Department of Geography and Earth Sciences

WRM 625-Hydrological Modeling

Lecture 4- Time Series Analysis and Stochastic Modeling

Test 1: Thursday, 1/10/2020, 10.00am-11.00am

Introduction

- Whatever is going on around us are processes occurring in certain systems. Some obvious examples are:

  - ✓the change of weather (system: Earth atmosphere)
  - ✓the change of illumination during the day (system: Earth atmosphere)

- In lay terms: **process** is the change in time of the state of the system.

- Note: the **state of the same system** can be characterized by one or several variables.

Introduction

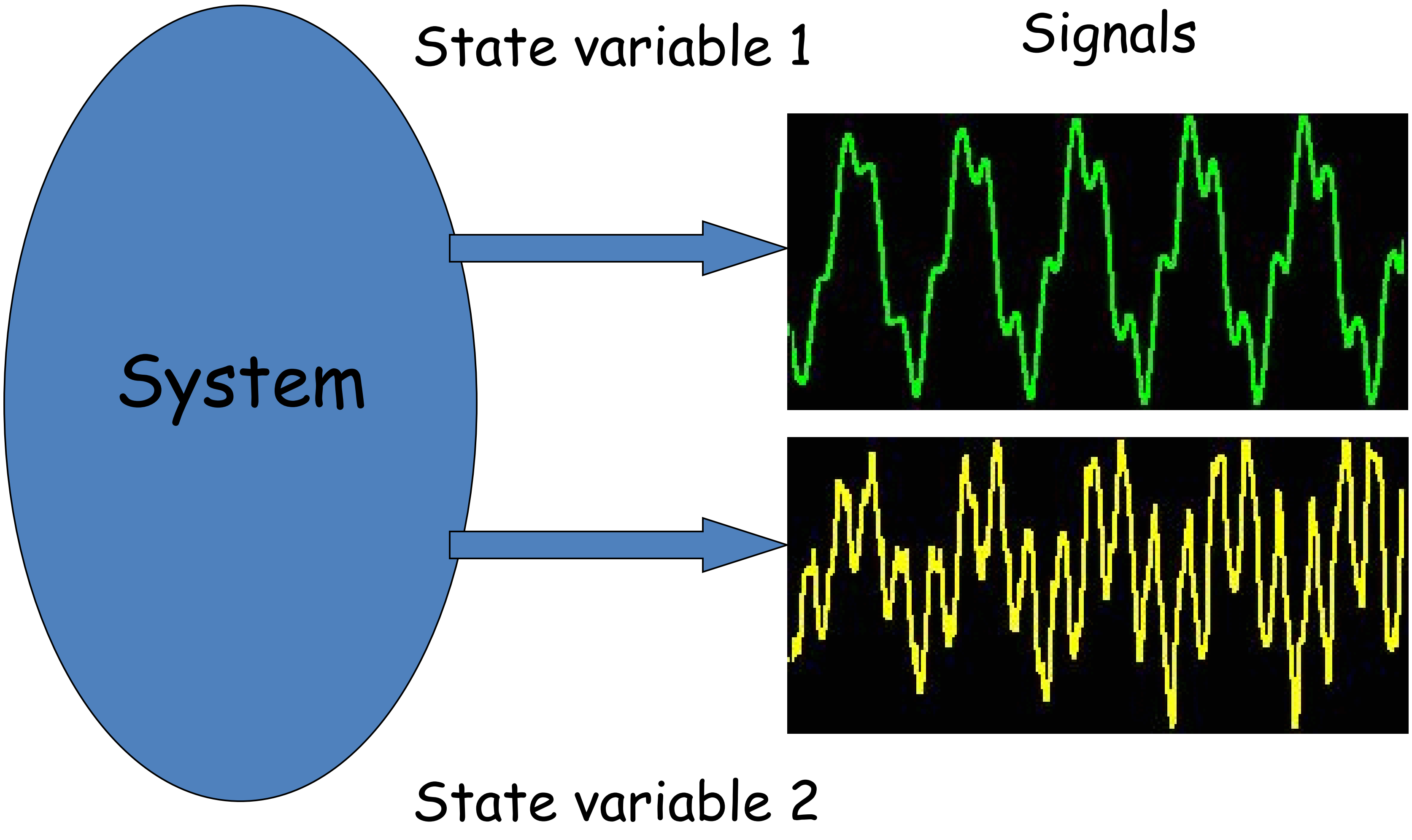
Examples of variables:

- Current weather at Chancellor College can be characterized by air temperature, humidity, wind velocity, atmosphere pressure, etc.

