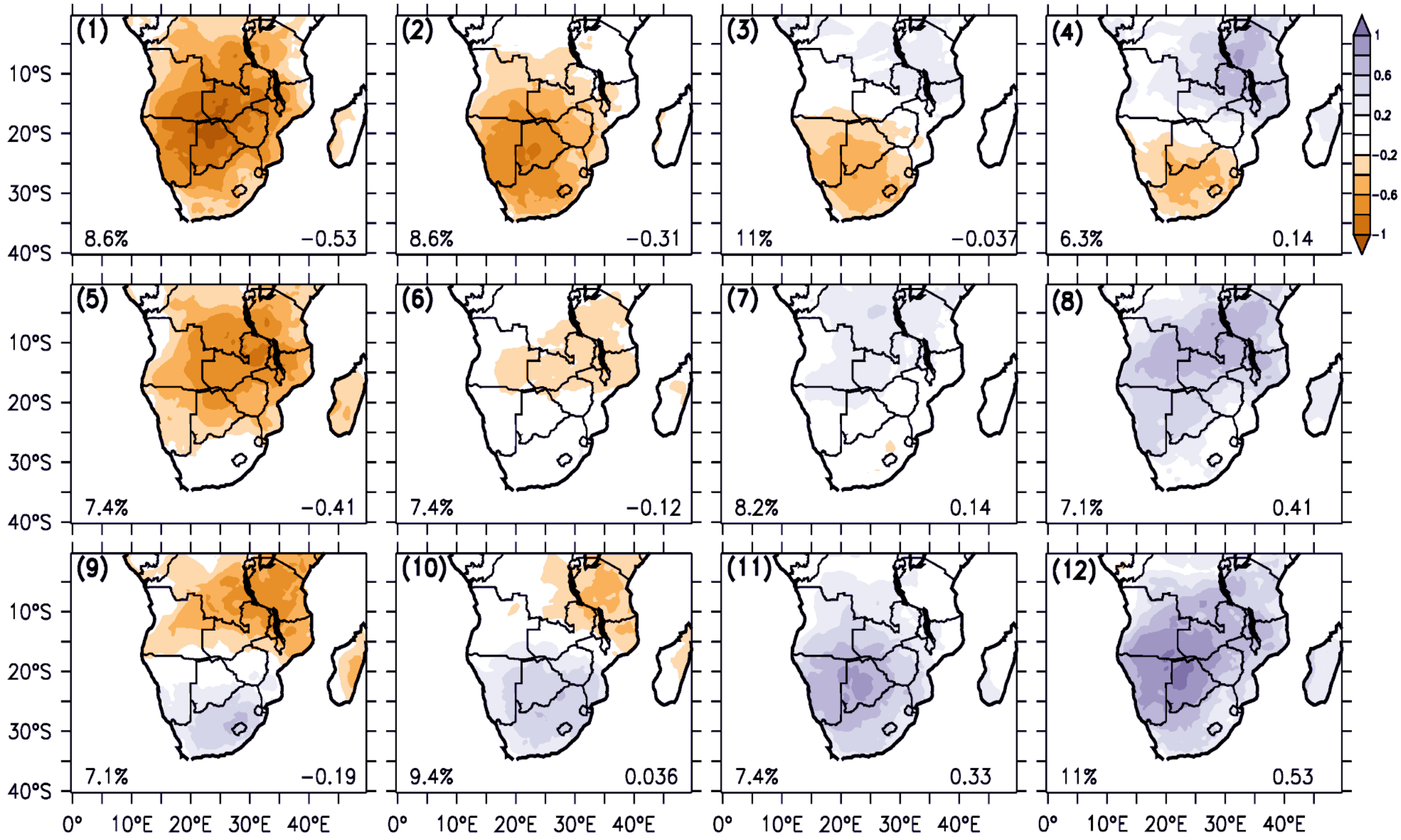
University of Cape Town

WATER RESEARCH COMMISSION

University of Cape Town • iYunivesithi yaseKapa • Universiteit van Kaapstad

SPES BONA

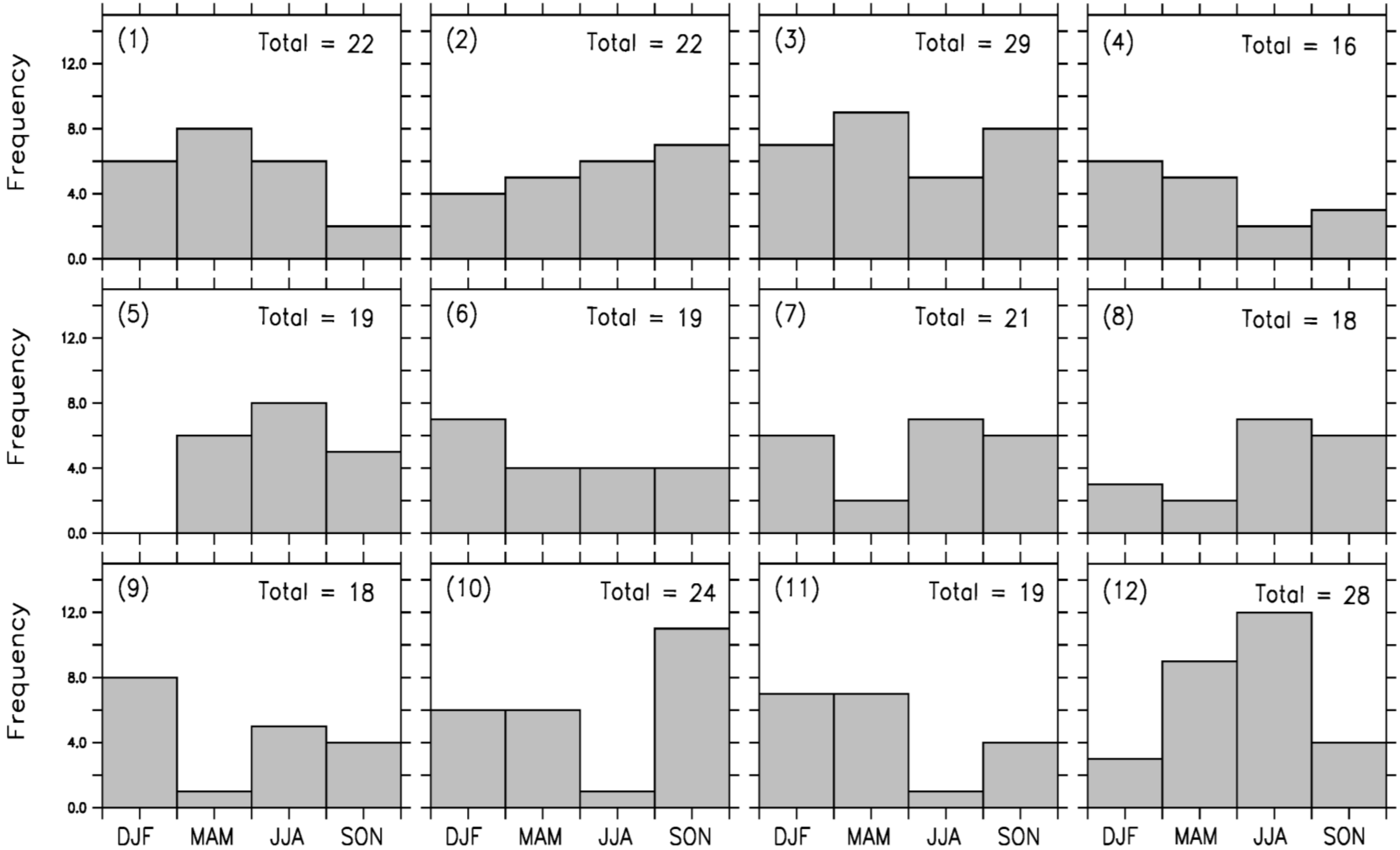
# Drought patterns over Southern Africa



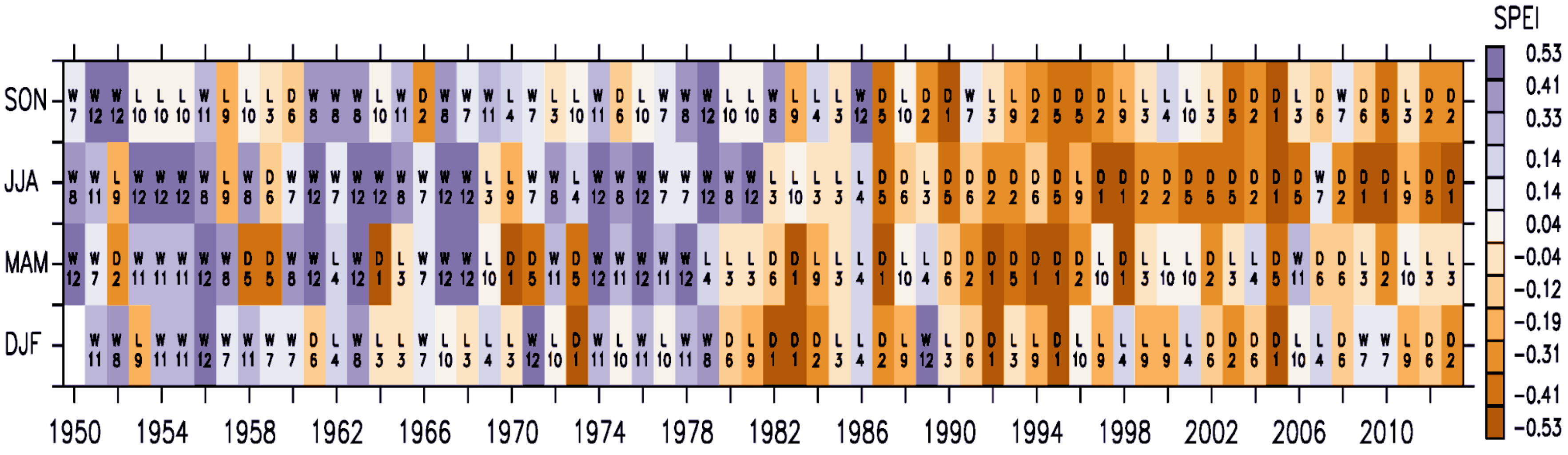

WATER RESEARCH COMMISSION

University of Cape Town · iYunivesithi yaseKapa · Universiteit van Kaapstad

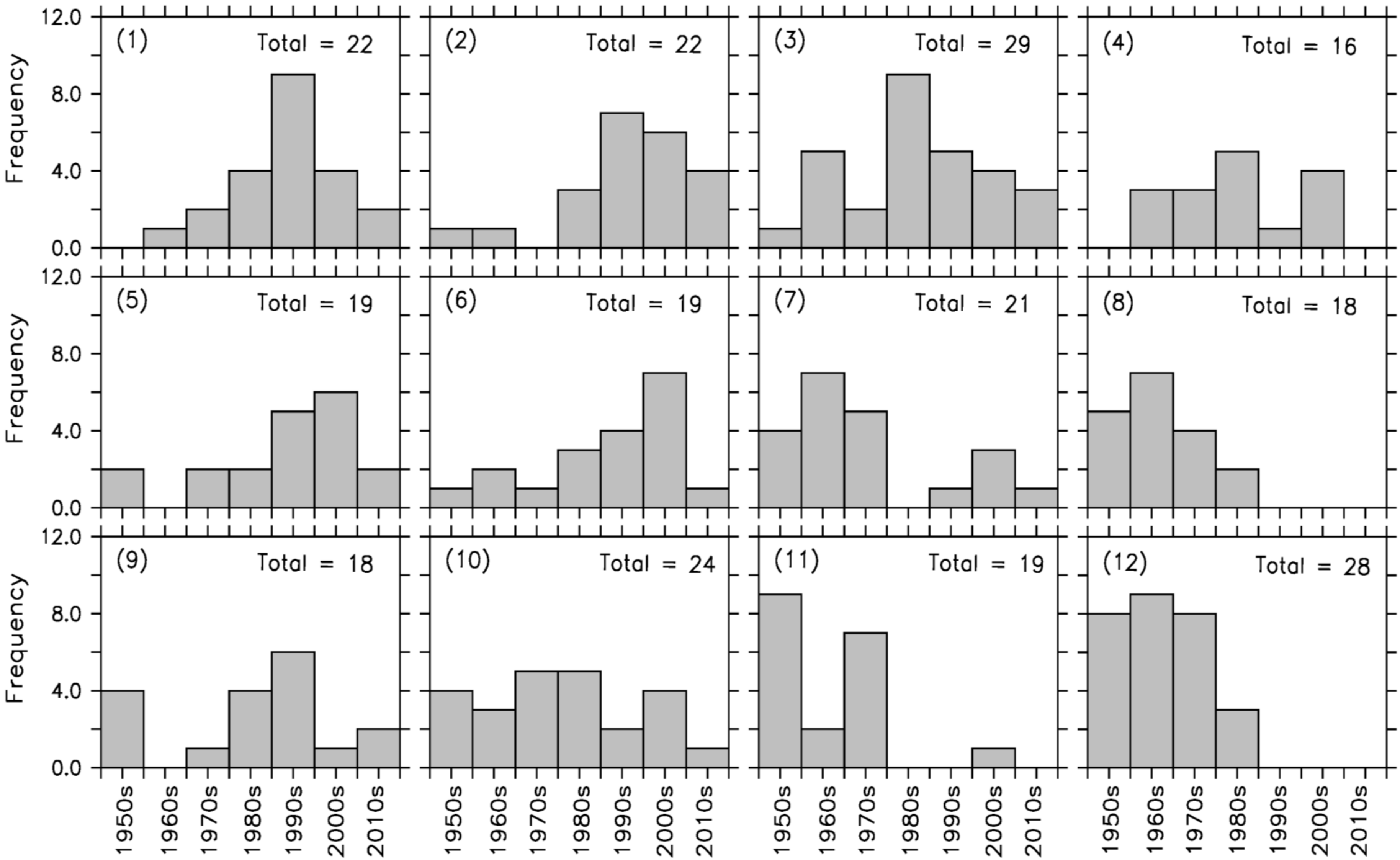
# Seasonal variation of the drought patterns



# Transition and Persistence of the Drought Patterns



# Decadal variation of the drought patterns



# Seminar Topics

- The usage of principal component analysis in climate science [Rebecca Efua Guisgo]

- Application of Artificial Neural Network methods in climate science [Grace Adu Aborah]

- A literature review on bias correction methods in climate science[Akeresola Rebecca]

# Time Series Analysis

# Time Series Analysis

- Time series data:
  - sequence of measurements that follow non-random orders.
  - measurements are at equally spaced time intervals.
- Goals of Time series analysis:
  - Identifying of patterns
  - Predicting future values

# Identifying patterns…

- Systematic pattern and random noise
  - Filtering …
- Trend and Seasonality (periodicity)

# Trend Analysis

- Smoothing
  - local averaging of data
- Curve fitting
  - Approximate data by a linear (non-linear) function

