INDIAN INSTITUTE OF SCIENCE

STOCHASTIC HYDROLOGY

Lecture -21

Course Instructor : Prof. P. P. MUJUMDAR

Department of Civil Engg., IISc.

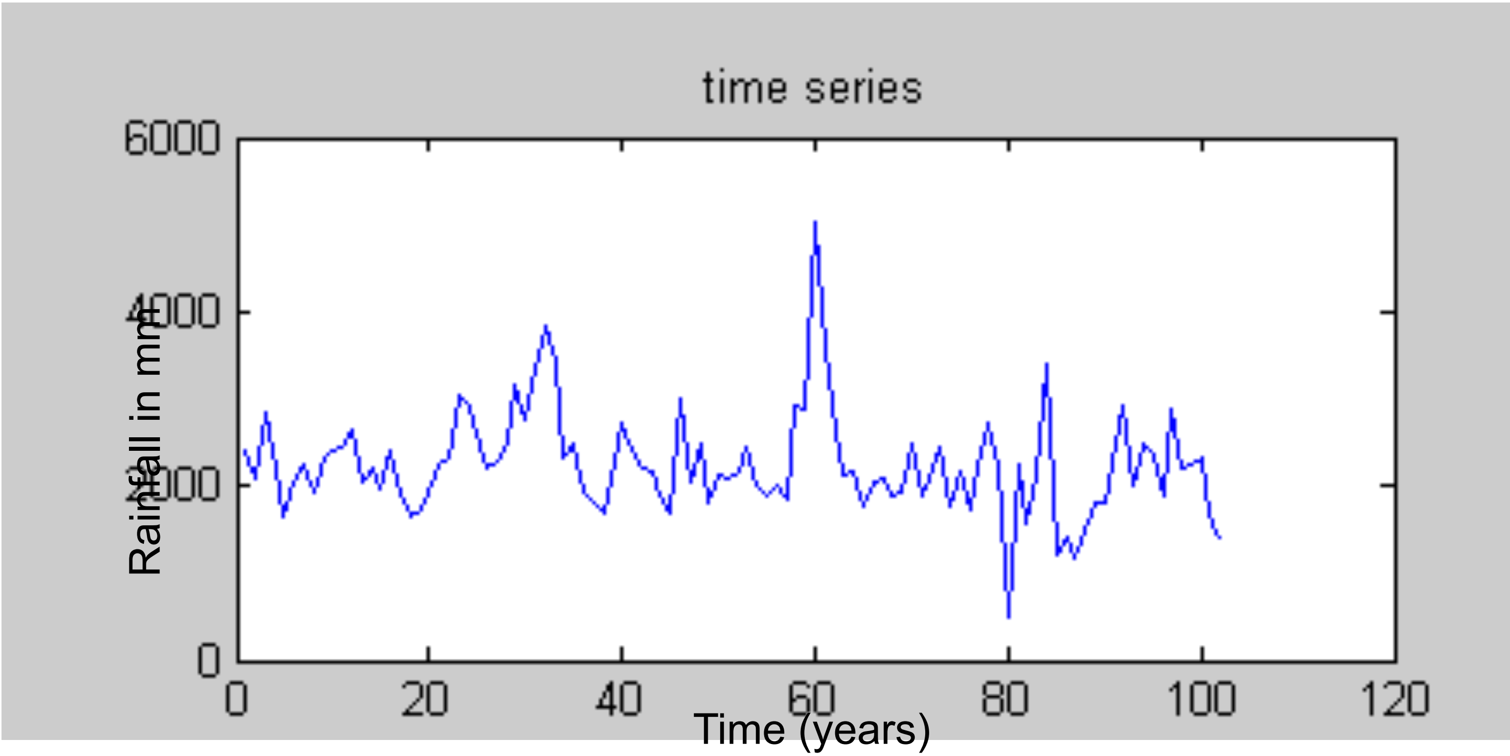
Summary of the previous lecture

- Case study -3: Monthly streamflows at KRS reservoir
  - Validation of the model
- Case study -4: Monthly streamflow of a river
  - Plots of Time series, Correlogram, Partial Autocorrelation function and Power spectrum
  - Candidate ARMA models
    - Log Likelihood
    - Mean square error
    - Validation test (Residual mean)