- One may record and observe the change in time of several, or of just one variable characterizing the system state.

- The recorded dependence of some variable in time is also called a **realization.**

System, Process and Signal



Time Series

- **Time series**: a collection of observations of state variables made sequentially in time.

- **Sequential generation**: means they are dependent to each other such that we do NOT have random sample.
  - ✓We assume that observations are equally spaced in time.
  - ✓We also assume that closer observations might have stronger dependency.

- **Univariate (bivariate, multivariate) time series**: collection of observations of one (two, several) state variables, each made at sequential time moments.

Note: the order of observations is important!

Time Series

Notations

- continuous signal $a(t)$
- sampling step Δt
- sampling frequency $f_s=1/\Delta t$
- time series $a(t_i)=a(i\Delta t)=a_i$, $i=1,2,\ldots,L$
- length of time series L

Synonyms:

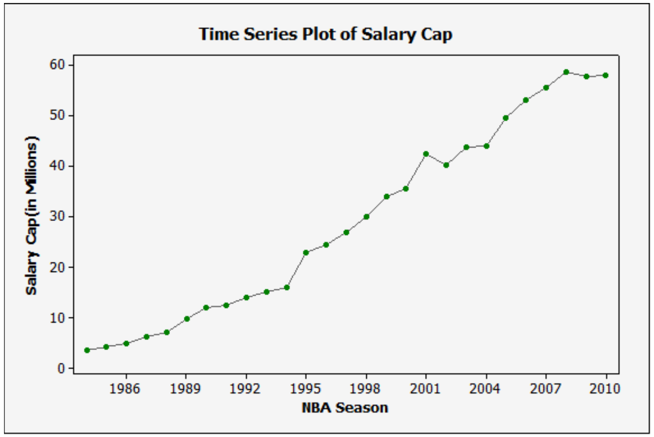
- Time series, (experimental) data, sampled signal, discretize signal
- Sampling rate (step), discretization rate (step)
- Time Series Analysis, Data Analysis, Signal Processing, Data Processing

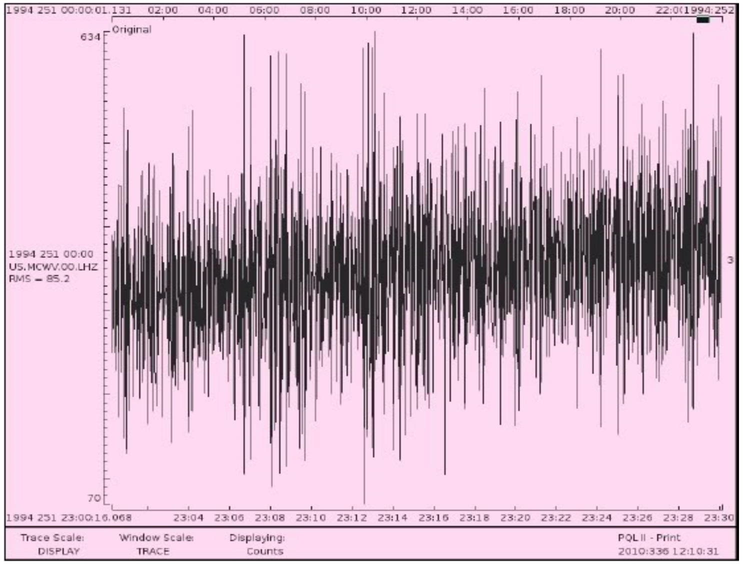
Note:

Mathematically, "time series" is not a SERIES, but a SEQUENCE!

Time series

- Discrete time series is one in which the set T_0 at which observations are made is a discrete set.
- Continuous time series are obtained when observations are recorded continuously over some time interval.