# Analysis of seasonality

- Autocorrelation
  - Autocorrelation
  - Cross - correlation

- Fourier Analysis

# Autocorrelation

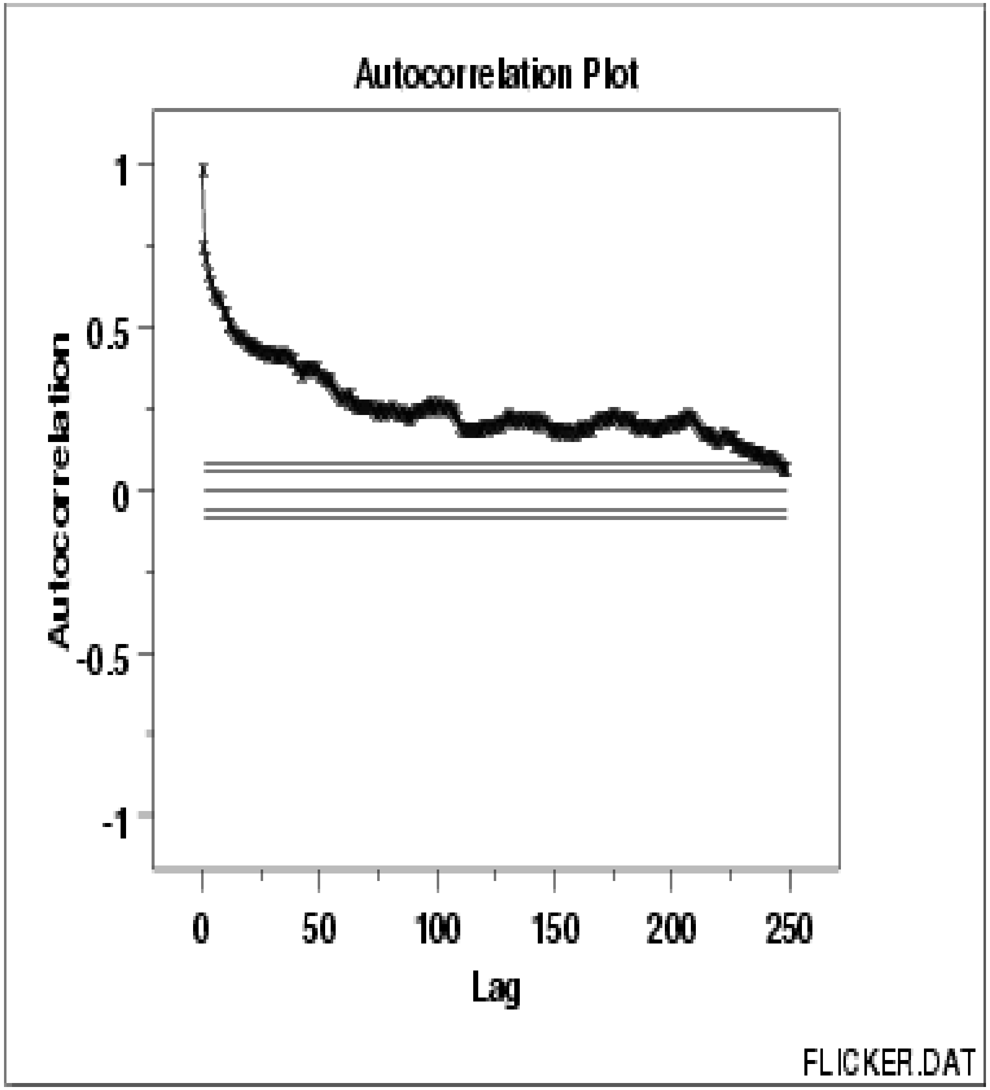
- When the correlation is calculated between a series and a lagged version of itself it is called autocorrelation.
- A high correlation is likely to indicate a periodicity in the signal of the corresponding time duration.

