INDIAN INSTITUTE OF SCIENCE

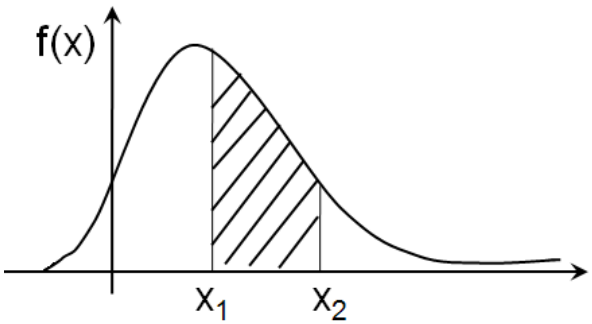
STOCHASTIC HYDROLOGY

Lecture -40

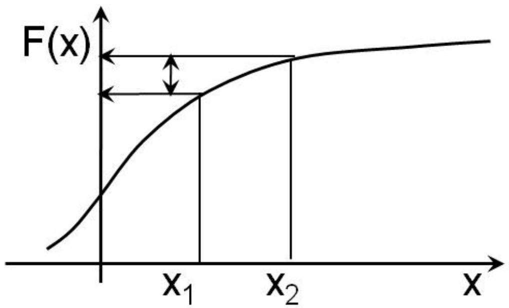
Course Instructor : Prof. P. P. MUJUMDAR
Department of Civil Engg., IISc.

Summary of the Course

- Concept of a random variable
- Discrete and continuous random variables
- Probability mass function, density function and cumulative distribution functions



PDF



CDF

$f(x) \geq 0$

$$\int_{-\infty}^{\infty} f(x) = 1$$

Summary of the Course (contd.)

- Bivariate distributions, Joint pmf and pdf
- Marginal density functions

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y)dxdy = 1$$

f(x, y) ≥ 0

$$g(x) = \int_{-\infty}^{y} f(x, y)dy$$

$$F(x) = \int_{-\infty}^{x} g(x)dx$$

- Conditional distributions

$$g(x/y) = \frac{f(x, y)}{h(y)} \quad h(y) > 0$$

Summary of the Course (contd.)

- Independent random variables

$$g\left(x/y\right) = \frac{f(x,y)}{h(y)} \quad \textbf{h(y) > 0}$$

$$g(x) \quad = \frac{f(x,y)}{h(y)}$$

$$f(x,y) = g(x).h(y)$$

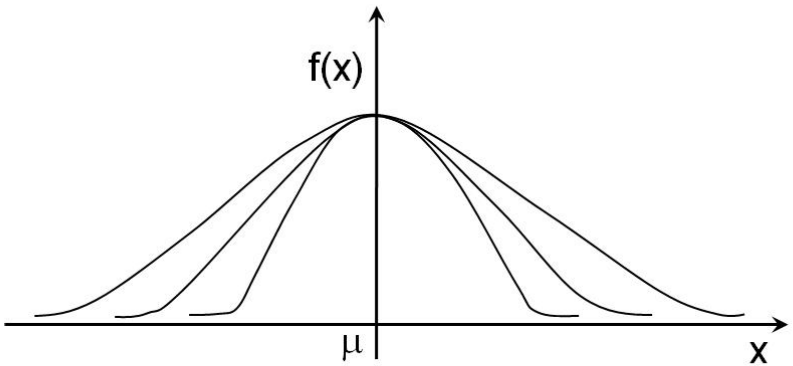
- Functions of random variables
- Moments of a distribution $\mu_n^0 = \int_{-\infty}^{\infty} x^n f(x)dx$
- Expected value

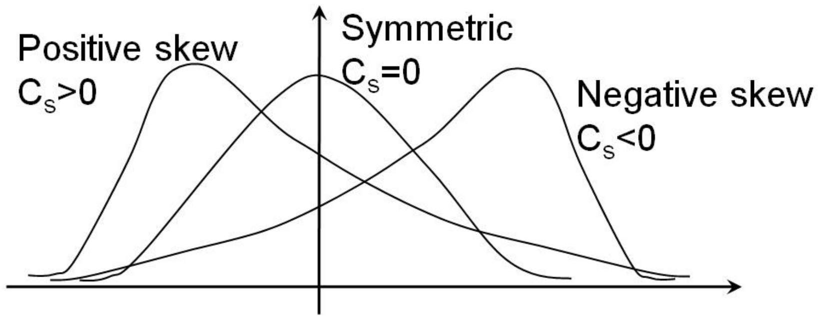
$$\mu = E(X) = \int_{-\infty}^{\infty} x f(x)dx$$