How time series can arise

1. Given a continuous signal, one can sample its values at equal time intervals. **Example:** river discharge

2. The value of the state variable aggregates (accumulates) during some time interval. **Example:** daily rainfall

3. Some processes are inherently discrete.
   Example: amount of rainfall recorded per hour

Kinds of processes

- Random (stochastic) process
- Deterministic process
- Mixed

Aims of Time Series Analysis

1. **Description**
   Describe (characterize) a generating process using its time series.

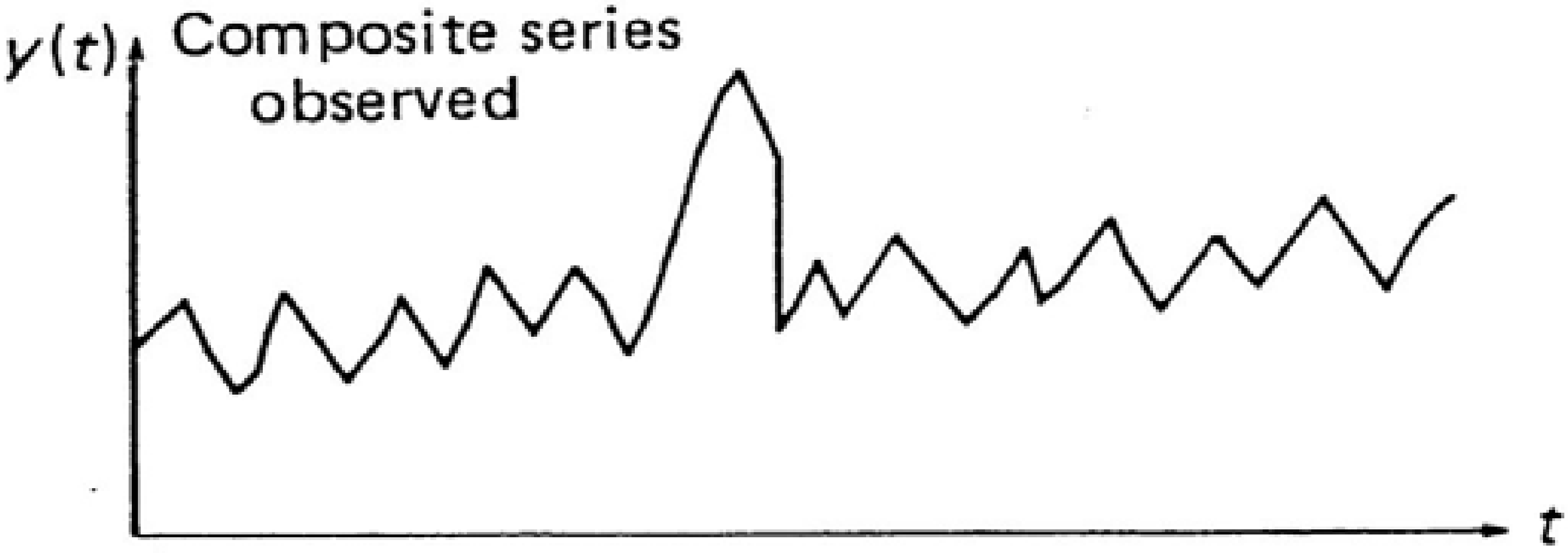
2. **Explanation**
   If time series is bi- or multi-variate, then it may be possible to use variations in one variable to explain the variations in another variable.

3. **Prediction (forecasting)**
   Use the knowledge of the past of the time series to predict its future.

4. **Control**
   To change deliberately the properties of the process by influencing it and observing the changes introduced by our intervention.

   - One can then learn to make the needed effort to achieve control.