# Autocorrelation (auto-correlogram)

$$R_h = C_h / C_0$$

$$C_0 = \frac{\sum_{t=1}^{N} (Y_t - \bar{Y})^2}{N}$$

$$C_h = \frac{1}{N} \sum_{t=1}^{N-h} (Y_t - \bar{Y})(Y_{t+h} - \bar{Y})$$

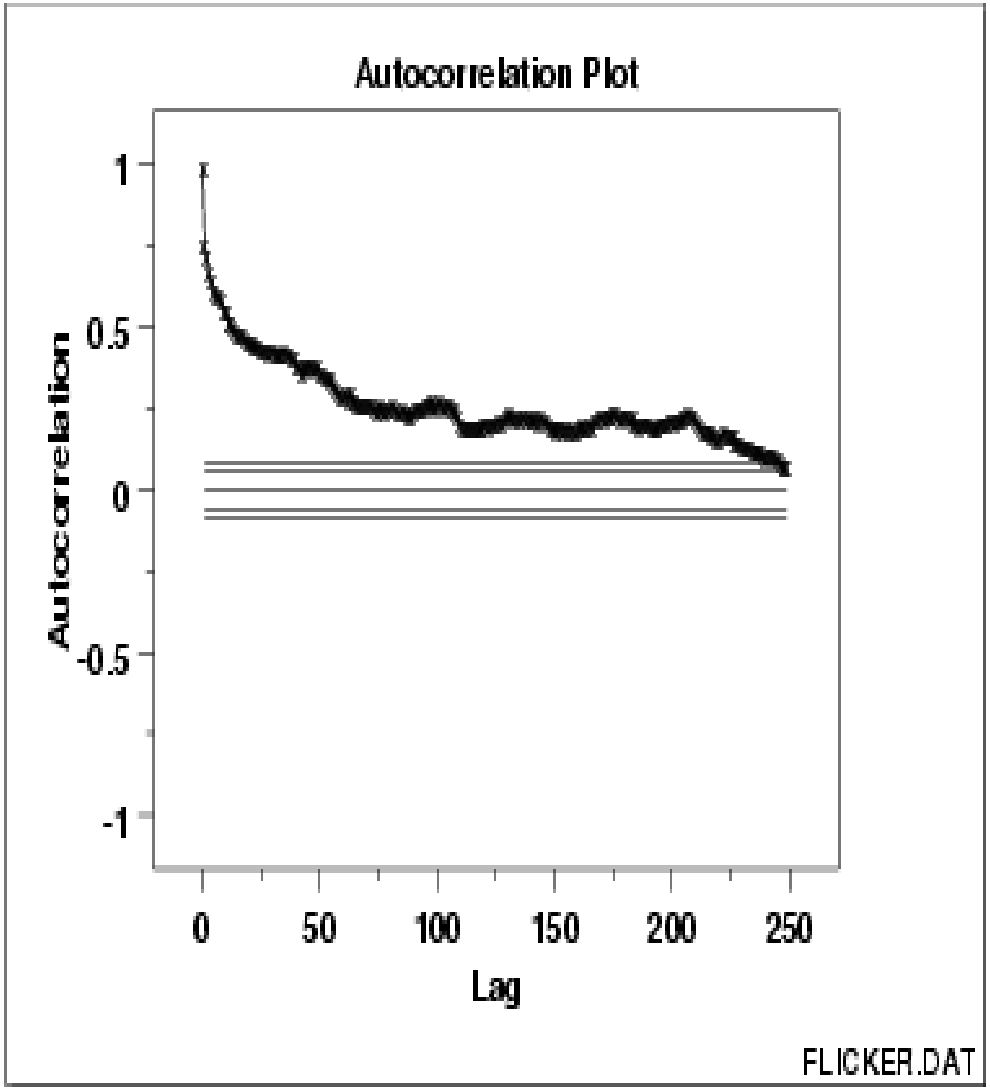


# Autocorrelation

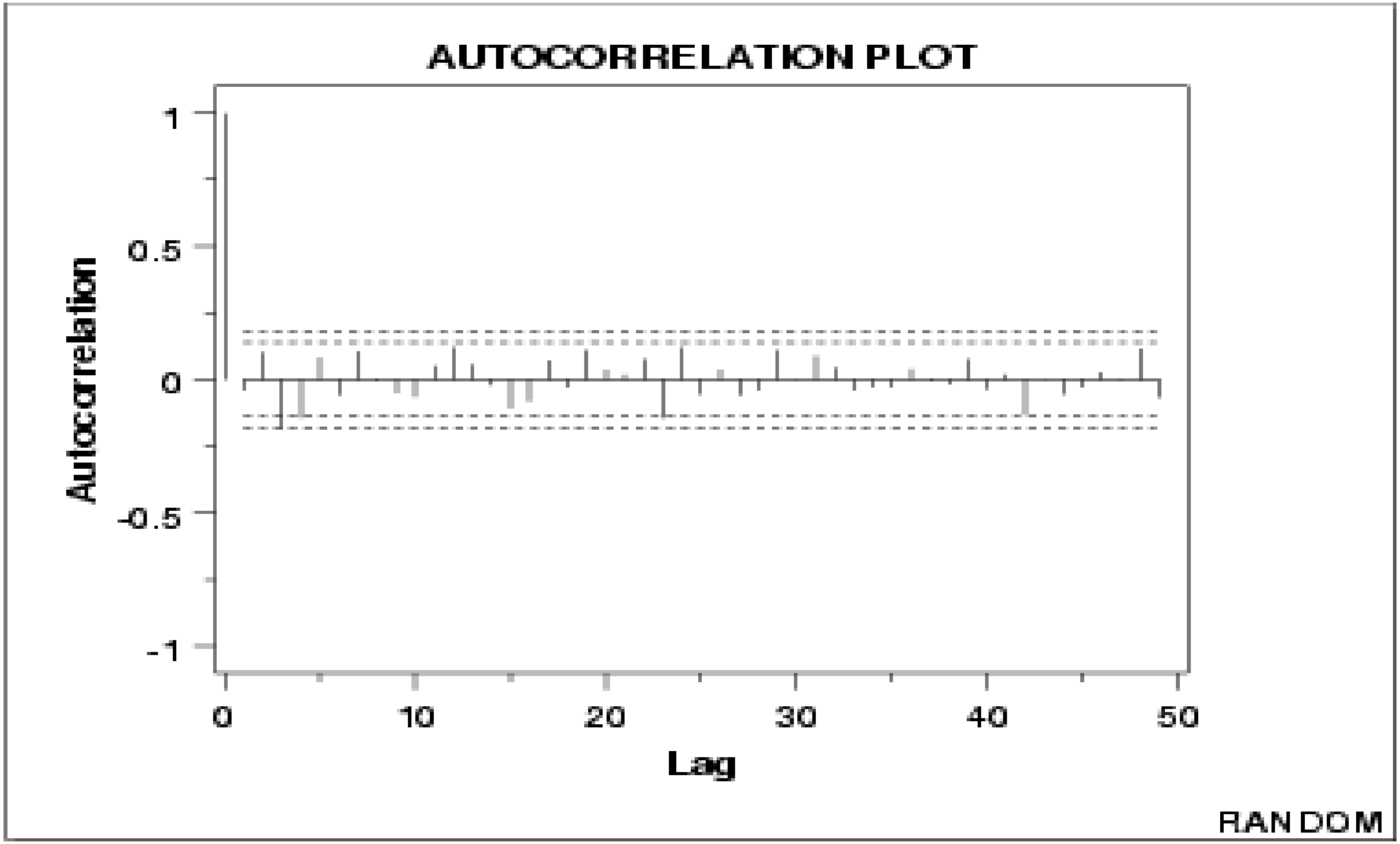
$$R_h = C_h / C_0$$

$$C_0 = \frac{\sum_{t=1}^{N} (Y_t - \bar{Y})^2}{N}$$

$$C_h = \frac{1}{N} \sum_{t=1}^{N-h} (Y_t - \bar{Y})(Y_{t+h} - \bar{Y})$$

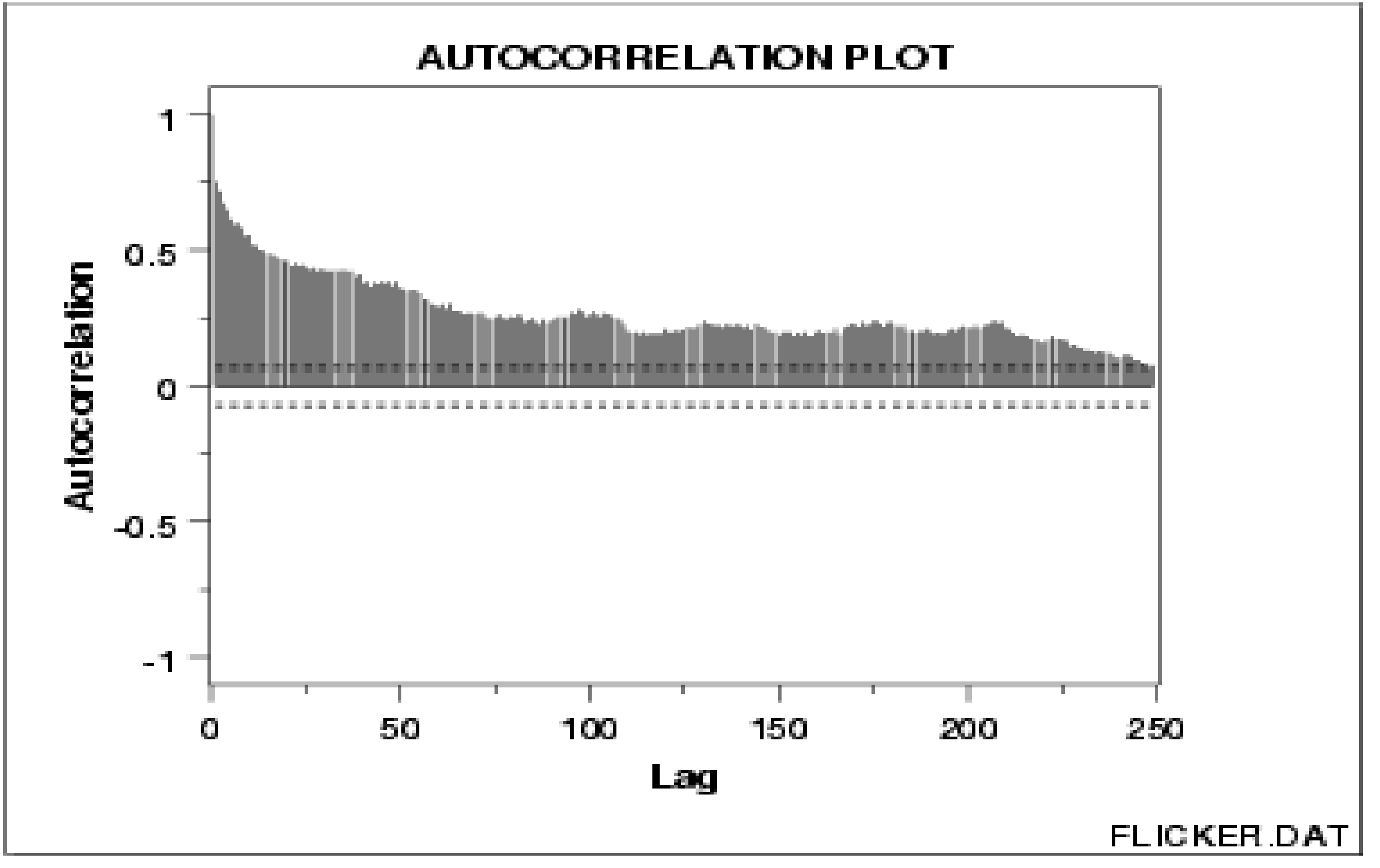


# Autocorrelation Plot: Random (white noise)

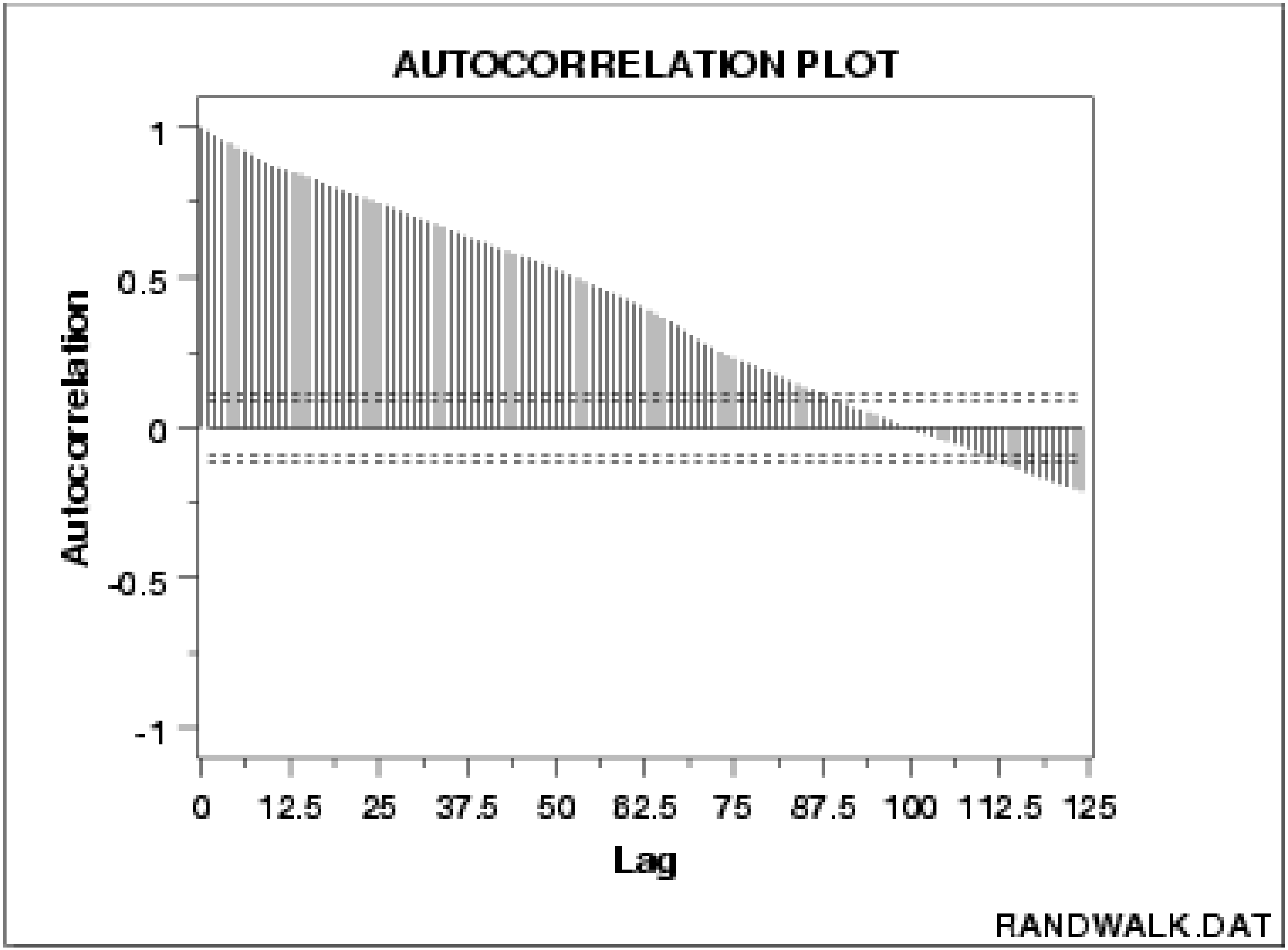



- There are no significant autocorrelations.
- The data are random.

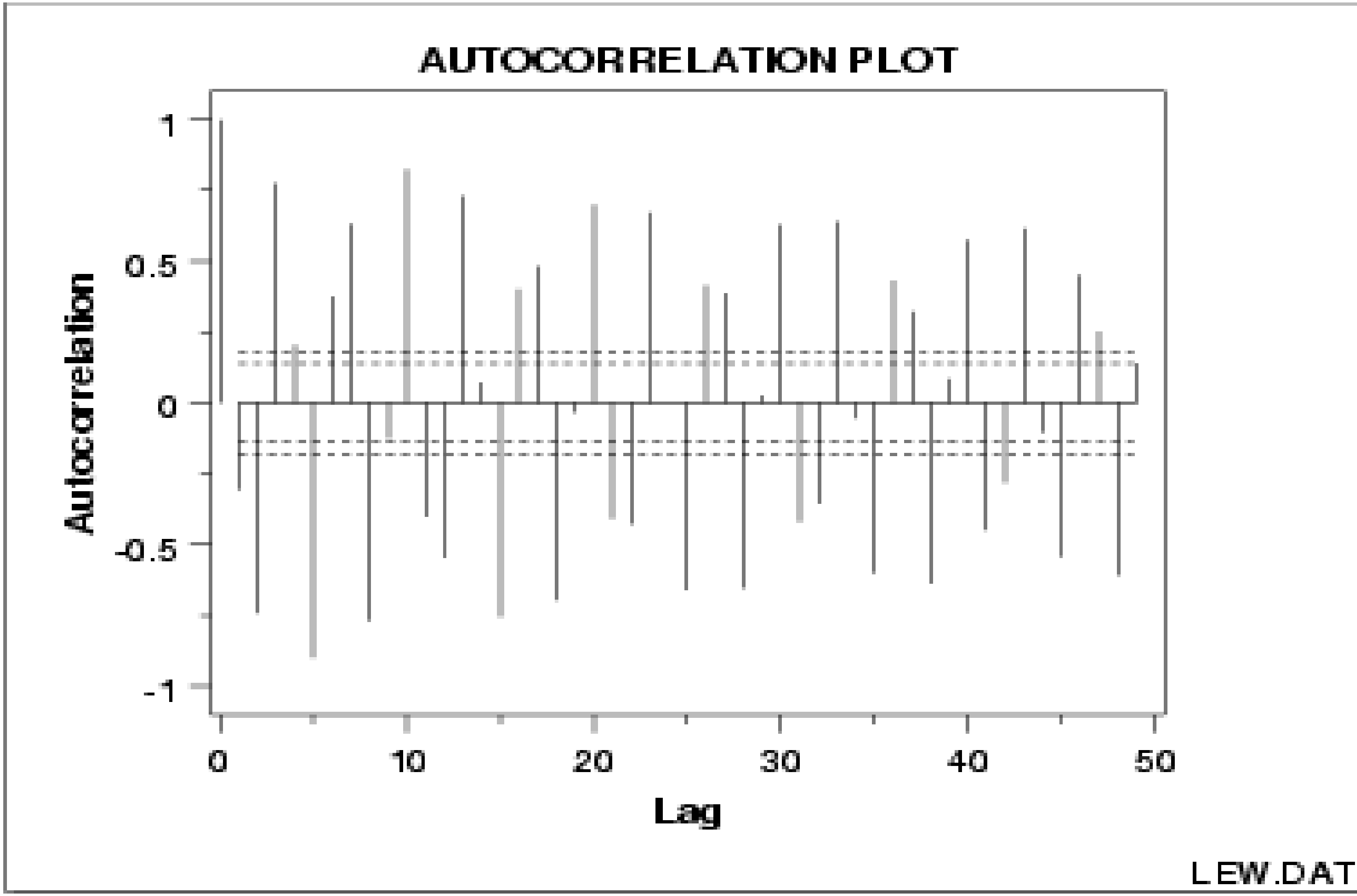
# Autocorrelation Plot: Moderate Autocorrelation



# Autocorrelation Plot: Strong Autocorrelation



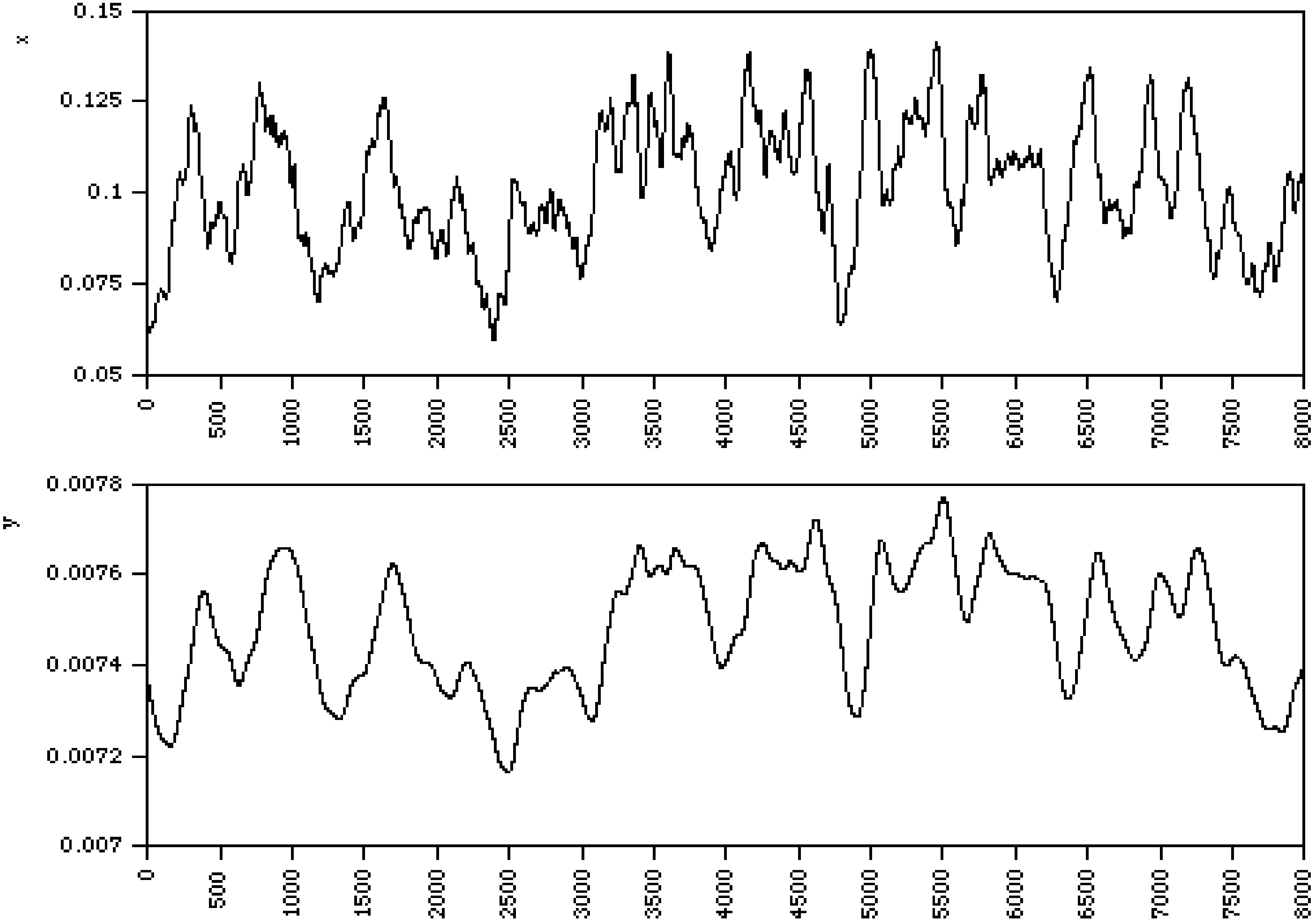
# Autocorrelation Plot: Sinusoidal



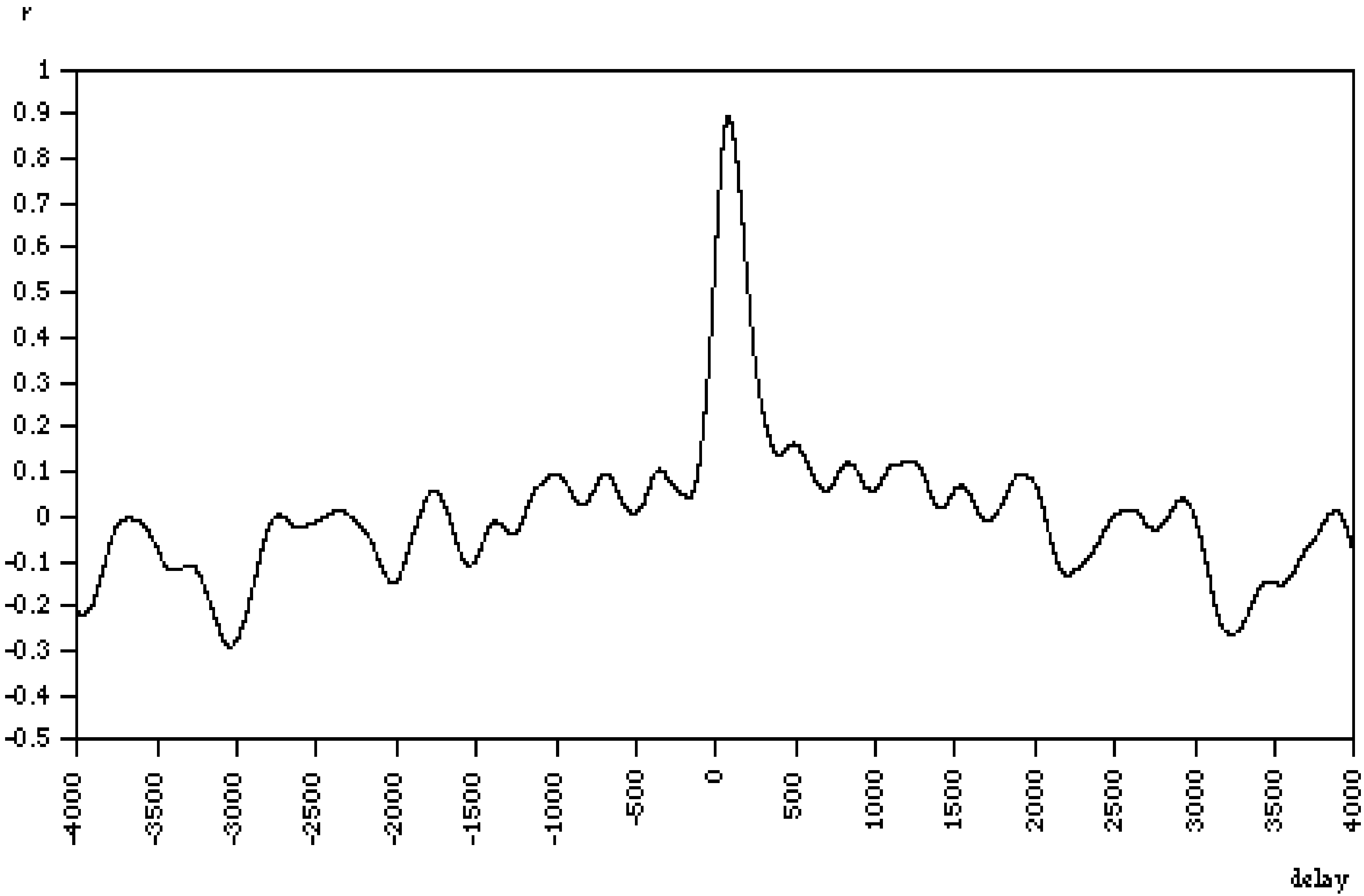
# Cross-correlation

- Cross correlation is a standard method of estimating the degree to which two series are correlated.