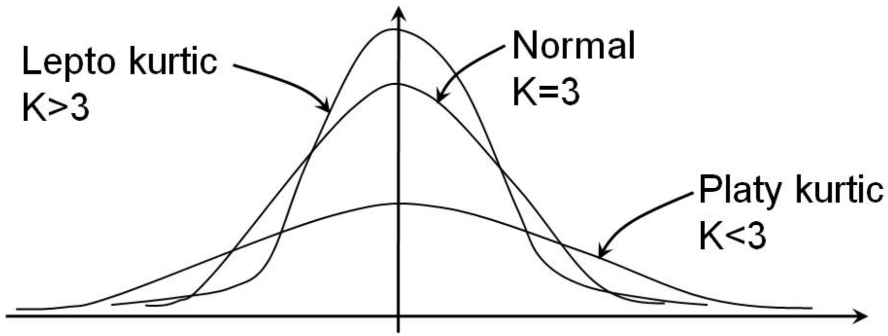
$$\mu_n = \int_{-\infty}^{\infty} \left(x-\mu\right)^n f(x)dx$$

Summary of the Course (contd.)

- Measures of central tendency, dispersion, symmetry and peakedness

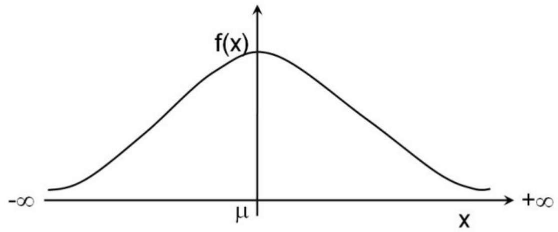






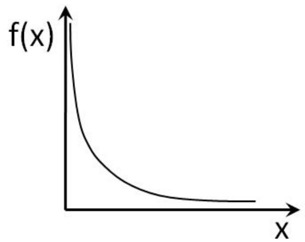
Summary of the Course (contd.)

- Normal distribution $Z = \frac{X - \mu}{\sigma}$



$$F(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-z^2/2} dz \qquad -\infty < z < +\infty$$

  - Central limit theorem
- Log-normal distribution
- Exponential Distribution
- Gamma Distribution



- Extreme Value Distributions
  - Extreme Value Type-I Distribution (Gumbel's Extreme Value Distribution)
  - Extreme Value Type-III Minimum Distribution (Weibull's Distribution)

Summary of the Course (contd.)

- Parameter estimation
  - Method of matching points
  - Method of moments

Equate the first 'm' moments of the population to the sample estimates of the first 'm' moments

Results in 'm' equations; solve to get the 'm' unknown parameters of the distribution

  - Method of maximum likelihood

$$L = f(x_1;\theta_1;\theta_2 \ldots \theta_m) \times f(x_2;\theta_1;\theta_2 \ldots \theta_m) \times f(x_n;\theta_1;\theta_2 \ldots \theta_m)$$

$$= \prod_{i=1}^{n} f\left(x_i, \theta_1, \ldots\ldots\ldots \theta_m\right)$$

$$\frac{\partial L}{\partial \theta_i} = 0 \quad \forall\ i$$

Summary of the Course (contd.)

- Jointly Distributed Random Variables
- Covariance

$$\mu_{1,1} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(x-\mu_x\right)\left(y-\mu_y\right)f(x,y)\,dx\,dy$$

$$= E\left[\left(x-\mu_x\right)\left(y-\mu_y\right)\right]$$

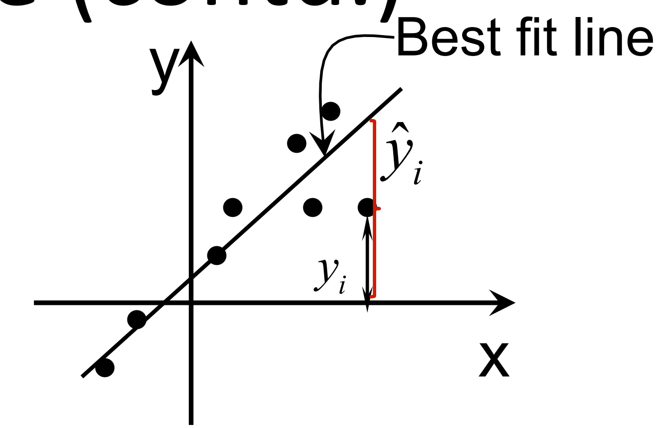
$$s_{X,Y} = \frac{\sum_{i=1}^{n}\left(x_i-\overline{x}\right)\left(y_i-\overline{y}\right)}{n-1}$$

- Correlation coefficient

$$r_{X,Y} = \frac{s_{X,Y}}{s_X s_Y}$$

Summary of the Course (contd.)