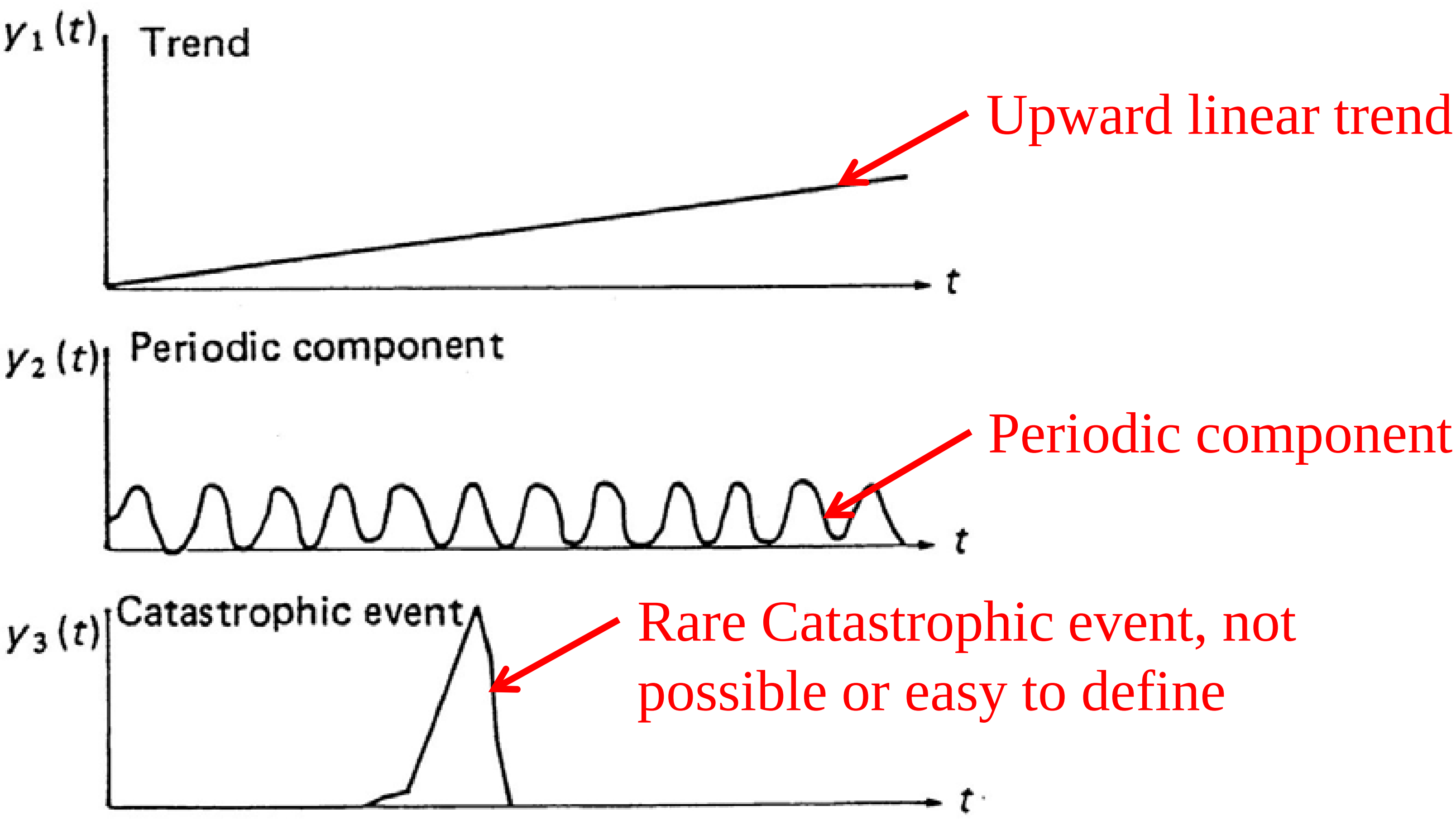
Example



A Series of Observations (y(t)) changing with time (t)

Example

Three discernible features in the pattern of the observations.



Example

- $y_4(t)$- a hidden feature of the series is the random stochastic component representing an irregular but continuing variation within the measured values and may have some persistence.
    - ✓chaotic noisy residuals left overs after all three have been accounted for.

- It may be due to:
    - instrumental of observational sampling errors
    - random unexplainable fluctuations in a natural physical process.

- A time series is said to be a random or stochastic process if it contains a stochastic component.