# Cross-correlation: x and y



# Cross-correlation Plot: r



# Wavelet Analysis

# Past Climate Change: The last 1000 Years

- Earth has always experienced climate variability on time scales ranging from seasonal-to-interannual to those occurring over hundreds to thousands of years.
- Notice that for most of the last 1000 years, the climate has been cooler than normal.
- The period from about 1550 - 1850 is referred to as the **Little Ice Age.**
- Notice a dramatic increase in temperature since 1860

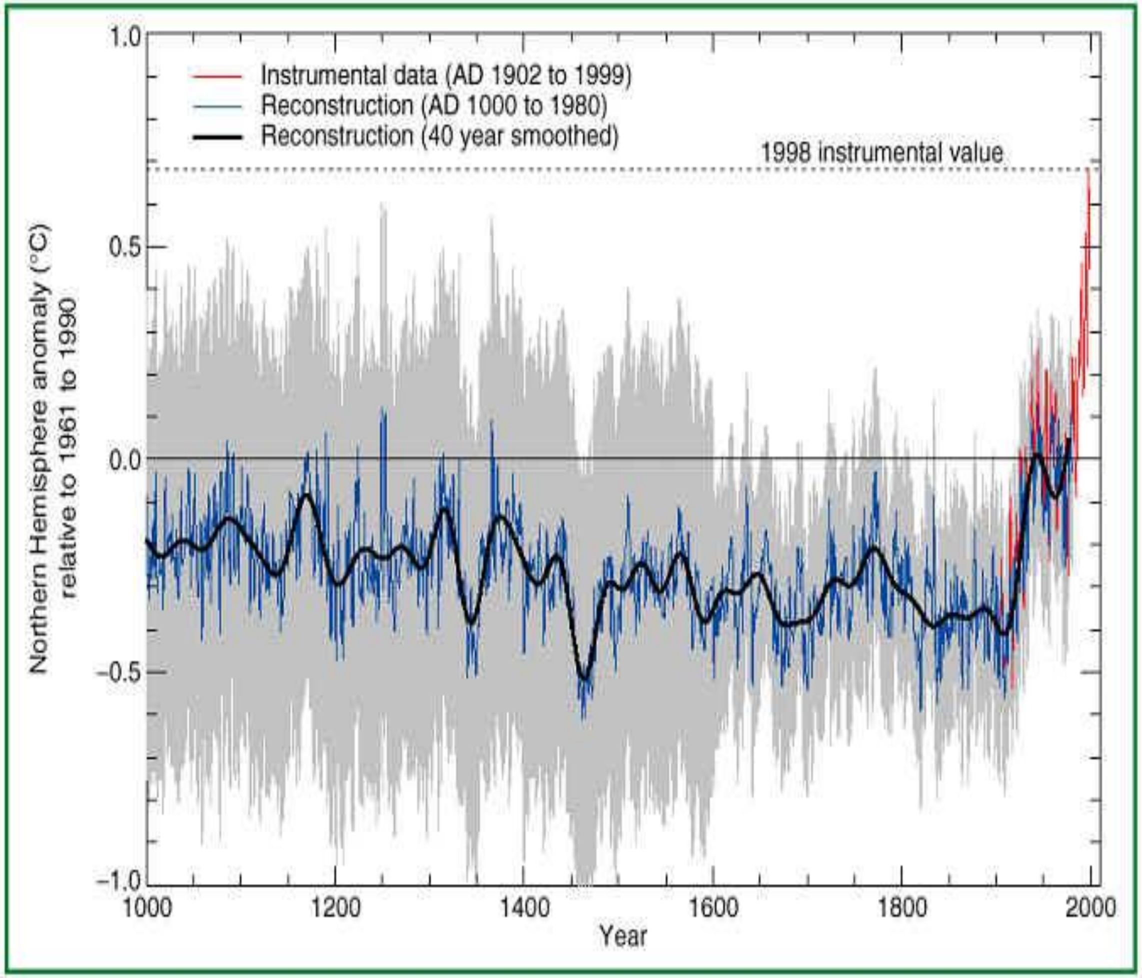
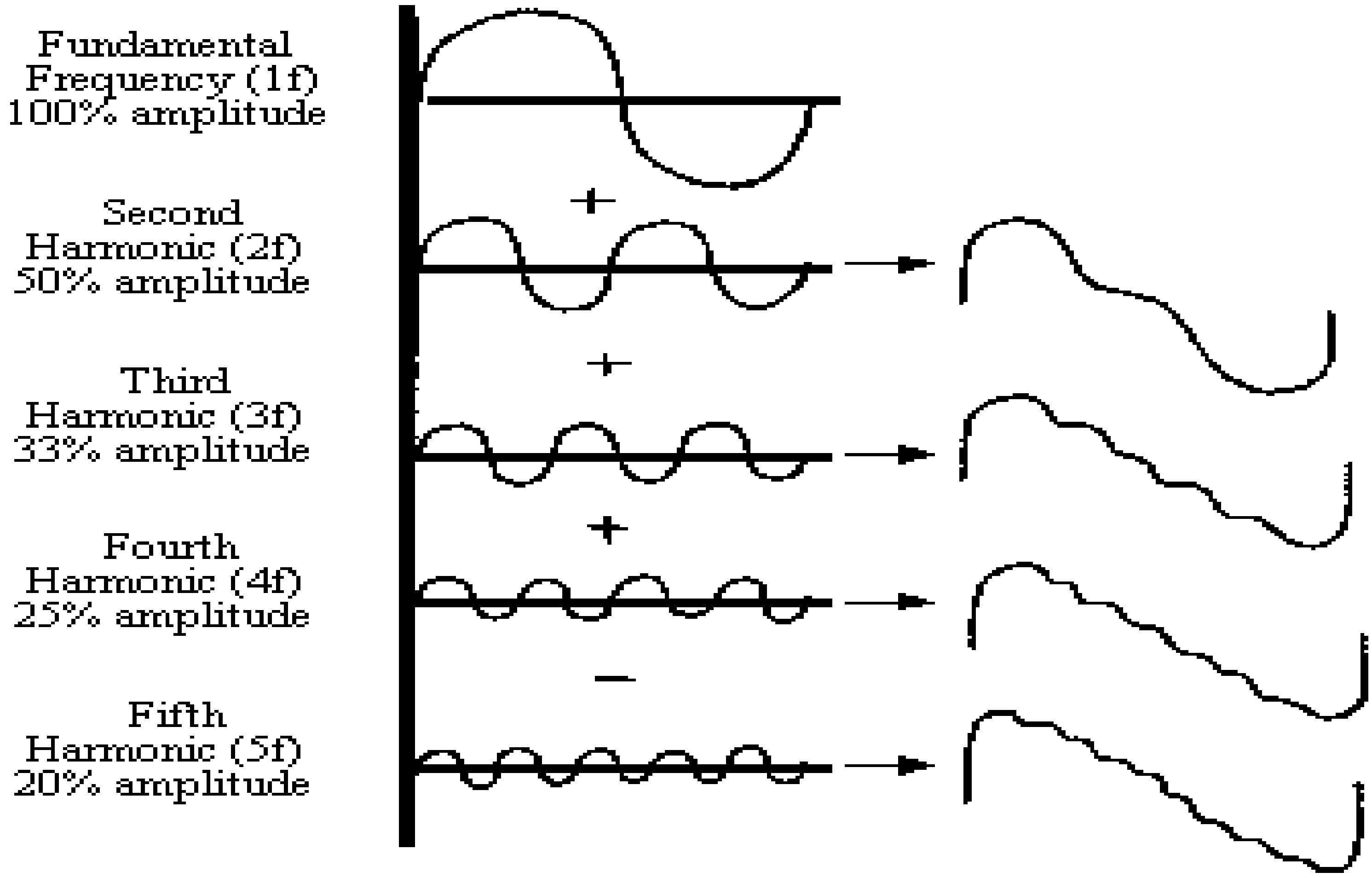



**Figure 5**: Millennial Northern Hemisphere (NH) temperature reconstruction (blue – tree rings, corals, ice cores, and historical records) and instrumental data (red) from AD 1000 to 1999. Smoother version of NH series (black), and two standard error limits (gray shaded) are shown. [Based on Figure 2.20]

# Trend and Variability




The Making of a Complex Waveform

# Fourier Analysis

- Also known as spectral analysis
- Use the exploration of cyclical patterns of data
- Purpose:
  - to decompose a complex time series with cyclical components into a few underlying sinusoidal (sine and cosine) functions of particular wavelengths
  - to identify the seasonal fluctuations of different lengths
- More efficient than autocorrelation
- Uses variance (not correlation)

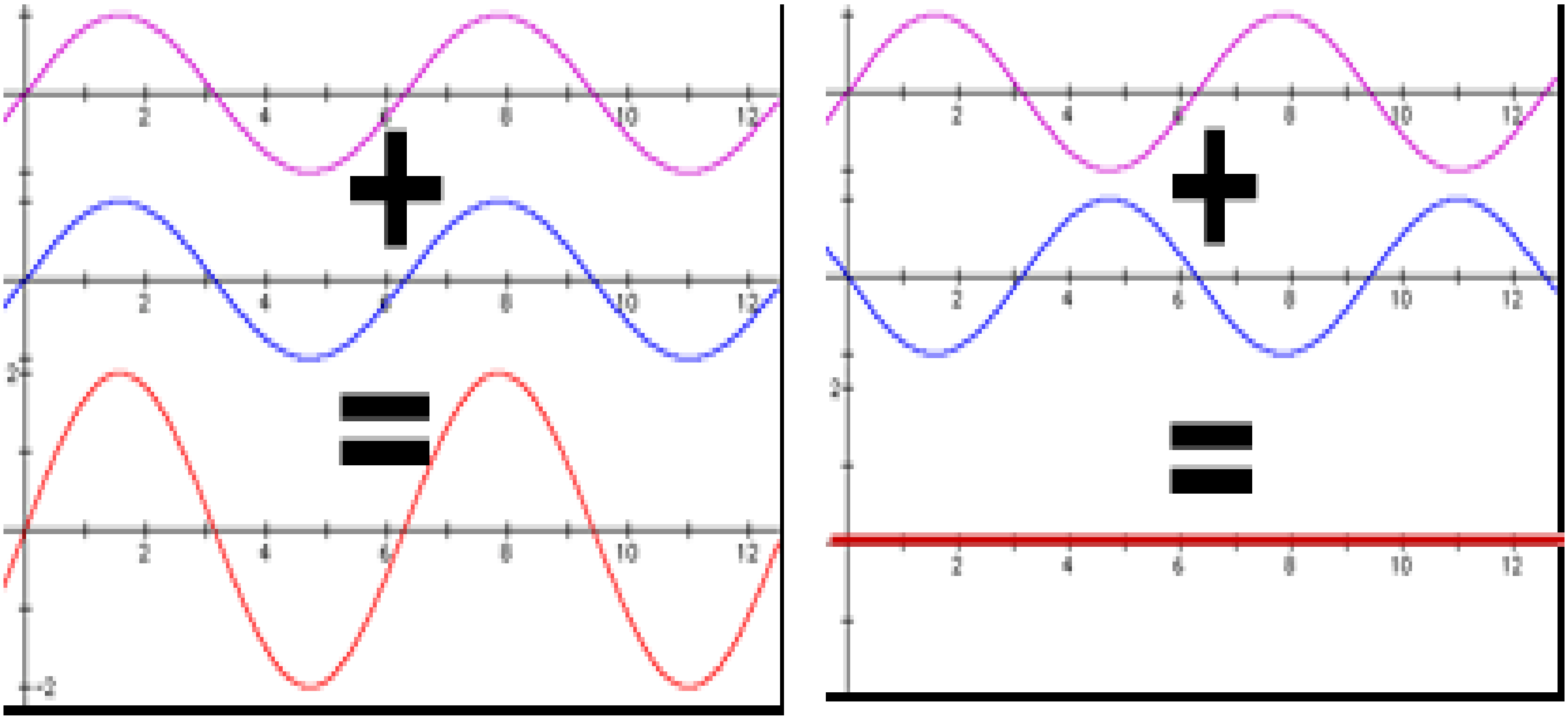
# Basic Notations

- Wave length
- Amplitude
- Frequency (f)
- Period  (1/f)
- Harmonics (n*f)
- Nyquist frequency (highest frequency)
- Lowest frequency determined by n/2

# Principle

- **Equation**
  - $x_t = a_0 + \sum [a_k * \cos(l_k * t) + b_k * \sin(l_k * t)]$ (for k = 1 to q)
- **Periodogram / Power spectrum**
  - $P_k = (a_k^2 + b_k^2) * N/2$
- **Continuous spectrum / spectral density**

# Fourier Analysis Example (1)



# Fourier Analysis Example (2)

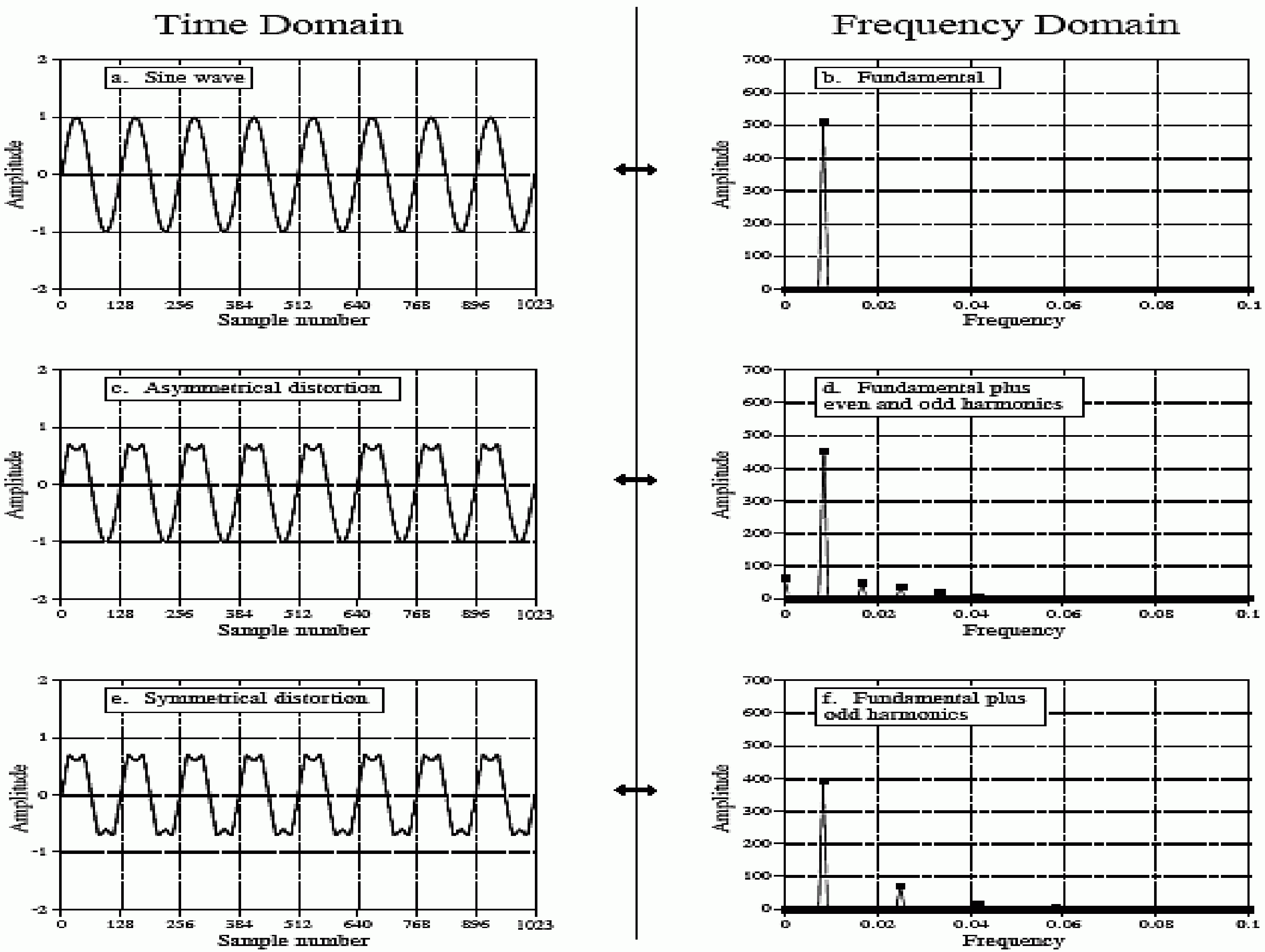
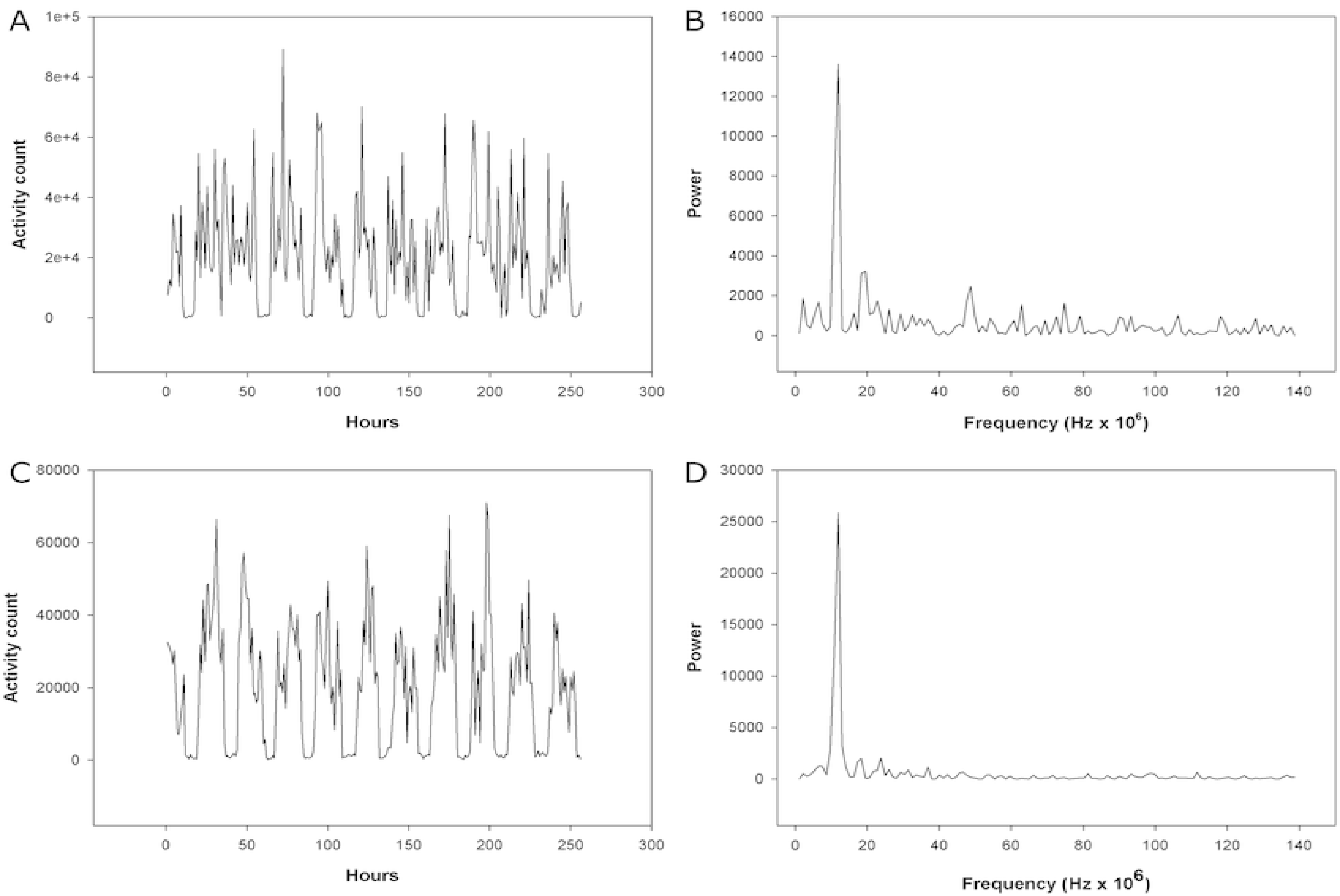