CASE STUDIES - ARMA MODELS

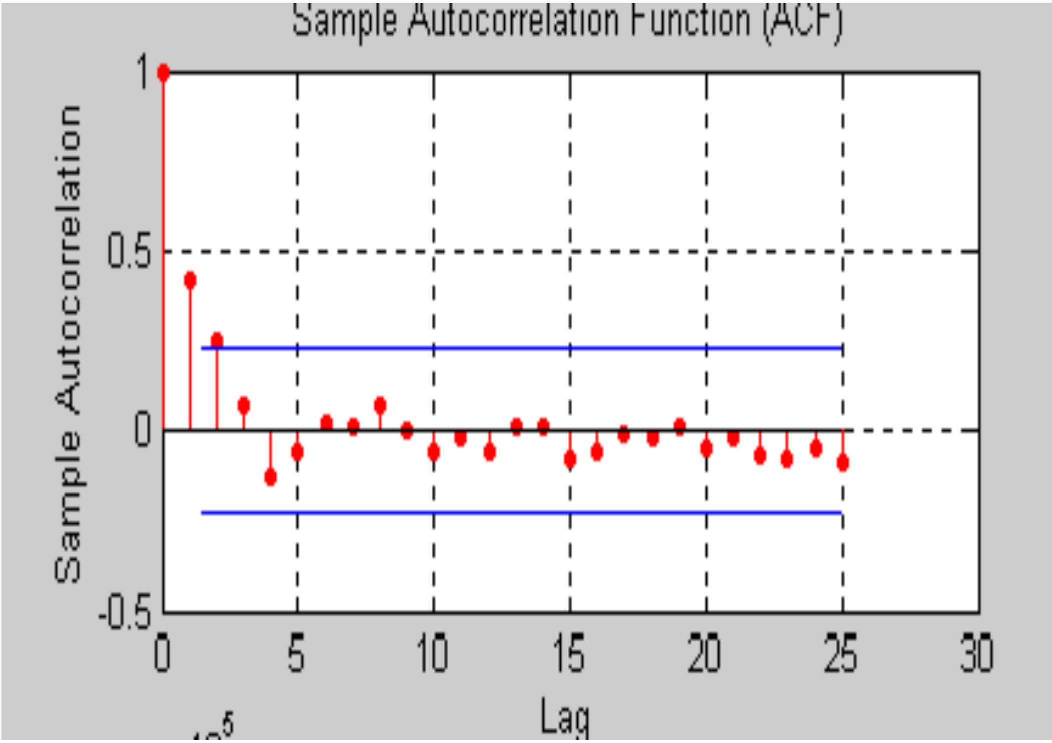
Case study – 5

Sakleshpur Annual Rainfall Data (1901-2002)



Case study – 5 (Contd.)

Correlogram



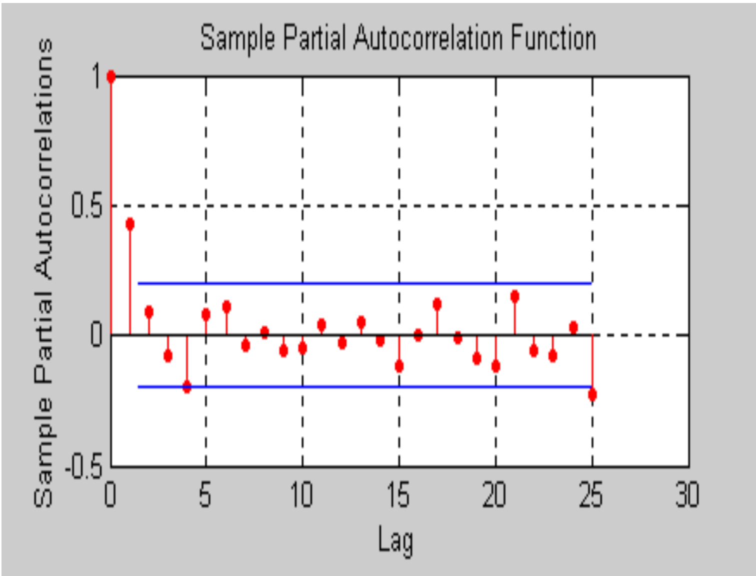
$$c_k = \frac{1}{N}\sum_{i=1}^{n-k}\left(X_t - \bar{X}\right)\left(X_{t+k} - \bar{X}\right)$$

$$r_k = \frac{c_k}{c_0}$$

$$c_0 = S_X^2$$

Case study – 5 (Contd.)

PAC function



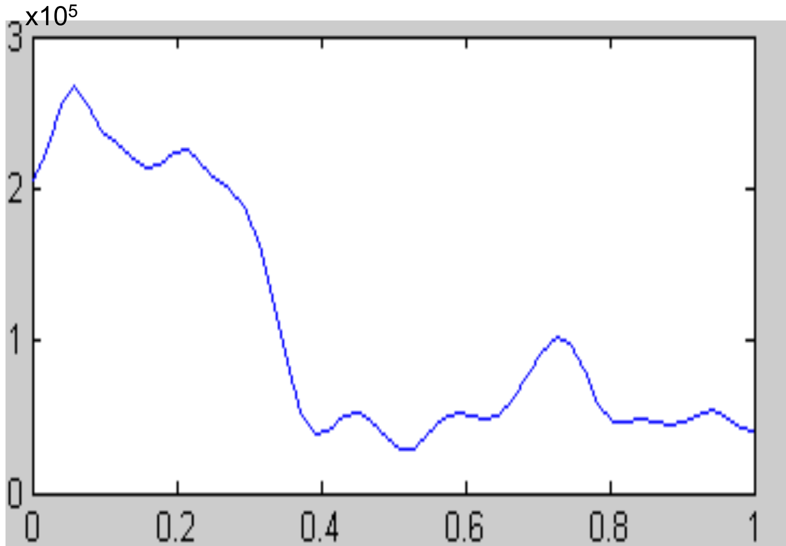
$P_p * \phi_p = \rho_p$

Auto Correlation function

Partial Auto Correlation

Auto Correlations

Power spectrum



$$I(k) = \frac{N}{2}\left[\alpha_k^2 + \beta_k^2\right] \qquad \omega_k = \frac{2\pi k}{N}$$

$$\alpha_k = \frac{2}{N}\sum_{t=1}^{n} x_t \cos(2\pi f_k t) \qquad \beta_k = \frac{2}{N}\sum_{t=1}^{n} x_t \sin(2\pi f_k t)$$

Case study – 5 (Contd.)