Simple Linear Regression



(x_i, y_i) are observed values

$\hat{y}_i$ is predicted value of x_i

$$\hat{y}_i = a + bx_i$$

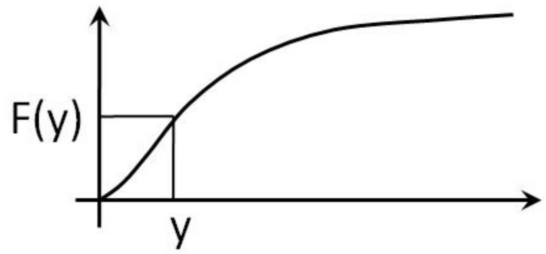
Error, $e_i = y_i - \hat{y}_i$

Estimate the parameters a, b such that the square error is minimum

Summary of the Course (contd.)

- Data Generation

Randomly picked up values of F (y) follow a uniform distribution u(0, 1)



$$F(y) = \int_{-\infty}^{y} f(y)dy$$

$$F(y) = R_u = \int_{-\infty}^{y} f(y)dy$$

Gamma distribution

$$y = \frac{-\sum_{i=1}^{\eta} \ln R_{u_i}}{\lambda}$$

(for integer values of η)

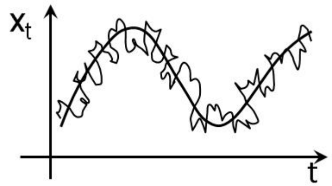
Normal distribution: Given R_N, data is generated by

$$y = \sigma R_N + \mu$$

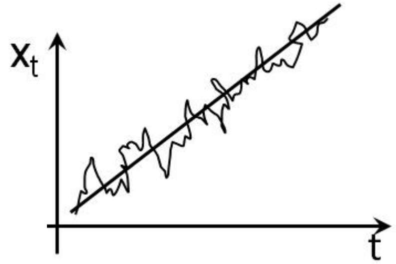
Summary of the Course (contd.)

$$X_t = d_t + \varepsilon_t$$

- Time series analysis
  - Realization ; Ensemble
  - Stationarity
  - Auto covariance Function
  - Auto corrleation and correlogram

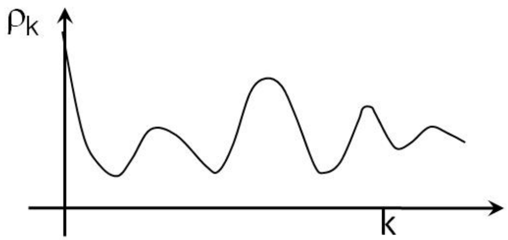


Stochastic + Periodic

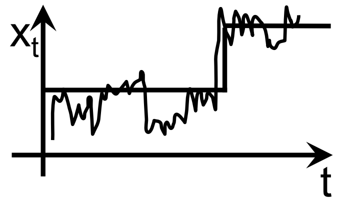


Stochastic + Trend

Stochastic



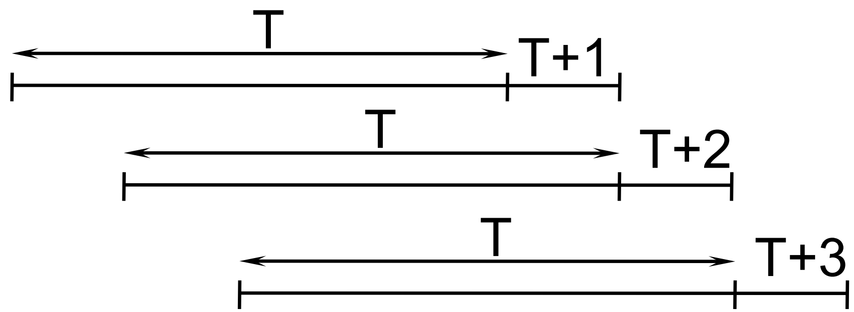
Correlogram



Stochastic + Jump

Summary of the Course (contd.)

- Data Extension & Forecasting
  - Moving average
  - Double moving average



- Data Generation – Uncorrelated Data

Summary of the Course (contd.)