FIGURE 11-7
Example of harmonics. Asymmetrical distortion, shown in (c), results in even and odd harmonics, (d), while symmetrical distortion, shown in (e), produces only odd harmonics, (f).

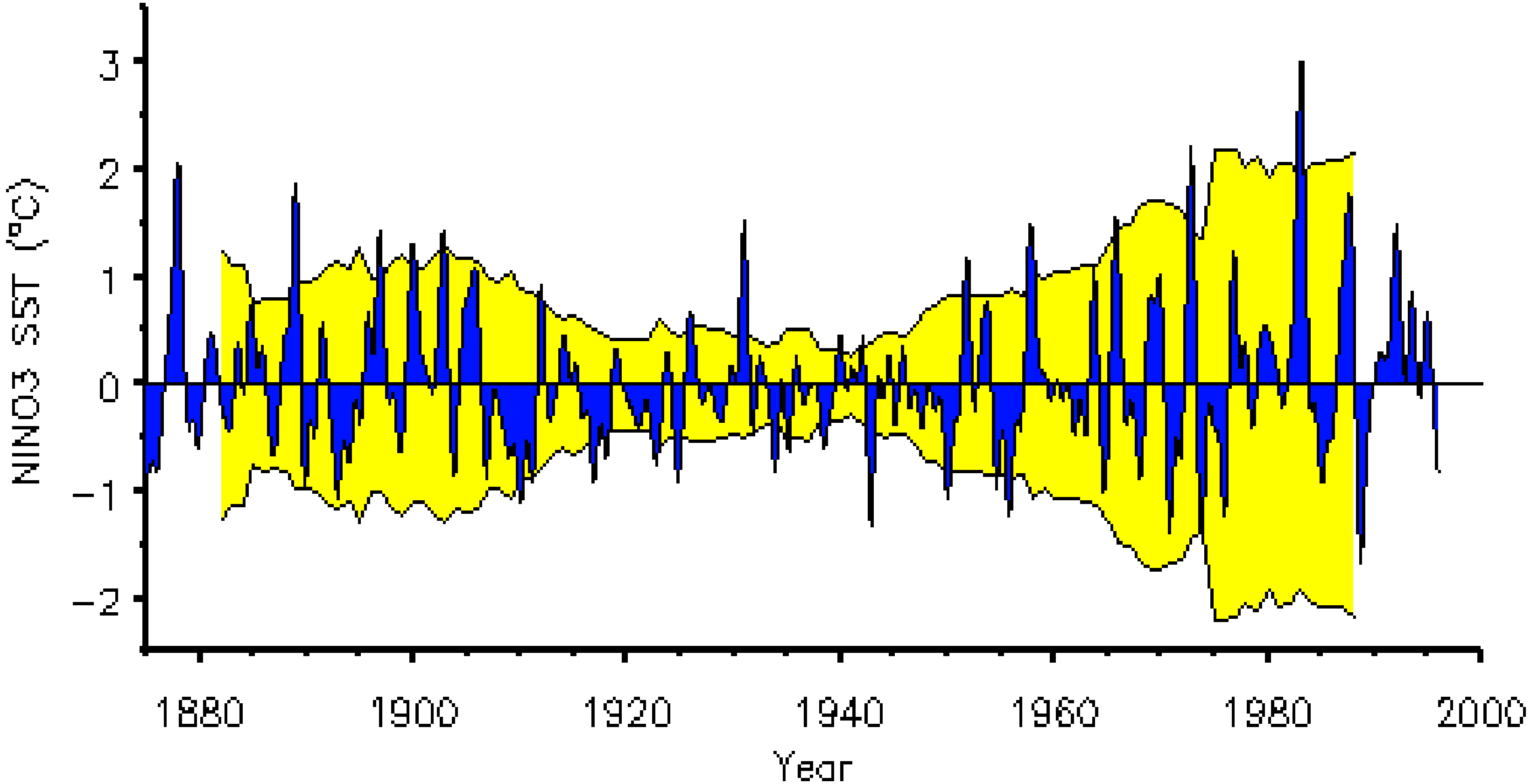
# Real Example



# Wavelet Analysis

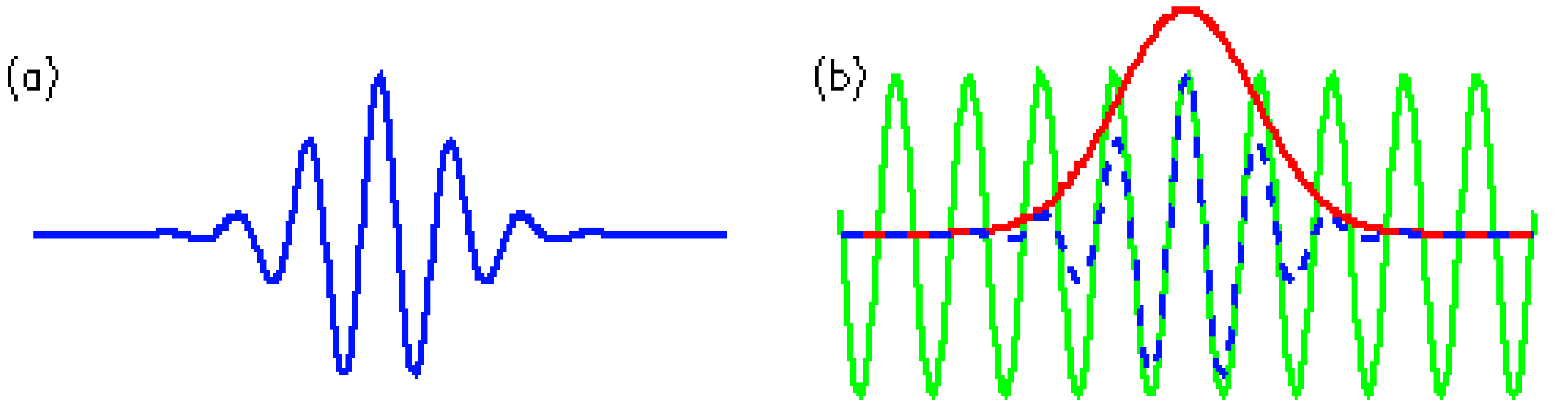
http://paos.colorado.edu/research/wavelets/

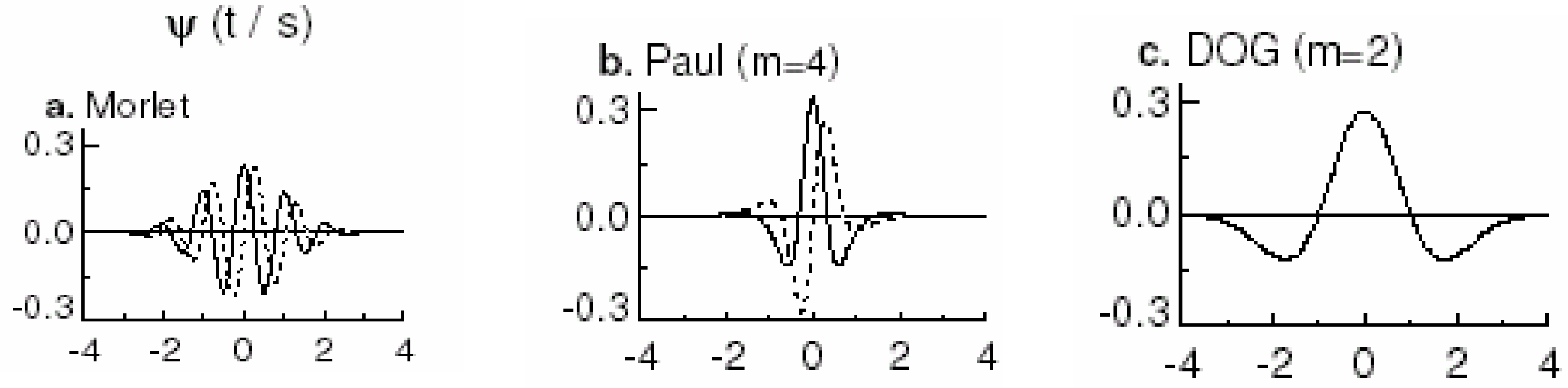
# Data: NINO3 SST




Sea surface temperatures averaged over the NINO3 region with 15 years moving variance