- Most Hydrologic time series may be thought of as stochastic processes since they contain both deterministic and stochastic components

Times series

- Can be expressed as

$$y(t) = y_1(t) + y_2(t) + y_3(t) + y_4(t)$$

Deterministic components, easier to identify or quantify

Stochastic or Random components, with persistence effects, less easier to identify or quantify

Properties of time series

- Mean
  - ✓ Sample estimation $$E(X_t) = \overline{X} = \frac{1}{n}\sum_{t=1}^{n} X_t$$
  - ✓ Population notation μ
- Variance
  - ✓Sample estimation $$S^2 = \frac{1}{n-1}\sum_{t=1}^{n}(X_t - \overline{X})^2$$
  - ✓Population notation σ^2
- Co-variance

$$Cov(X_t, X_{t+L}) = \frac{1}{n-L}\sum_{t=1}^{n}\left(X_t - \overline{X}\right)\left(X_{t+L} - \overline{X}\right)$$

Where L is the time lag.

Stationary time series

- A time series is stationary if the sample statistics are not functions of the timing or the length of the sample
- Can be stationary to the second moment, weakly stationary or strongly stationary.

$$E(X_t) = \mu$$
$$Var(X_t) = \sigma^2$$
$$Cov(X_t, X_{t+L}) = \lambda_L$$

- Basic assumption in time series as all the other components can be removed leaving them stationary.

Non-stationary time series

- A time series is non-stationary if the sample statistics are dependent on the *timing* or the *length* of the sample.
- If series have a definite trend

$$E(X_t) = \mu_t$$

$$Var(X_t) = \sigma_t^2$$

$$Cov(X_t, X_{t+L}) = \lambda_{L,t}$$

White noise time series

- For a stationary times series, if the process is purely random and stochastically independent, the time series is called a white noise series
- Represented as

$$E(X_t) = \mu$$

$$Var(X_t) = \sigma^2$$

$$Cov(X_t, X_{t+L}) = 0 \; for \, all \, L \neq 0$$

Gaussian time series

- A process (not necessarily stationary) of which all random variables are normally distributed, and of which all simultaneous distributions of random variables of the process are normal.

- A Gaussian random process that is weekly stationary, it is also strictly stationary, since the normal distribution is completely characterised by its first and second order moments

Analysis of Hydrologic Time Series

- Rainfall and river flow data are ideal for time series analysis.
- Extremes frequency analysis.
- Synthesizing long time series from short records
- Evaluation of data quality, outliers etc.
- Cross correlation to understand hydrological processes.
- Identification of the several components of a time series.
- Mathematical description (modelling) different components identified.

Analysis of Hydrologic Time Series

- Suppose you have a hydrological time series X_t

$$X_1, X_2, X_3, \ldots, X_t, \ldots$$

- Structure of X_t can be represented as

$$X_t \Leftrightarrow [T_t, P_t, E_t].$$

Basic model for time series analysis

Trend Component, Deterministic

Periodic Component, Deterministic

Stochastic Component, random with self correlated persistence

Aims of time series analysis

(1) Description and understanding of the mechanism

(2) Monte-Carlo simulation-repeated random sampling to solve deterministic problems

(3) Forecasting future evolution

Trend component

- Can be caused by
  - long-term climatic change or, in river flow, by gradual changes in a catchment's response to rainfall owing to land use changes.

- Sometimes, the presence of a trend cannot be readily identified

Methods of trend identification

- Parametric and non-Parametric approaches
- Mann (1945) and Kendall (1975)
- Mann-Kendall test
  - Non-Parametric test (Distribution free)
  - Uses the raw (un-smoothed) hydrologic data to detect possible trends.
  - The Kendall statistic was originally devised by Mann (1945)
  - Statistic Approaches normal when $n \geq 8$
  - Only calculates the significance and direction of trend

Linear regression method

- Can be used to identify if there exists a linear trend in a hydrologic time series.

- Fitting a linear regression equation with the time t as independent variable and the hydrologic data and Y as dependent variable

$$y = \alpha + \beta . t$$

testing the statistical significance of the regression coefficient β.