- Data Generation – Serially Correlated Data
  - First order Markov Model
    - Annual flow generation

$$X_{t+1} = \mu_x + \rho_1 \left(X_t - \mu_x \right) + u_{t+1} \sigma_x \sqrt{1 - \rho_1^2}$$

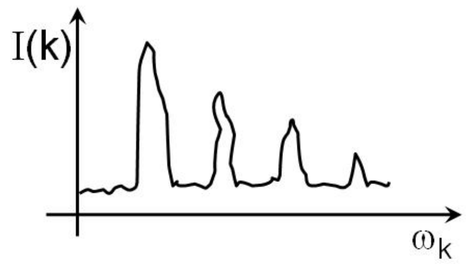
  - First order Markov model with non-stationarity
    - Thoma Fiering model for monthly and seasonal flow generation

$$X_{i,j+1} = \mu_{j+1} + \rho_j \frac{\sigma_{j+1}}{\sigma_j} \left(X_{ij} - \mu_j \right) + t_{i,j+1} \sigma_{j+1} \sqrt{1 - \rho_j^2}$$

Summary of the Course (contd.)

- Frequency domain analysis
  - Spectral density
  - Test for significance of periodicities
  - Removing periodicities
  - Standardizing the data

$$Z_t = \frac{\left(X_t - \bar{X}_i\right)}{S_i}$$



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

$$\omega_k = \frac{2\pi k}{N}$$

Summary of the Course (contd.)

- ARIMA models

$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \ldots + \phi_p X_{t-p} + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \ldots + \theta_q e_{t-q} + e_t$

$\{e_t\}$ is the residual series
Assumptions : $\{e_t\}$ has zero mean with uncorrelated terms

$$\hat{X}_{t+1} = \sum_{j=1}^{p} \phi_j X_{t-j} + \sum_{j=1}^{q} \theta_j e_{t-j}$$

- Partial Auto Correlation function

Calculation of Partial Auto Correlations:
(Yule Walker equations)

ϕ_p is the PAC of order p

$$P_p * \phi_p = \rho_p$$

Summary of the Course (contd.)

Box Jenkins Time series models

Differencing: $Y_t = X_t' = X_t - X_{t-1}$

X_t' is First order differencing

$$X_t'' = X_t' - X_{t-1}'$$

X_t'' is Second order differencing

Operator 'B':

The effect of operator 'B' is to shift the argument to that one step behind.

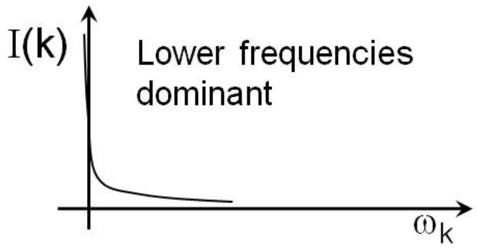
$BX_t = X_{t-1}$

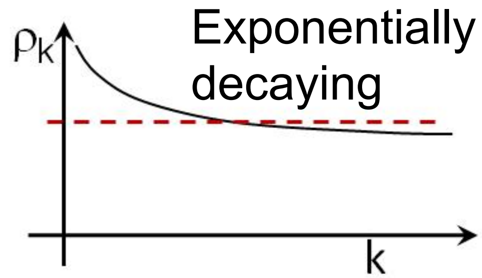
$BX_{t-1} = X_{t-2}$

Summary of the Course (contd.)

- Behavior of AR and MA process

  AR(2) process



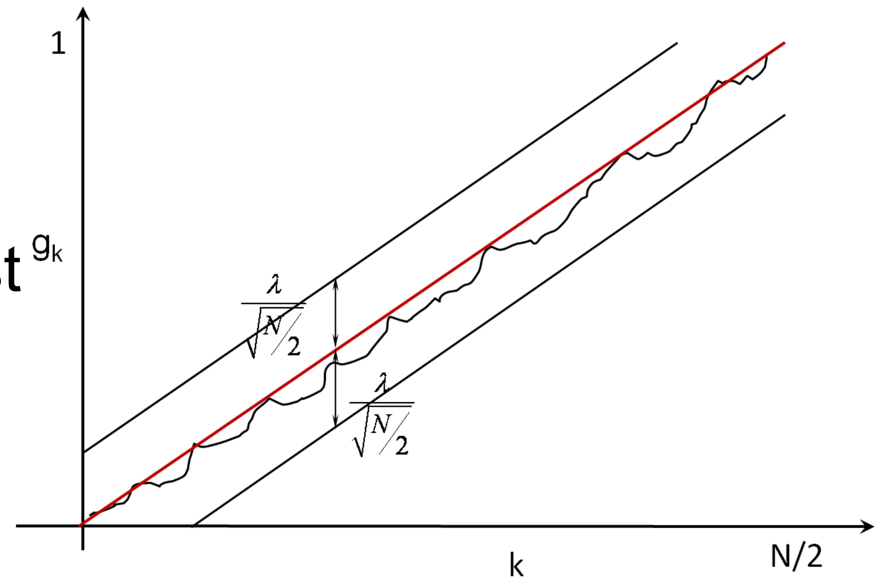


- Parameter estimation
  - Marquadt's algorithm
  - Matlab function "armax" m = armax(data, orders)
- Model selection
  - Maximum likelihood rule
  - Mean square error

Summary of the Course (contd.)