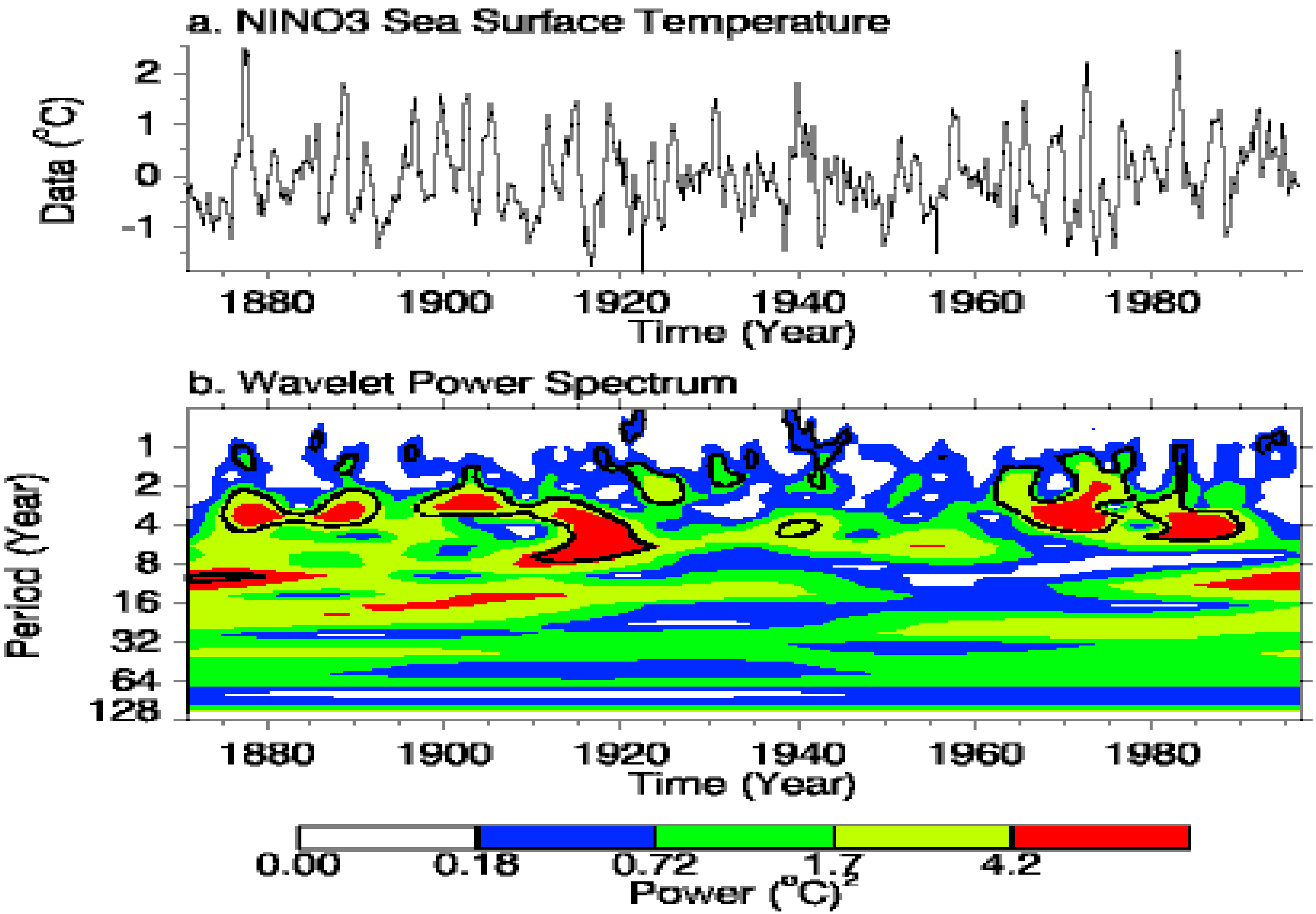
# Morlet wavelet



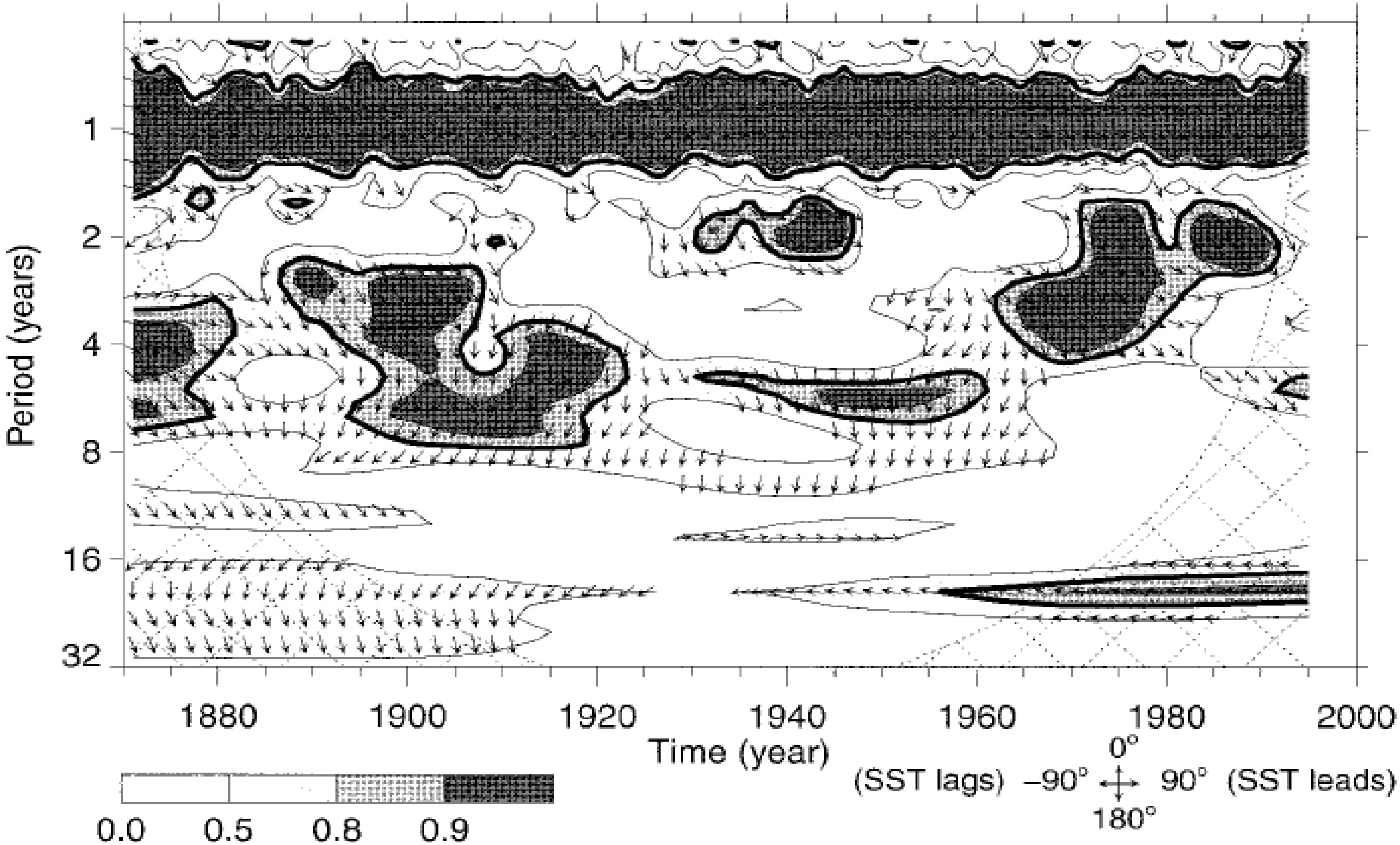




Morlet wavelet of arbitrary width and amplitude

# Waveleght analysis of NINO3 SST



# The wavelet coherency and phase



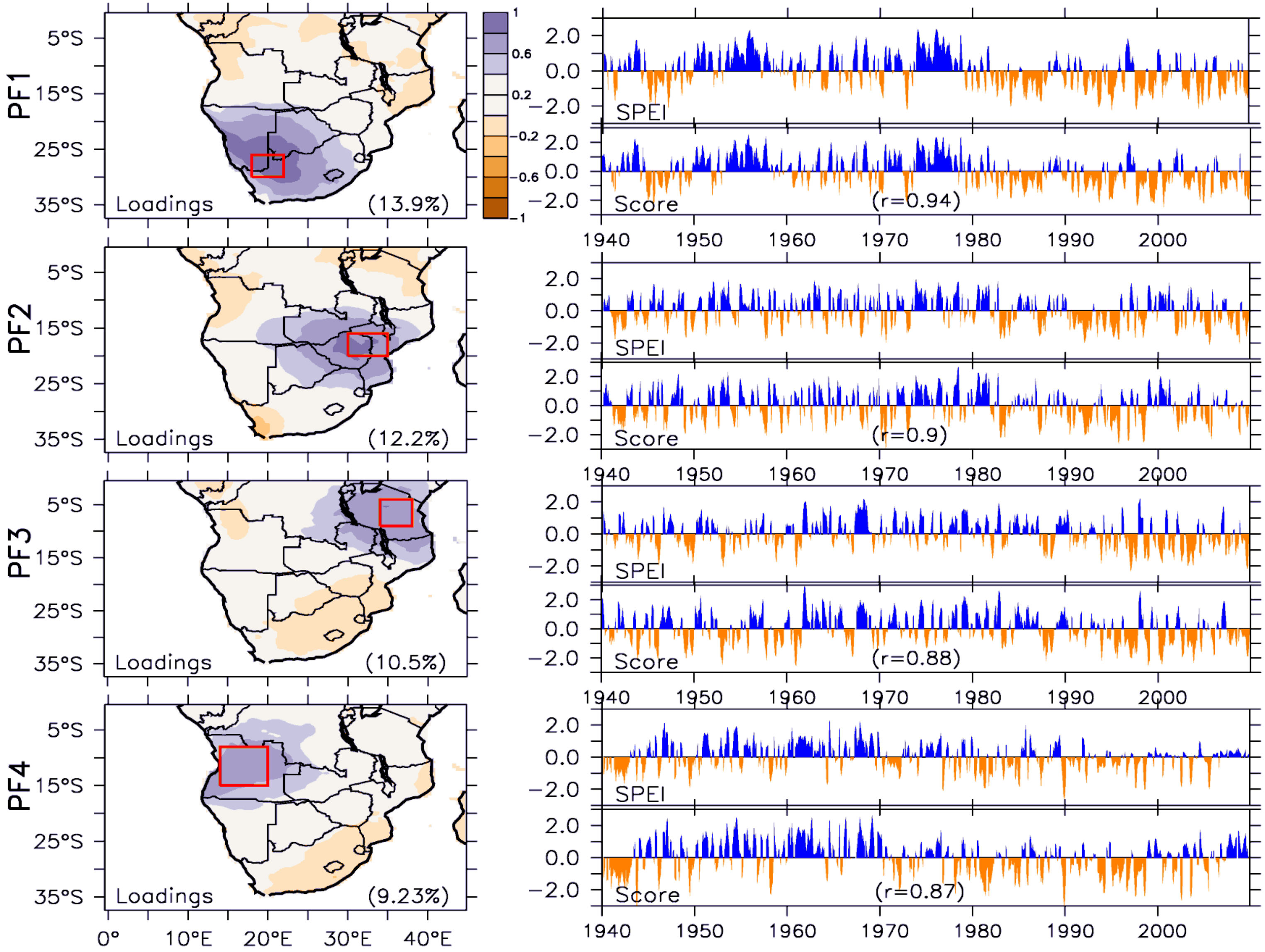

The wavelet coherency and phase between NNO3 and Indian Rainfall

Torreence at al 1999

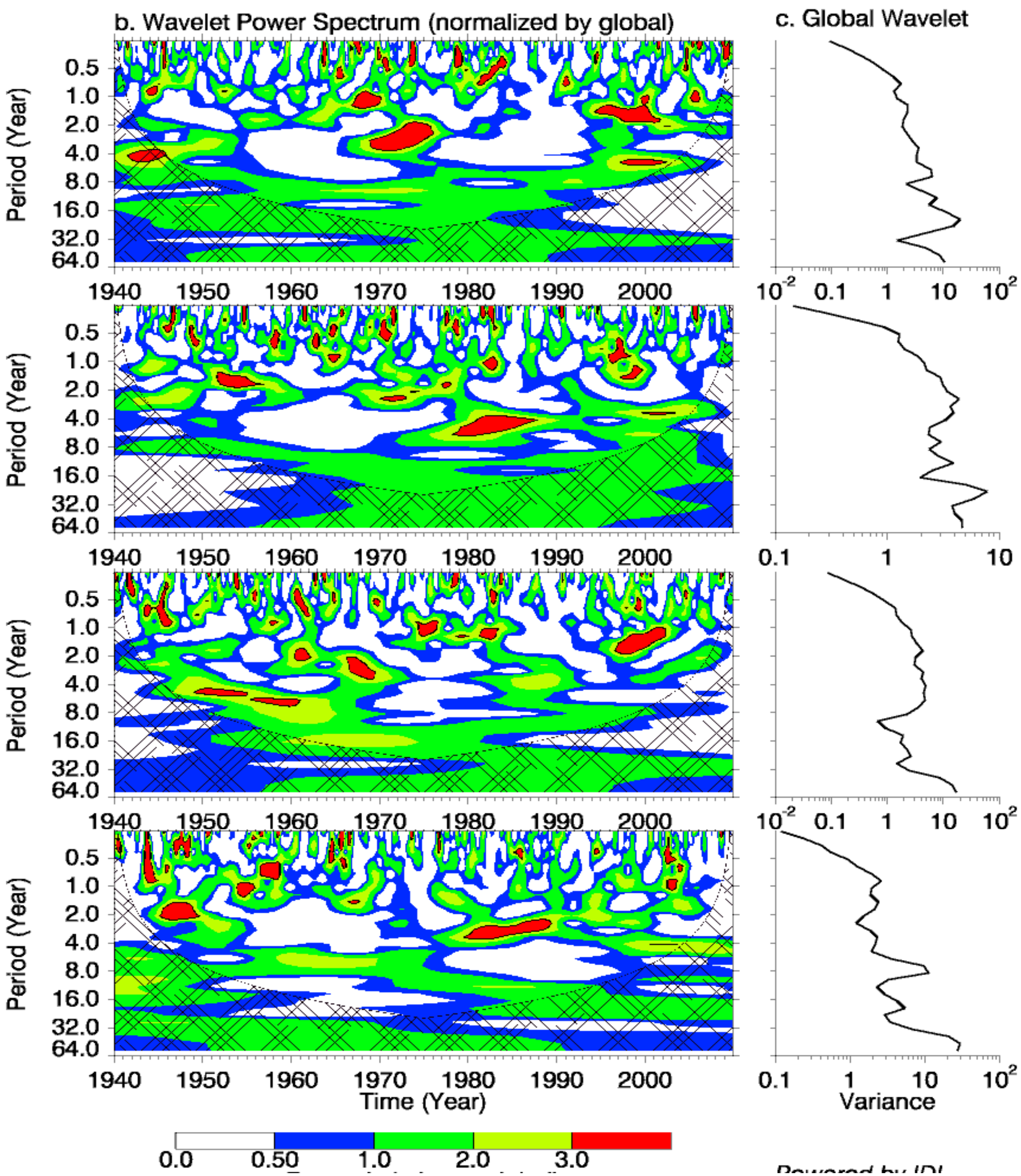
# Example from a student's thesis

(Ujeneza and Abiodun, 2014)