Model	Likelihood
AR(1)	9.078037
AR(2)	8.562427
AR(3)	8.646781
AR(4)	9.691461
AR(5)	**9.821681**
AR(6)	9.436822
ARMA(1,1)	8.341717
ARMA(1,2)	8.217627
ARMA(2,1)	7.715415
ARMA(2,2)	5.278434
ARMA(3,1)	6.316174
ARMA(3,2)	6.390390

Case study – 5 (Contd.)

- ARMA(5,0) is selected with highest likelihood value
- The parameters for the selected model are as follows

  ϕ_1 = 0.40499

  ϕ_2 = 0.15223

  ϕ_3 = -0.02427

  ϕ_4 = -0.2222

  ϕ_5 = 0.083435

  Constant = -0.000664

Case study – 5 (Contd.)

- Significance of residual mean

152

Model	$\eta(e)$	$t_{0.95}(N)$
ARMA(5,0)	0.000005	1.6601

Case study – 5 (Contd.)

Significance of periodicities:

Periodicity	η	$F_{0.95}(2, N\text{-}2)$
1^{st}	0.000	3.085
2^{nd}	0.00432	3.085
3^{rd}	0.0168	3.085
4^{th}	0.0698	3.085
5^{th}	0.000006	3.085
6^{th}	0.117	3.085

Case study – 5 (Contd.)

- Whittle's white noise test:

Model	η	$F_{0.95}(n1, N-n1)$
ARMA(5,0)	0.163	1.783

Case study – 5 (Contd.)

Model	MSE
AR(1)	1.180837
AR(2)	1.169667
AR(3)	1.182210
AR(4)	1.168724
AR(5)	1.254929
AR(6)	1.289385
ARMA(1,1)	1.171668
ARMA(1,2)	**1.156298**
ARMA(2,1)	1.183397
ARMA(2,2)	1.256068
ARMA(3,1)	1.195626
ARMA(3,2)	27.466087

$$e_t = x_t - \hat{x}_t$$

$$MSE = \frac{\sum_{t=1}^{N} e_t^2}{N}$$

N → Validation period

Case study – 5 (Contd.)

- ARMA(1, 2) is selected with least MSE value for one step forecasting
- The parameters for the selected model are as follows

  $\phi_1 = 0.35271$

  $\theta_1 = 0.017124$

  $\theta_2 = -0.216745$

  Constant = -0.009267

ARMA(1, 2)

$$X_t = \phi_1 X_{t-1} + \theta_1 e_{t-1} + \theta_2 e_{t-2} + e_t$$

Case study – 5 (Contd.)

- Significance of residual mean

Model	$\eta(e)$	$t_{0.95}(N\ )$
ARMA(1, 2)	-0.0026	1.6601

Case study – 5 (Contd.)

Significance of periodicities:

Periodicity	η	$F_{0.95}(2, N-2)$
1st	0.000	3.085
2nd	0.0006	3.085
3rd	0.0493	3.085
4th	0.0687	3.085
5th	0.0003	3.085
6th	0.0719	3.085

Case study – 5 (Contd.)

- Whittle's white noise test:

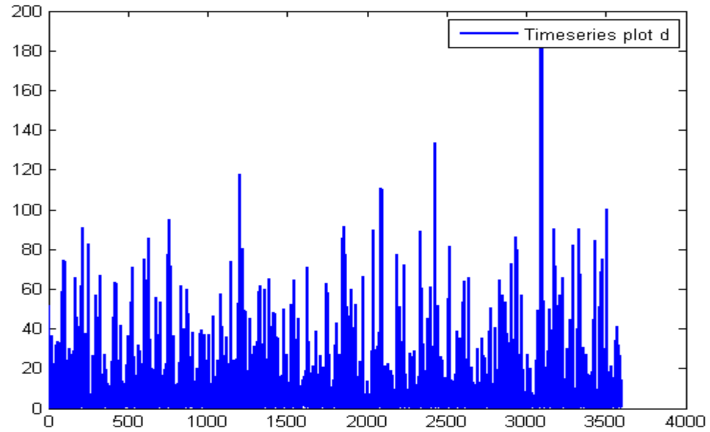
25 (= Kmax)