Models for trend

- Linear trend

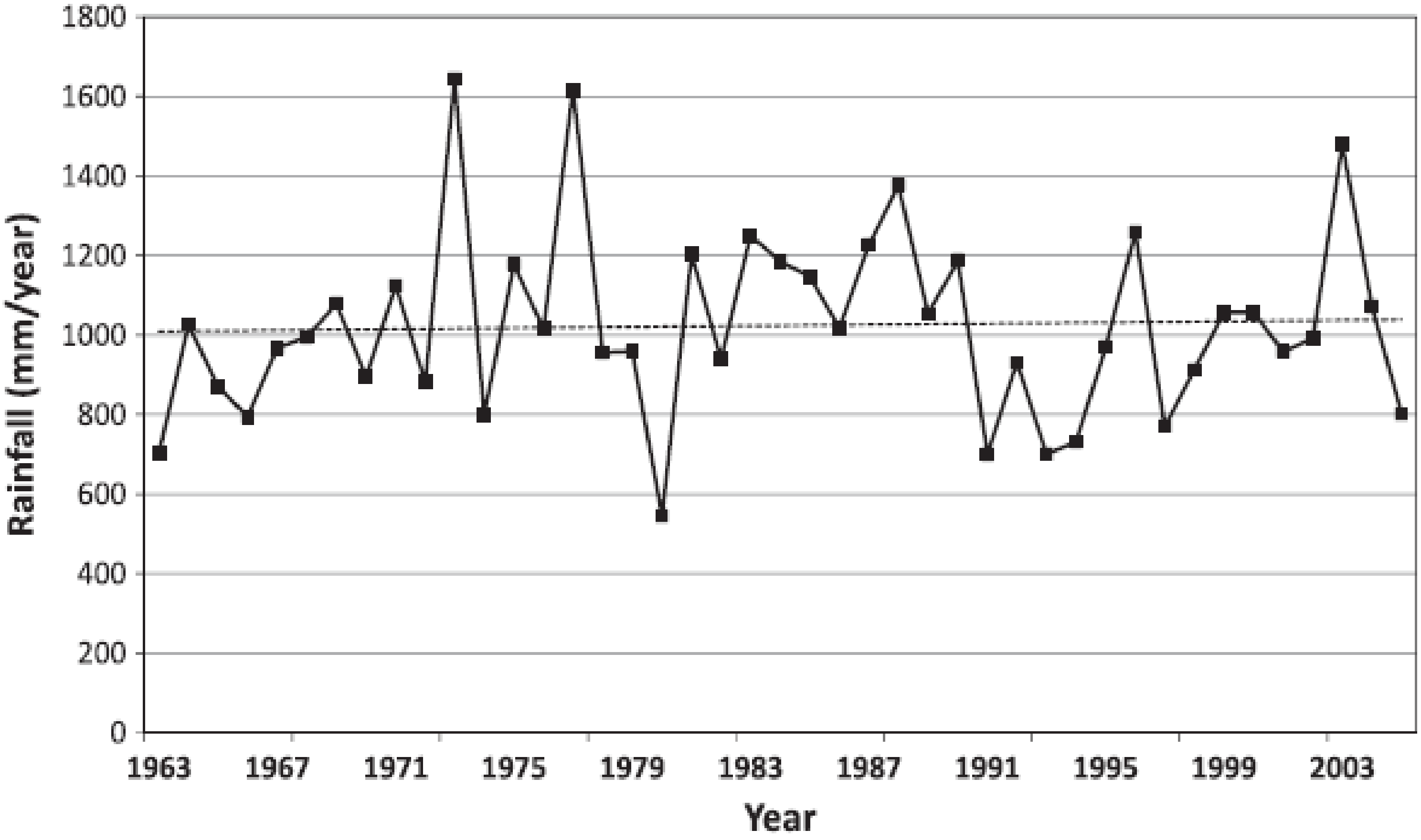
$$T_t = a + bt$$

- Non-Linear trend

$$T_t = a + bt + ct^2 + dt^4 + \ldots$$

- Coefficients a, b, c, d, ... are usually evaluated by least-squares fitting
- Number of terms restricted by phenomena, least number of parameters desired