# PCA of SPEI over South Africa



# Wavelet Power Spectrum of the PFs



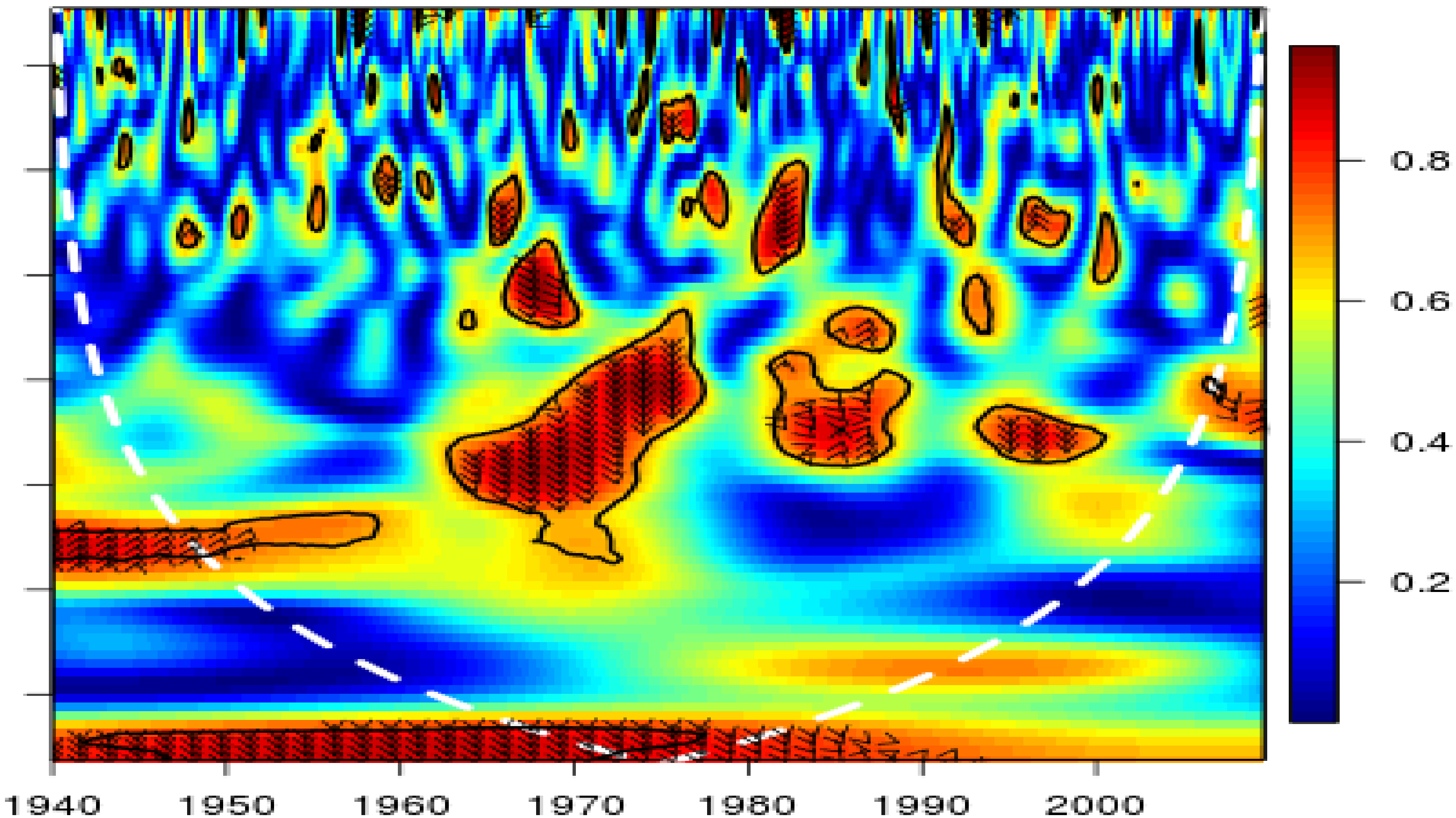

PF1

PF2

PF3

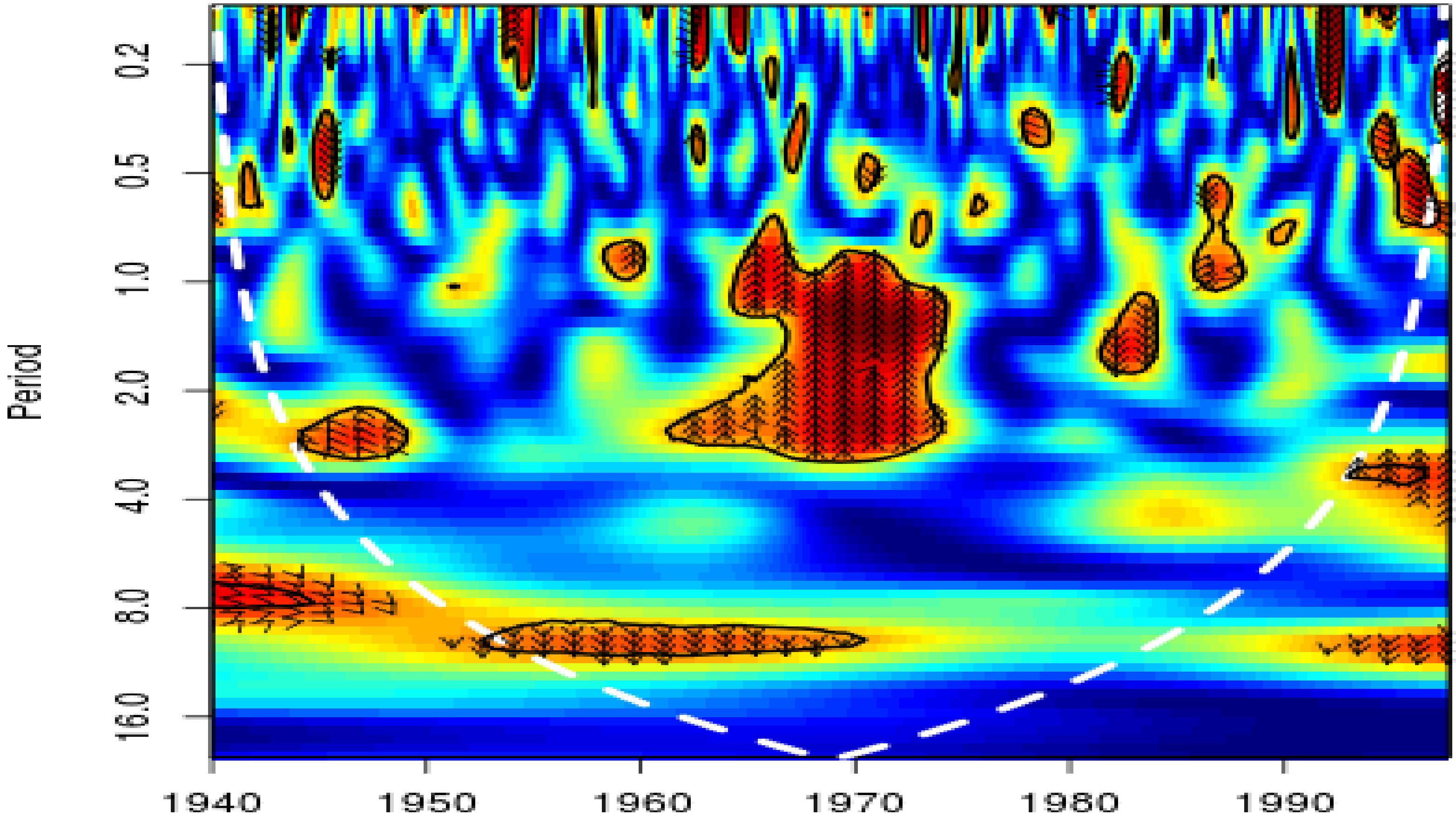
PF4

# PFI and SOI



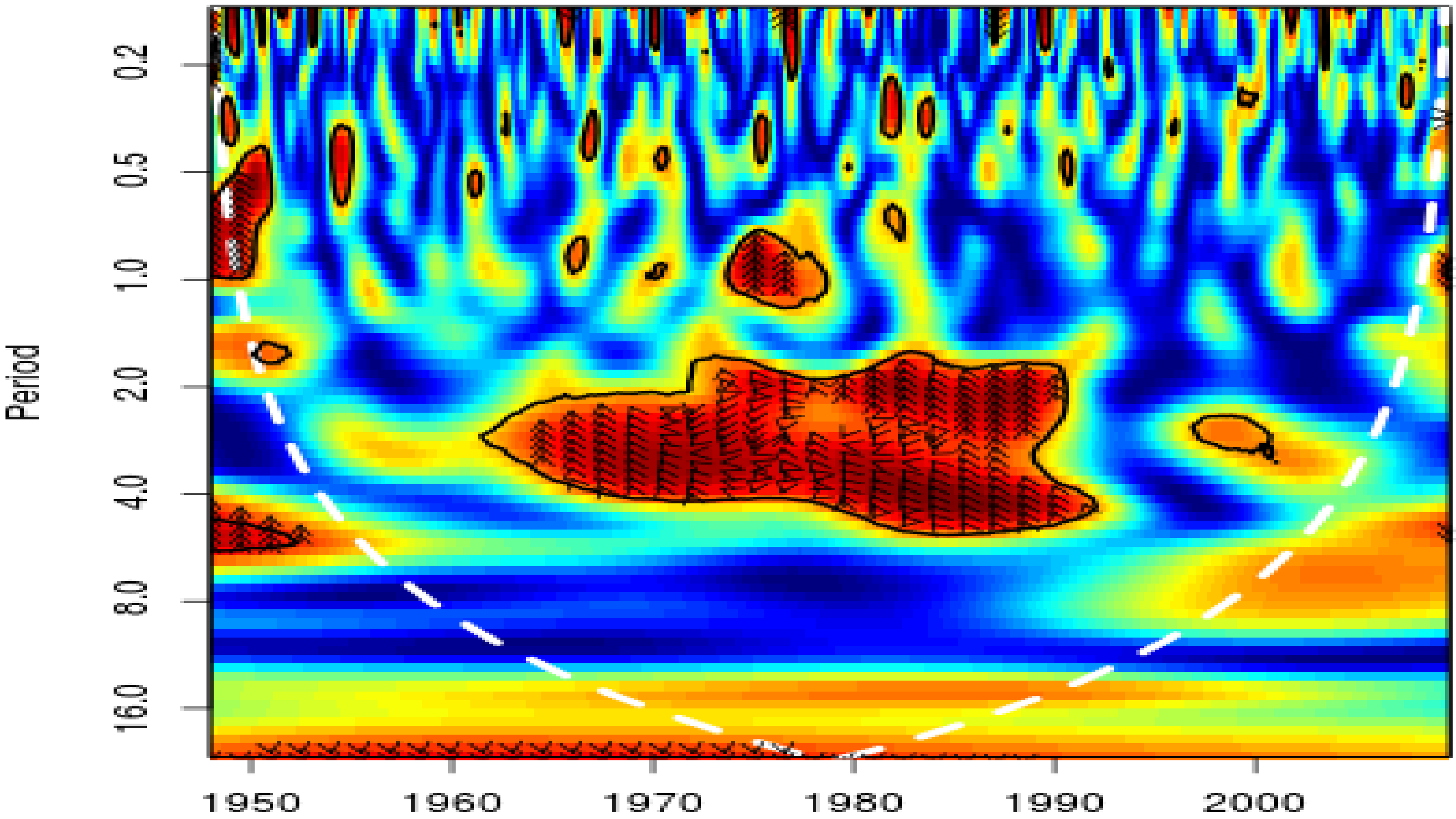

SOI: Southern Oscillation Index

# PF2 and IOD



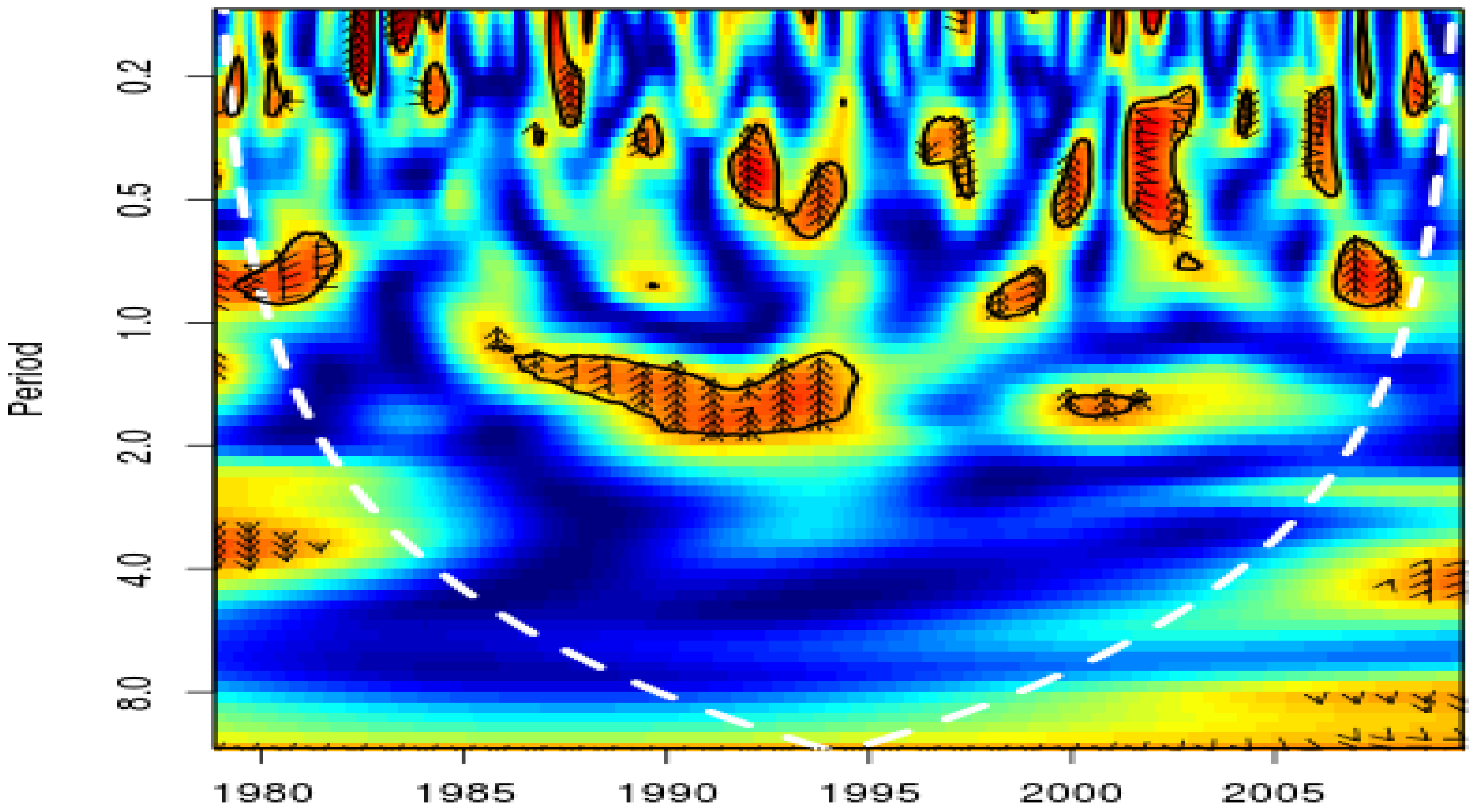

IOD: Indian Ocean Oscillation Dipole

# PF3 and TSA



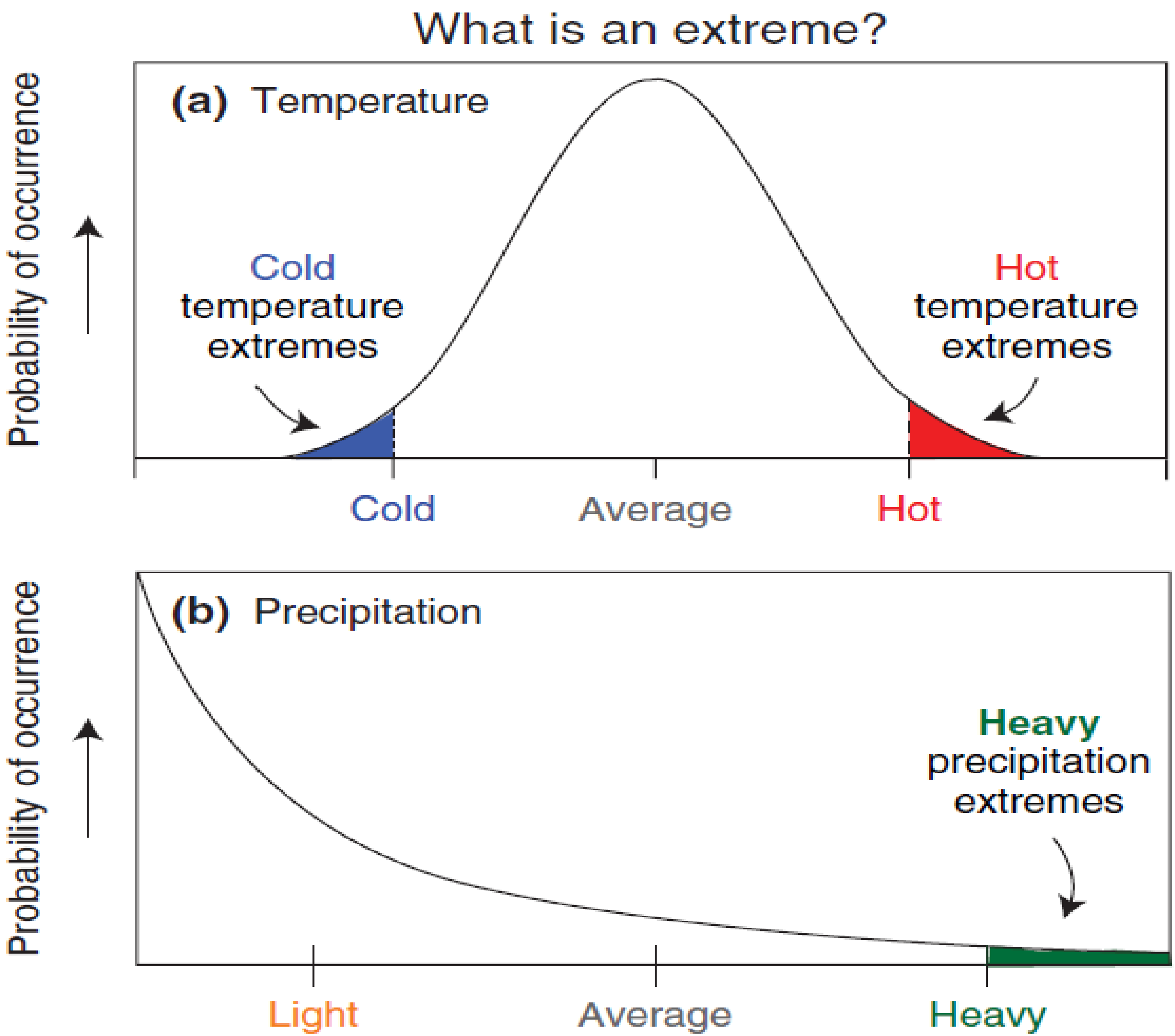

TSA: Tropical Southern Atlantic Oscilation

# PF4 and AAO



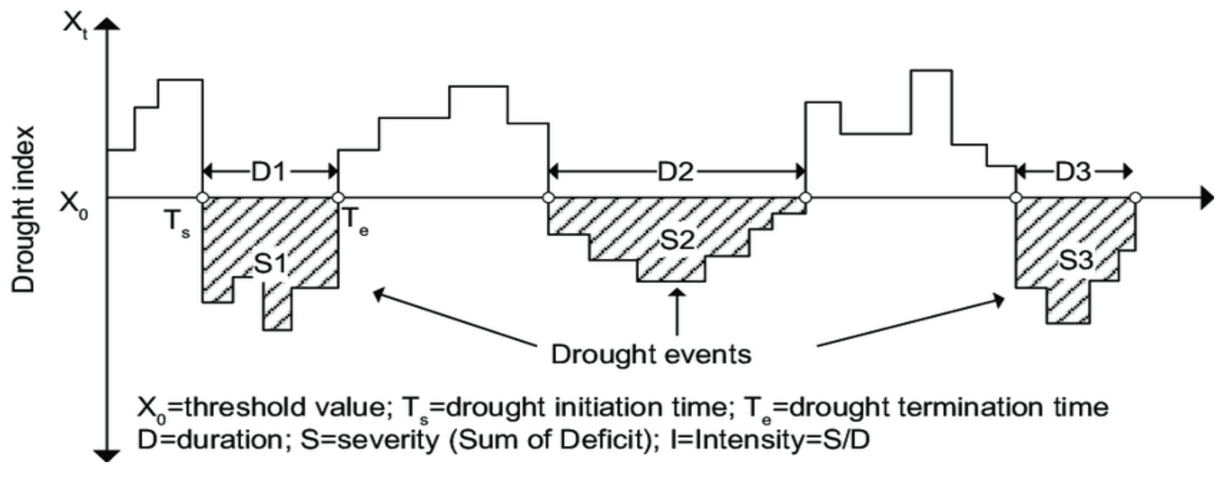

AAO: North Atlantic Oscillation

# Analysis of Extreme Weather and Climate Events


What is an extreme?
(a) Temperature
Probability of occurrence
Cold
temperature
extremes
Hot
temperature
extremes
Cold
Average
Hot
(b) Precipitation
Probability of occurrence
Heavy
precipitation
extremes
Light
Average
Heavy

- **Extreme Weather and Climate Events (EWCE)**
  - Heat wave
  - Cold wave
  - Extreme precipitation
  - Drought

- **Guidance on characterizing EWCE**
  - Location
  - Magnitude
  - Duration
  - Extent



# Applications of Extreme Event Analysis

- Risk management
  - Risk identification
  - Risk reduction
  - Risk transfer
- Research
  - Tracking trends in event frequency, severity and distribution
  - On causal contributions of hazards, exposure and vulnerability to losses

Joint WMO CLIVAR/CCl/JCOMM/GEWEX

# Expert Team on Climate Change Detection and Indices

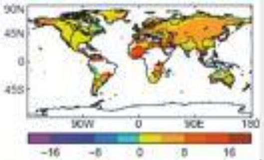
- Develop indices relevant to climate change monitoring and detection
- Create observational dataset(s) of indices with (ideally) global coverage (www.climdex.org)
- Indices are:
  - ❑ Statistically robust
  - ❑ Easy to understand
  - ❑ Globally valid
  - ❑ Enable comparison of modeled data and observations

http://www.clivar.org/panels-and-working-groups/etccdi/etccdi.php

# ETCCDI indices

http://etccdi.pacificclimate.org/list_27_indices.shtml

## ETCCDI/CRD Climate Change Indices



Indices
Definition
Calculation
Homogeneity
Examples
Software
Data
Workshops
Home

### Climate Change Indices

#### Definitions of the 27 core indices

1. FD, *Number of frost days*: Annual count of days when TN (daily minimum temperature) < 0°C.

   Let $TN_{ij}$ be daily minimum temperature on day *i* in year *j*. Count the number of days where:

   $TN_{ij} < 0^oC$.