Model	η	$F_{0.95}(n1, N-n1)$
ARMA(1, 2)	0.3605	1.783

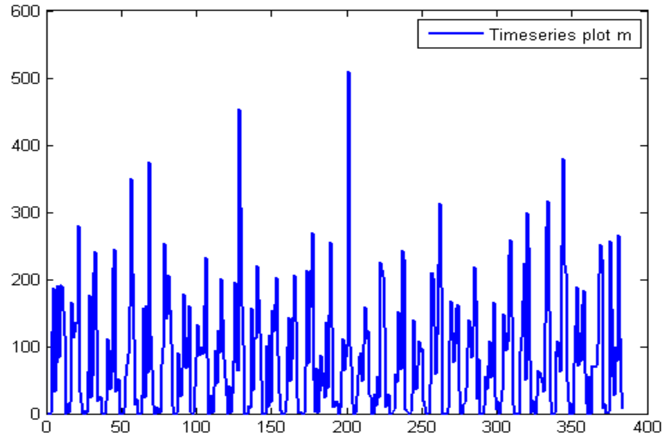
SUMMARY OF CASE STUDIES

Summary of Case studies

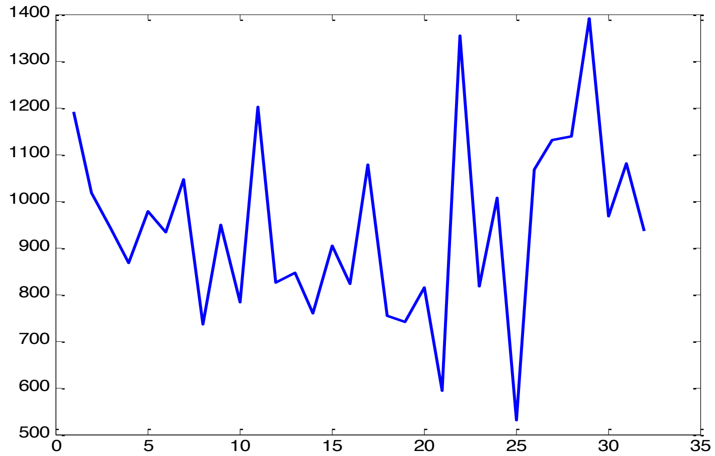
Case study-1:Time series plot



Daily rainfall data of Bangalore city



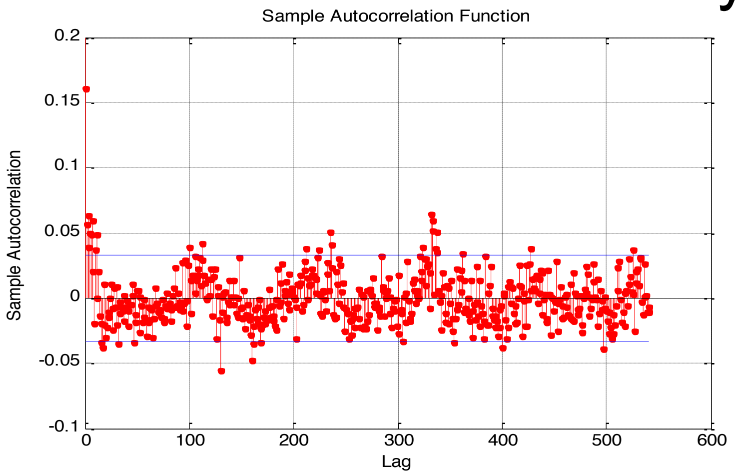
Monthly rainfall data of Bangalore city



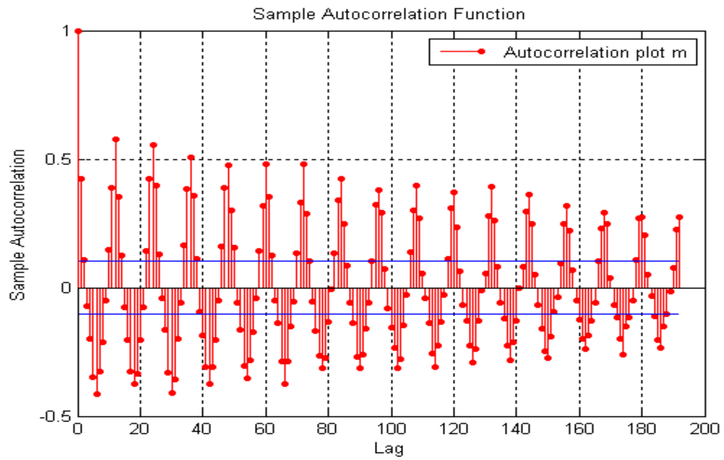
Yearly rainfall data of Bangalore city

Summary of Case studies

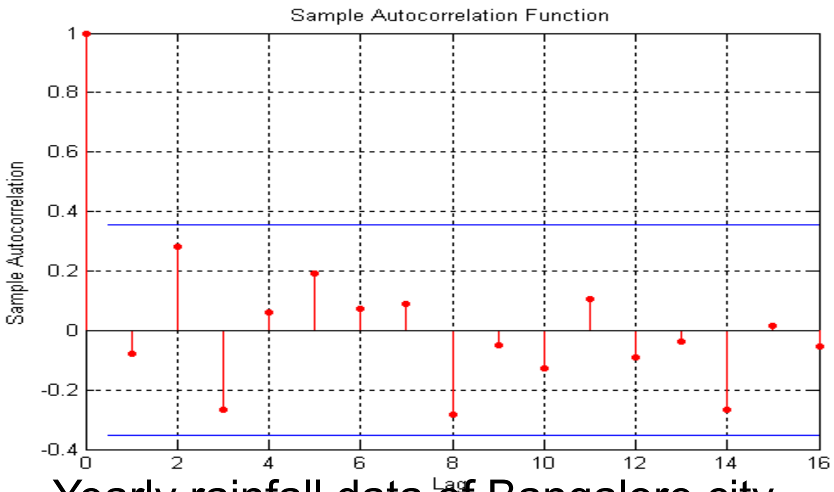
Case study-1: Correlogram



Daily rainfall data of Bangalore city



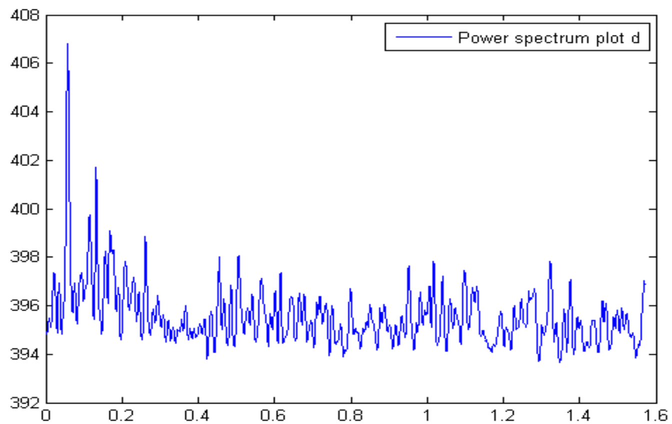
Monthly rainfall data of Bangalore city



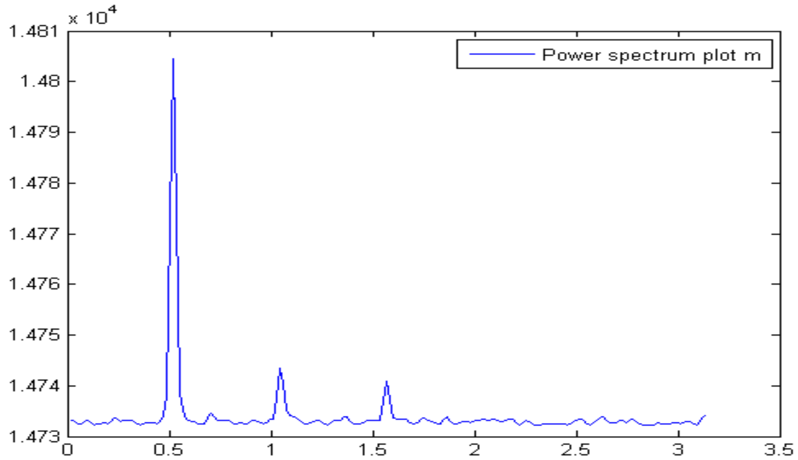
Yearly rainfall data of Bangalore city

Summary of Case studies

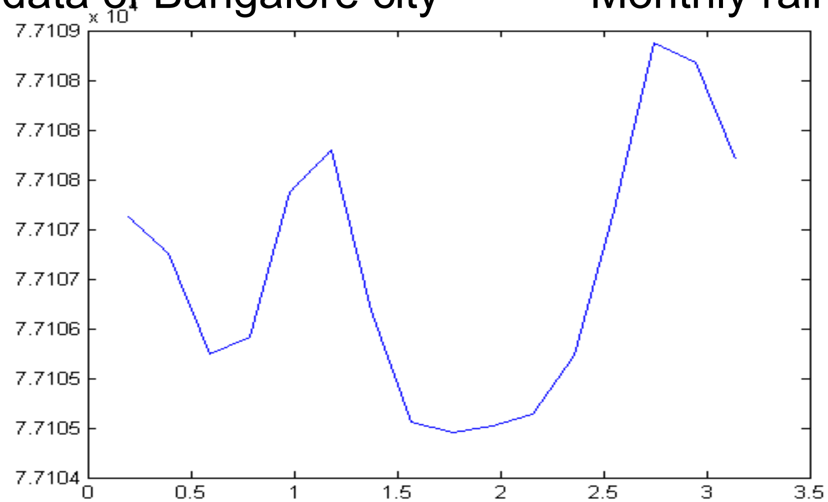
Case study-1: Power spectrum



Daily rainfall data of Bangalore city



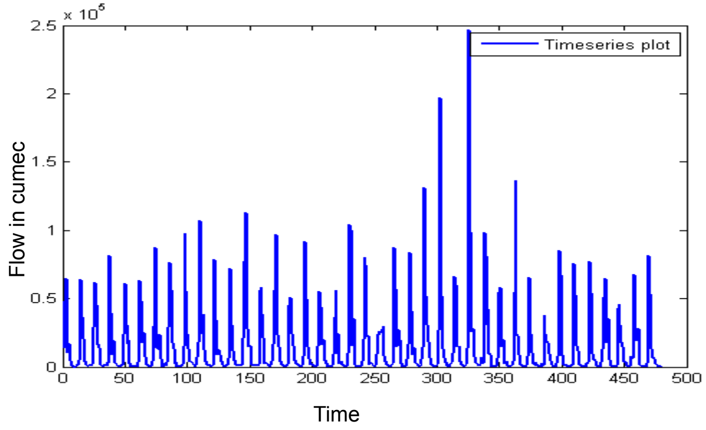
Monthly rainfall data of Bangalore city

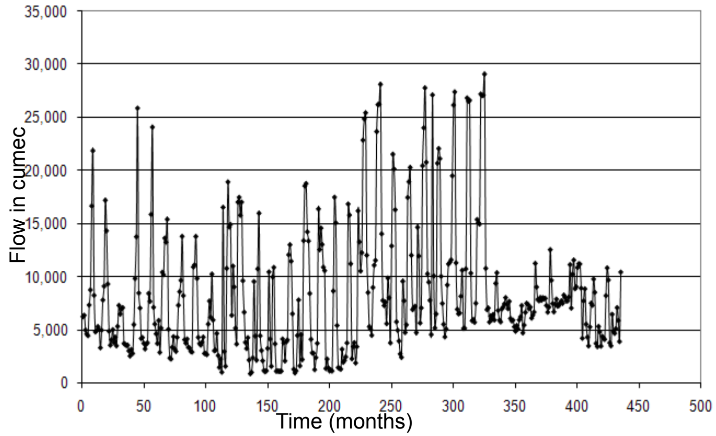


Yearly rainfall data of Bangalore city