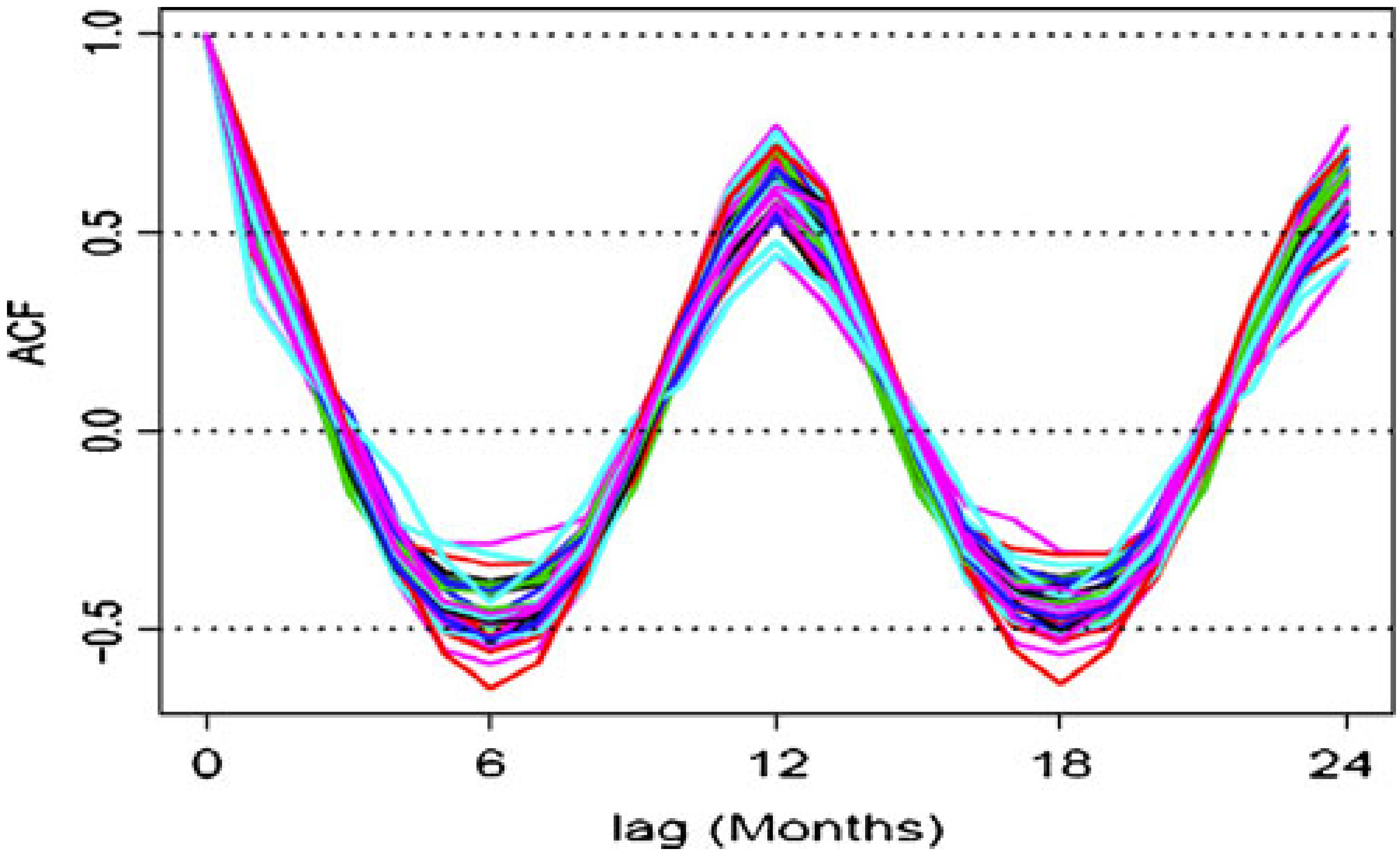
Periodic component



Annual series- No periodic component

Source: Chimtengo, Ngongondo et al. 2014, JPCE

Periodic component



Monthly series- periodic component, seasonal effects

Source: Ngongondo et al. 2011, Theor App. Clim.