2. SU, *Number of summer days*: Annual count of days when TX (daily maximum temperature) > 25°C.

   Let $TX_{ij}$ be daily maximum temperature on day *i* in year *j*. Count the number of days where:

   $TX_{ij} > 25^oC$.

3. ID, *Number of icing days*: Annual count of days when TX (daily maximum temperature) < 0°C.

   Let $TX_{ij}$ be daily maximum temperature on day *i* in year *j*. Count the number of days where:

   $TX_{ij} < 0^oC$.

4. TR, *Number of tropical nights*: Annual count of days when TN (daily minimum temperature) > 20°C.

   Let $TN_{ij}$ be daily minimum temperature on day *i* in year *j*. Count the number of days where:

   $TN_{ij} > 20^oC$.

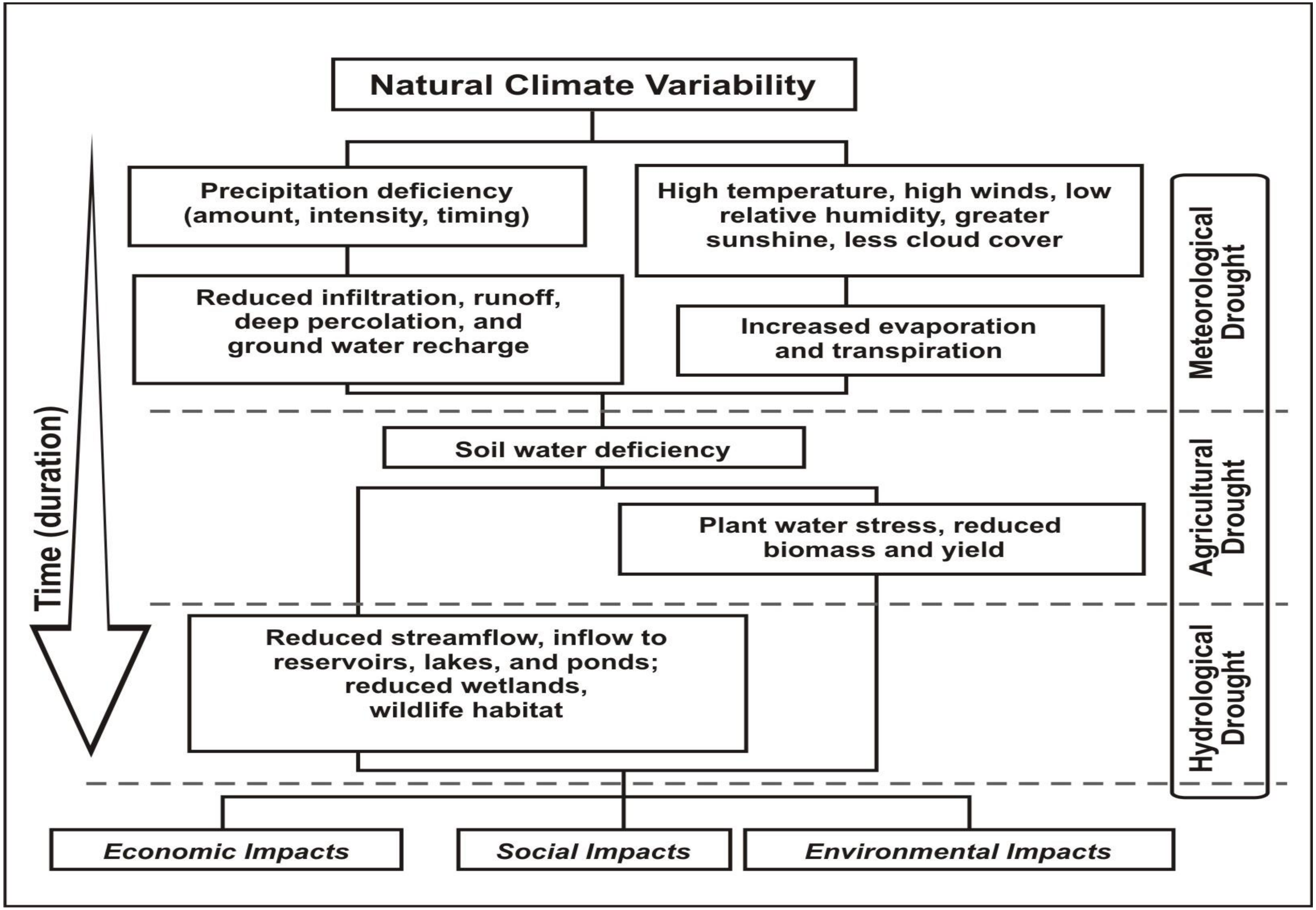
# DROUGHT ANALYSIS AND MONITORING

# WHAT IS DROUGHT?

**"Drought is caused by not only lack of precipitation and high temperatures but by overuse and overpopulation,"**

**- David Miskus.**

# TYPES OF DROUGHT



# EFFECTS OF DROUGHT

# DROUGHT INDICATIORS VS INDICES

- Indicators: variables or parameters used to describe drought conditions.
  - **Examples**: precipitation, temperature, stream groundwater and reservoir levels, soil moisture and snowpack.

- Indices: typically computed numerical representations of drought severity or intensity using climatic, hydrologic, or remotely sensed inputs.

# CONCLUSION

- Drought is a complex natural phenomena without a general and commonly accepted definition

- It is very difficult to objectively quantify their characteristics in terms of intensity, magnitude, duration and spatial extent.

- Due to this reason several drought indices have been developed

- A list of indices used https://library.wmo.int/pmb ged/wmo_1173_en.pdf

# The Palmer Drought Severity Index(PDSI)

- This uses soil moisture as a measure of drought conditions
- Based on a simplified water balance equation.

| Palmer Classifications | |
|---|---|
| 4.0 or more | extremely wet |
| 3.0 to 3.99 | very wet |
| 2.0 to 2.99 | moderately wet |
| 1.0 to 1.99 | slightly wet |
| 0.5 to 0.99 | incipient wet spell |
| 0.49 to -0.49 | near normal |
| -0.5 to -0.99 | incipient dry spell |
| -1.0 to -1.99 | mild drought |
| -2.0 to -2.99 | moderate drought |
| -3.0 to -3.99 | severe drought |
| -4.0 or less | extreme drought |

❖ **Strengths**

- ➢ Effective in determining long-term drought
- ➢ Takes into account the basic global warming effect

❖ **Weaknesses**

- ➢ **Lacks multi-scale features of indices**
- ➢ **Does not account for snow or ice**

- **It is mainly used to identify agricultural drought**

# The Standardized Precipitation Index (SPI)

- Based on probability of precipitation for any time scale
- SPI = (X-Xm)/σ

  where X=Precipitation for the station

  Xm=Mean precipitation

  σ =Standard deviation

Table II. Classification of the standardized precipitation index (SPI).

| SPI values | Drought class |
|---|---|
| ≥ 2 | Extremely wet |
| 1.5-1.99 | Very wet |
| 1-1.49 | Moderately wet |
| –0.99-0.99 | Near-normal |
| –1-1.49 | Moderately dry |
| –1.5-1.99 | Severely dry |
| ≤ 2 | Extremely dry |

❖ **Strengths**

- ➢ uses precipitation only
- ➢ more comparable across with different climates than the PDSI
- ➢ less complex to calculate than the PDSI

❖ **Weaknesses**

- ➢ Sensitive to the quantity and reliability of data
- ➢ Does not account for evapotranspiration

- Can identify any type of drought depending on the timescale:
- (1-month SPI, 3-month SPI etc)

## The Standardized Precipitation Evapotransipiration Index (SPEI)

- **Extension of the SPI**
- **Uses the difference between precipitation and reference evapotranspiration**

**(P - PET),**

**where PET =the potential evapotranspiration**

- **Strengths**
  - ➢ **account for the impact of temperature on a drought situation**
  - ➢ **The output is applicable for all climate regimes**
- **Weaknesses**
  - ➢ **Requires more data than the SPI**
  - ➢ **Sensitive to the method used to calculate PET**

- **Similar to SPI, it can identify variety of droughts**

# The Standardized Water-level Index(SWI)

- **It uses data from wells to investigate the impact of drought on groundwater recharge.**

$$SWI = (W-Wm)/\sigma$$

**where *W*= is the seasonal water level,**

***Wm*= its seasonal mean**

**$\sigma$ = standard deviation.**

**Results can be interpolated between points.**

❖ **Strengths**
- ➢ **Measures the impact of drought on groundwater**

❖ **Weaknesses**
- ➢ **Only takes groundwater into account**
- ➢ **interpolation between points may not be representative of the region or climate regime**

**For areas with frequent seasonal low flows on main rivers and streams**

# Soil Water Storage(SWS)

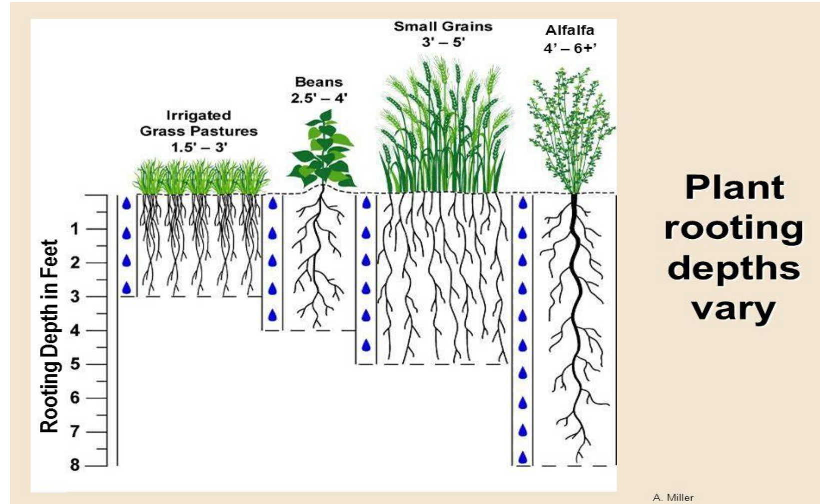
- **Identifies the amount of available moisture within a plant's root zone.**

  **SWS = RD - AWSC;**

  **where**

  **RD=Rooting Depth**

  **AWSC=Available Water Storage Capacity of the soil(mm/m)**




- **Maximum Soil Water Deficit(MWSD)**

  **MWSD = SWS - AC;**

**Where AC=Available Coefficient of the water of the crop**

❖ **Strengths**

➢ **Calculations are simple and easy to follow**

❖ **Weaknesses**

➢ **in areas where the soils are not homogeneous, there may be large changes in small distances**

**Available Water Capacities of Soils**

| Texture class | AWC (mm. water/ m. soil) |
|---|---|
| Clay | 200 |
| Clay loam | 200 |
| Silt loam | 208 |
| Clay loam | 200 |
| Loam | 175 |
| Fine sandy loam | 142 |
| Sandy loam | 125 |
| Loamy sand | 100 |
| Sand | 83 |

- **Used for monitoring drought in agricultural contexts**

# Choosing appropriate Drought Index

- Drought specific
- Available information
- Parameters used
- Spatial scale
- Impact of the drought vegetation, agriculture and water level

# Choosing appropriate Drought Index

| *Meteorology* | *Page* | *Ease of use* | *Input parameters* | *Additional information* |
|---|---|---|---|---|
| Aridity Anomaly Index (AAI) | 11 | Green | P, T, PET, ET | Operationally available for India |
| Deciles | 11 | Green | P | Easy to calculate; examples from Australia are useful |
| Keetch–Byram Drought Index (KBDI) | 12 | Green | P, T | Calculations are based upon the climate of the area of interest |
| Percent of Normal Precipitation | 12 | Green | P | Simple calculations |
| Standardized Precipitation Index (SPI) | 13 | Green | P | Highlighted by the World Meteorological Organization as a starting point for meteorological drought monitoring |
| Weighted Anomaly Standardized Precipitation (WASP) | 15 | Green | P, T | Uses gridded data for monitoring drought in tropical regions |
| Aridity Index (AI) | 15 | Yellow | P, T | Can also be used in climate classifications |
| China Z Index (CZI) | 16 | Yellow | P | Intended to improve upon SPI data |
| Crop Moisture Index (CMI) | 16 | Yellow | P, T | Weekly values are required |
| Drought Area Index (DAI) | 17 | Yellow | P | Gives an indication of monsoon season performance |

# Choosing appropriate Drought Index

| *Soil moisture* | *Page* | *Ease of use* | *Input parameters* | *Additional information* |
|---|---|---|---|---|
| Soil Moisture Anomaly (SMA) | 25 | Yellow | P, T, AWC | Intended to improve upon the water balance of PDSI |
| Evapotranspiration Deficit Index (ETDI) | 26 | Red | Mod | Complex calculations with multiple inputs required |
| Soil Moisture Deficit Index (SMDI) | 26 | Red | Mod | Weekly calculations at different soil depths; complicated to calculate |
| Soil Water Storage (SWS) | 27 | Red | AWC, RD, ST, SWD | Owing to variations in both soil and crop types, interpolation over large areas is challenging |

| *Hydrology* | *Page* | *Ease of use* | *Input parameters* | *Additional information* |
|---|---|---|---|---|
| Palmer Hydrological Drought Severity Index (PHDI) | 27 | Yellow | P, T, AWC | Serially complete data required |
| Standardized Reservoir Supply Index (SRSI) | 28 | Yellow | RD | Similar calculations to SPI using reservoir data |
| Standardized Streamflow Index (SSFI) | 29 | Yellow | SF | Uses the SPI program along with streamflow data |
| Standardized Water-level Index (SWI) | 29 | Yellow | GW | Similar calculations to SPI, but using groundwater or well-level data instead of precipitation |