Summary of Case studies

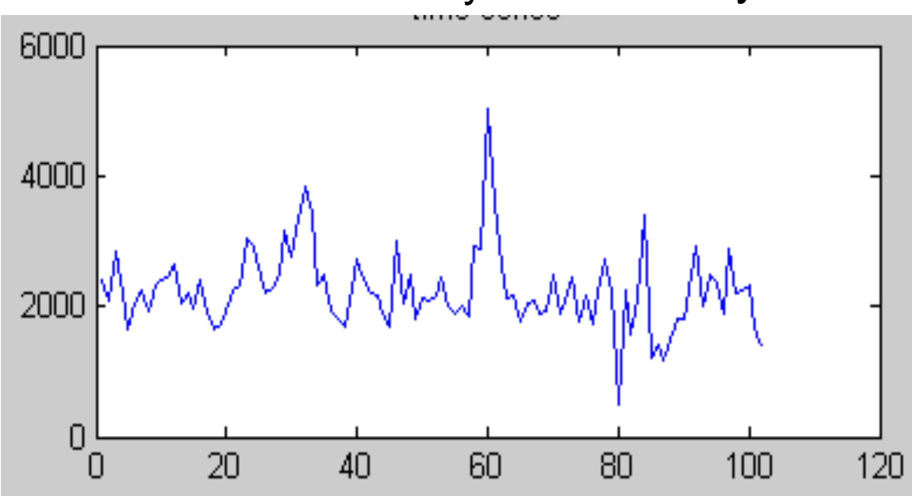
Time series plot





3. Monthly stream flow data for Cauvery

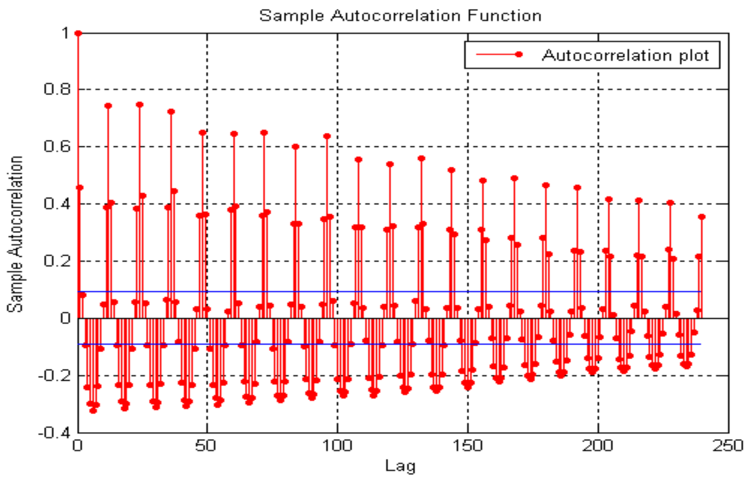
4. Monthly stream flow data of a river



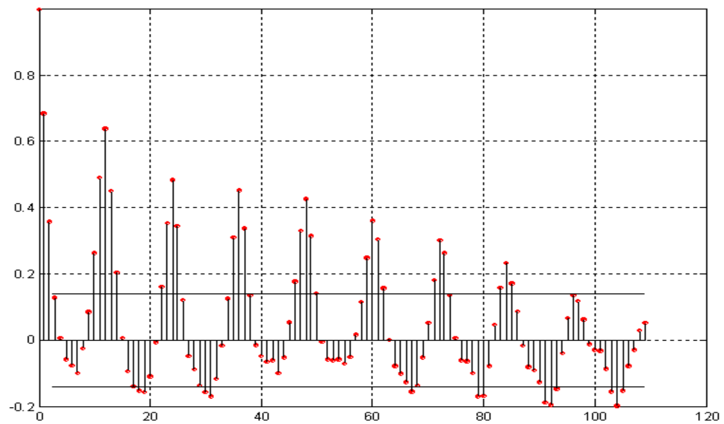
5. Sakleshpur Annual Rainfall Data

Summary of Case studies

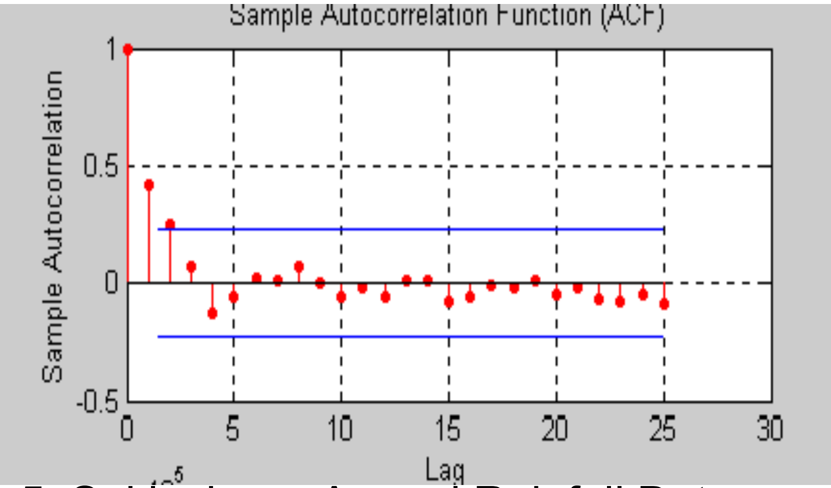
Correlogram



3. Monthly stream flow data for Cauvery



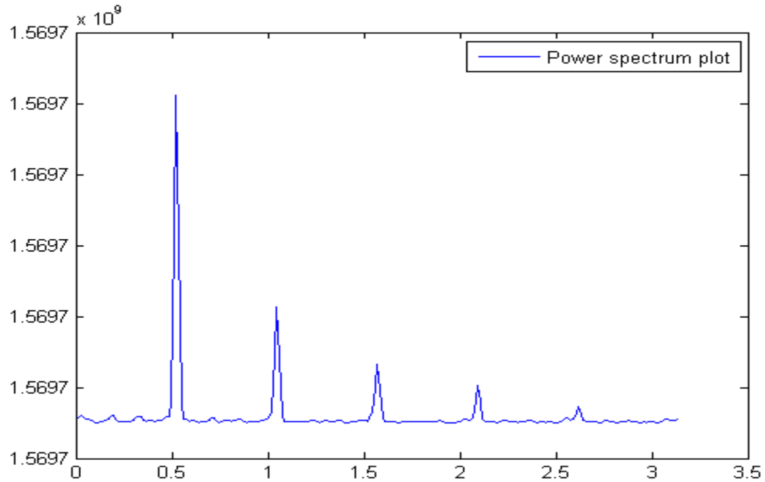
4. Monthly stream flow data of a river



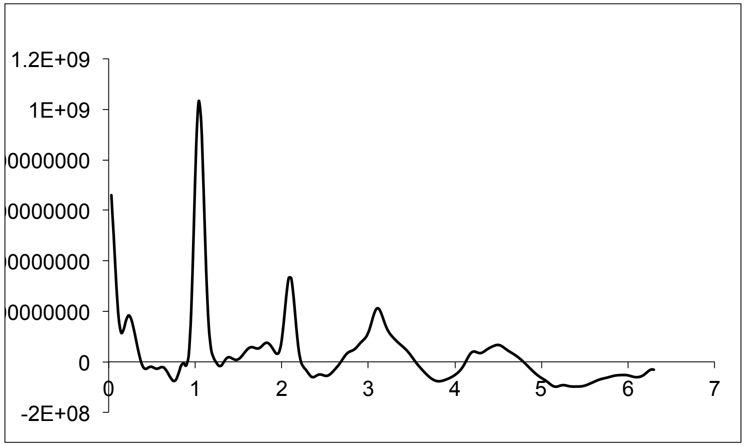
5. Sakleshpur Annual Rainfall Data

Summary of Case studies

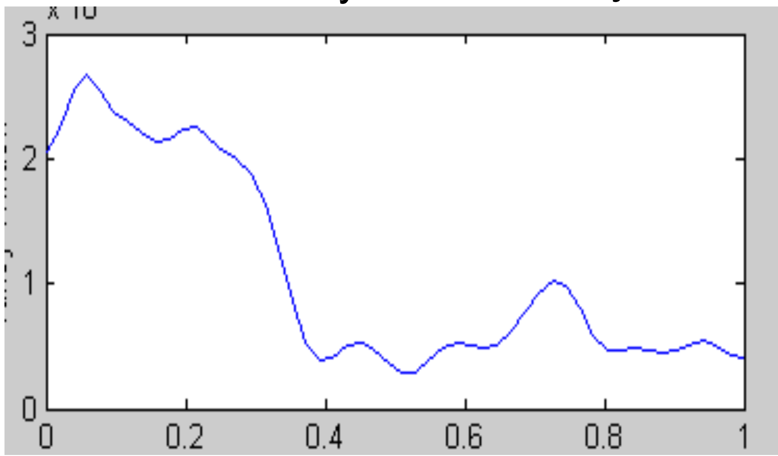
Power spectrum



3. Monthly stream flow data for Cauvery



4. Monthly stream flow data of a river



5. Sakleshpur Annual Rainfall Data

Summary of Case studies

ARMA Models

3. Monthly stream flow data for Cauvery

ARMA(4, 0) – For data generation

ARMA(1, 0) – For one step forecasting

USGS site.

4. Monthly stream flow data of a river

ARMA (8, 0) – For both data generation & one step forecasting