Periodic component

- The existence of periodic components may be investigated quantitatively by

(1) Fourier analysis,
(2) spectral analysis, and
(3)Wavelet analysis
(4) autocorrelation analysis- widely used in hydrology

Autocorrelation analysis

- Two steps
  - calculating the autocorrelation coefficients and
  - testing their statistical significance

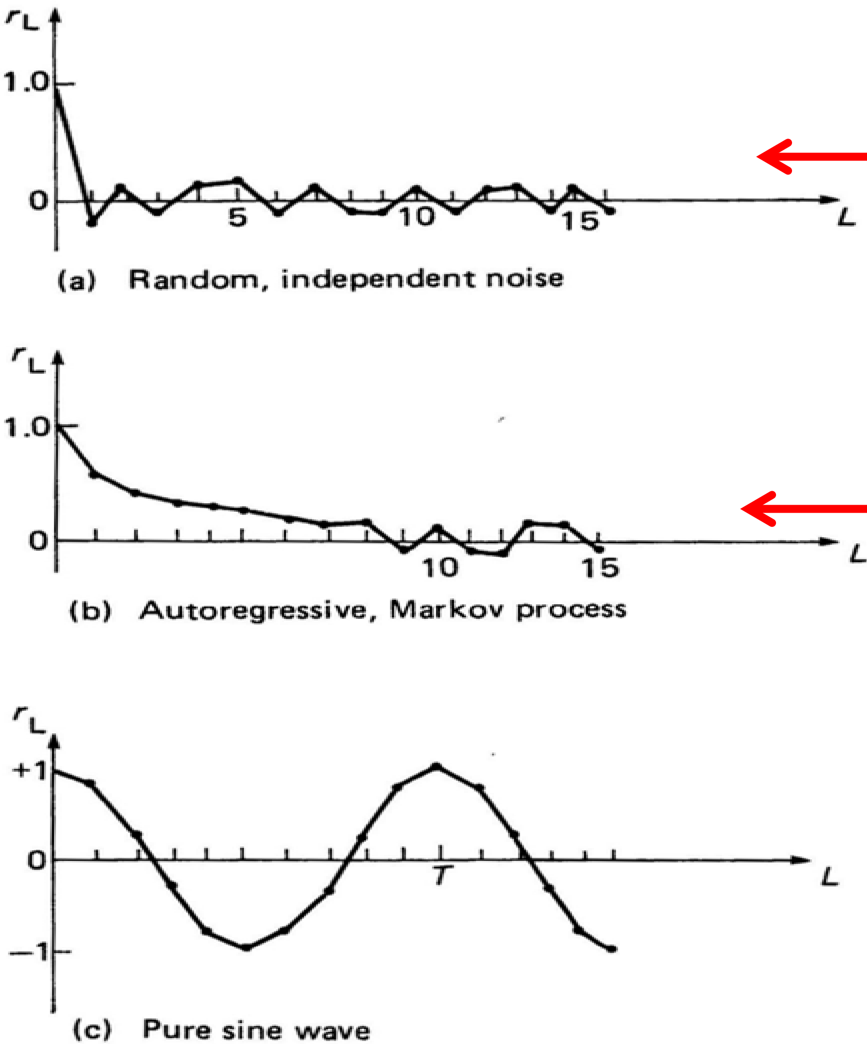
- For a time series X_t, the autocorrelation coefficient r_L between X_t and X_{t+L} are calculated and plotted against values of L for all pairs of data L time units apart in the series.

$$r_L = \frac{1}{n-L}\sum_{t=1}^{n-L}(X_t - \overline{X})(X_{t+L} - \underline{X}) \Big/ \frac{1}{n}\sum_{t=1}^{n}\left(X_t - \overline{X}\right)^2$$

- L is normally taken up to n/4

Autocorrelation analysis

- Plot on a correlogram R vs L



- If L = 0, r_L=1. That is, the correlation of an observation with itself is one.
- If $r_L \approx 0$ for all $L \neq 0$, the process is said to be a purely random process. This indicates that the observations are linearly independent of each other.

Deterministic $r_L \neq 0 \quad \mathit{for\ all}\ L \neq 0$

If $r_L \neq 0$ for some $L \neq 0$, but after $L > \tau$, $r_\tau \approx 0$, system is random

Modelling of periodic component

- A periodic function P^t is a function such that

$$P_{t+T} = P_t$$

- For all t and T is called the period i.e. smallest time it takes to complete one cycle
- Also for all integer n $P_{t+nT} = P_t$
- The frequency is defined as the number of periods per time-unit:

$$Frequency = \frac{1}{Period}$$

End of Lecture 4