5. Sakleshpur Annual Rainfall Data

ARMA(5, 0) – For data generation

ARMA(1, 2) – For one step forecasting

MARKOV CHAINS

Markov Chains

- A Markov chain is a stochastic process with the property that value of process X_t at time t depends on its value at time t-1 and not on the sequence of other values (X_{t-2} , X_{t-3},……. X_0) that the process passed through in arriving at X_{t-1}.

$$P\left[X_t / X_{t-1}, X_{t-2}, \ldots\ldots X_0\right] = P\left[X_t / X_{t-1}\right]$$

Single step Markov chain

Markov Chains

$$P\left[X_t = a_j \middle/ X_{t-1} = a_i\right]$$

- This conditional probability gives the probability at time t will be in state ‘j’, given that the process was in state ‘i’ at time t-1.
- The conditional probability is independent of the states occupied prior to t-1.
- For example, if X_{t-1} is a dry day, we would be interested in the probability that X_t is a dry day or a wet day.
- This probability is commonly called as transition probability

Markov Chains

$$P\left[X_t = a_j / X_{t-1} = a_i\right] = P_{ij}^t$$

- Usually written as P_{ij}^t indicating the probability of a step from a_i to a_j at time 't'.
- If P_{ij} is independent of time, then the Markov chain is said to be homogeneous.

i.e., $P_{ij}^t = P_{ij}^{t+\tau}$ ∀ t and τ

the transition probabilities remain same across time

Markov Chains

Transition Probability Matrix(TPM):

$$
P = \begin{array}{c} \\ \\ \end{array}
\begin{array}{c|cccccc}
 & t+1 \rightarrow 1 & 2 & 3 & . & . & m \\
t \downarrow & & & & & & \\
1 & P_{11} & P_{12} & P_{13} & . & . & P_{1m} \\
2 & P_{21} & P_{22} & P_{23} & . & . & P_{2m} \\
3 & P_{31} & & & & & \\
. & . & & & & & \\
. & . & & & & & \\
m & P_{m1} & P_{m2} & & & & P_{mm}
\end{array}_{m \times m}
$$

Markov Chains

$$\sum_{j=1}^{m} P_{ij} = 1 \quad \forall \text{ i}$$

- Elements in any row of TPM sum to unity
- TPM can be estimated from observed data by enumerating the number of times the observed data went from state 'i' to 'j'
- $P_j^{(n)}$ is the probability of being in state 'j' in time step 'n'.

Markov Chains

- $p_j^{(0)}$ is the probability of being in state 'j' in period t = 0.

$$p^{(0)} = \begin{bmatrix} p_1^{(0)} & p_2^{(0)} & . & . & p_m^{(0)} \end{bmatrix}_{1\times m}$$ …. Probability vector at time 0

$$p^{(n)} = \begin{bmatrix} p_1^{(n)} & p_2^{(n)} & . & . & p_m^{(n)} \end{bmatrix}_{1\times m}$$ …. Probability vector at time 'n'

- If $p^{(0)}$ is given and TPM is given

$$p^{(1)} = p^{(0)} \times P$$

Markov Chains

$$p^{(1)} = \begin{bmatrix} p_1^{(0)} & p_2^{(0)} & . & . & p_m^{(0)} \end{bmatrix} \begin{bmatrix} P_{11} & P_{12} & P_{13} & . & . & P_{1m} \\ P_{21} & P_{22} & P_{23} & . & . & P_{2m} \\ P_{31} & & & & & \\ . & & & & & \\ P_{m1} & P_{m2} & & & & P_{mm} \end{bmatrix}$$

$$= p_1^{(0)} P_{11} + p_2^{(0)} P_{21} + ... + p_m^{(0)} P_{m1}$$ …. Probability of going to state 1

$$= p_1^{(0)} P_{12} + p_2^{(0)} P_{21} + ... + p_m^{(0)} P_{m2}$$ …. Probability of going to state 2

And so on…

Markov Chains

Therefore

$$p^{(1)} = \begin{bmatrix} p_1^{(1)} & p_2^{(1)} & . & . & p_m^{(1)} \end{bmatrix}_{1\times m}$$

$$\begin{aligned} p^{(2)} &= p^{(1)} \times P \\ &= p^{(0)} \times P \times P \\ &= p^{(0)} \times P^2 \end{aligned}$$

In general,

$$p^{(n)} = p^{(0)} \times P^n$$

Markov Chains

- As the process advances in time, $p_j^{(n)}$ becomes less dependent on $p^{(0)}$
- The probability of being in state 'j' after a large number of time steps becomes independent of the initial state of the process.
- The process reaches a steady state at large n

$$p = p \times P^n$$

- As the process reaches steady state, the probability vector remains constant

Example – 1

Consider the TPM for a 2-state first order homogeneous Markov chain as

$$TPM = \begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix}$$

State 1 is a non-rainy day and state 2 is a rainy day

Obtain the

1. probability of day 1 is non-rainfall day / day 0 is rainfall day
2. probability of day 2 is rainfall day / day 0 is non-rainfall day
3. probability of day 100 is rainfall day / day 0 is non-rainfall day

Example – 1 (contd.)

1. probability of day 1 is non-rainfall day / day 0 is rainfall day

$$TPM = \begin{array}{c} \\ \text{No rain} \\ \text{rain} \end{array} \begin{array}{c} \begin{array}{cc} \text{No rain} & \text{rain} \end{array} \\ \begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix} \end{array}$$

The probability is 0.4

2. probability of day 2 is rainfall day / day 0 is non-rainfall day

$$p^{(2)} = p^{(0)} \times P^2$$

Example – 1 (contd.)

$$p^{(2)} = \begin{bmatrix} 0.7 & 0.3 \end{bmatrix} \begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix}$$

$$= \begin{bmatrix} 0.61 & 0.39 \end{bmatrix}$$

The probability is 0.39

3. probability of day 100 is rainfall day / day 0 is non-rainfall day

$$p^{(n)} = p^{(0)} \times P^n$$

Example – 1 (contd.)

$$P^2 = P \times P$$

$$= \begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix} \begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix} = \begin{bmatrix} 0.61 & 0.39 \\ 0.52 & 0.48 \end{bmatrix}$$

$$P^4 = P^2 \times P^2 = \begin{bmatrix} 0.5749 & 0.4251 \\ 0.5668 & 0.4332 \end{bmatrix}$$

$$P^8 = P^4 \times P^4 = \begin{bmatrix} 0.5715 & 0.4285 \\ 0.5714 & 0.4286 \end{bmatrix}$$

$$P^{16} = P^8 \times P^8 = \begin{bmatrix} 0.5714 & 0.4286 \\ 0.5714 & 0.4286 \end{bmatrix}$$

Example – 1 (contd.)

Steady state probability

$$p = \begin{bmatrix} 0.5714 & 0.4286 \end{bmatrix}$$

For steady state,

$$p = p \times P^n$$

$$= \begin{bmatrix} 0.5714 & 0.4286 \end{bmatrix} \begin{bmatrix} 0.5714 & 0.4286 \\ 0.5714 & 0.4286 \end{bmatrix}$$

$$= \begin{bmatrix} 0.5714 & 0.4286 \end{bmatrix}$$