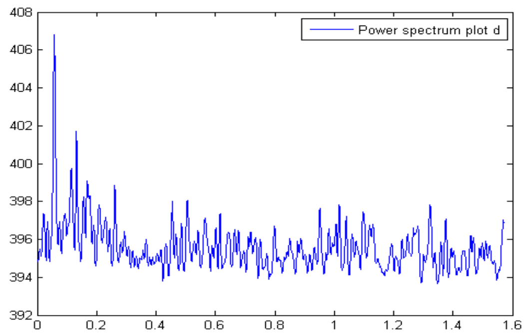
- Model testing/validation
  - Significance of residual mean
  - Significance of periodicities
    - Cumulative periodogram test or Bartlett's test
  - White noise test
    - Whittle's test
    - Portmanteau test

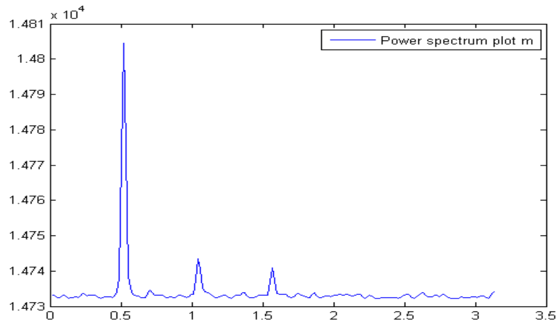
$$\eta(e) = \frac{N^{1/2}\overline{e}}{\hat{\rho}^{1/2}}$$

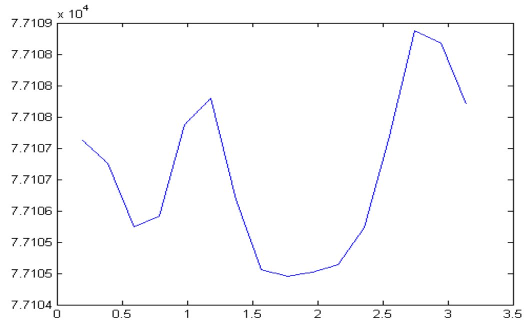


Summary of the Course (contd.)

- Use of ARMA models for data generation and one-time step ahead forecasting
- Case studies 1 and 2
  - Daily, monthly and annual rainfall
  - Annual stramflow
  - Monthly streamflow

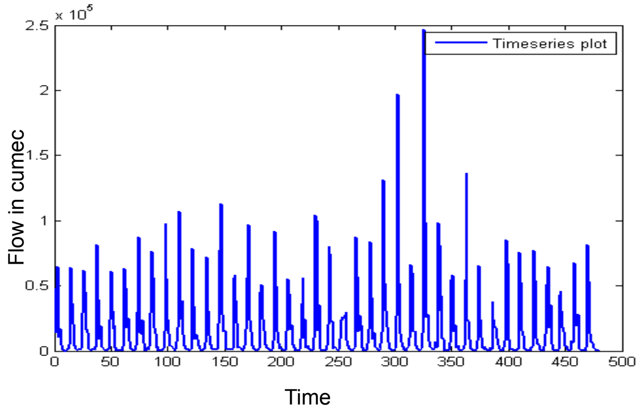




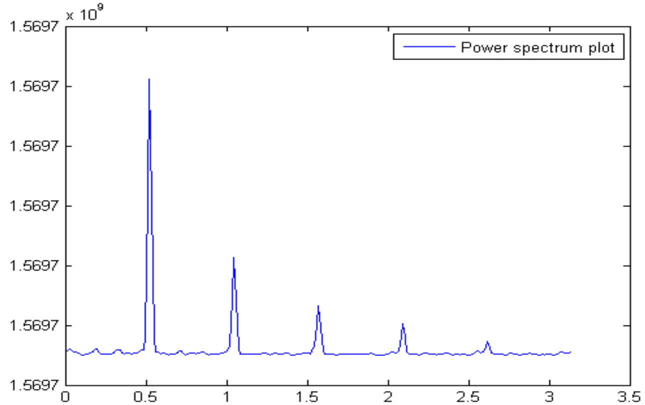


Summary of the Course (contd.)

– Case study -3: Monthly streamflows at KRS reservoir
  - Plots of Time series, Correlogram, Partial Autocorrelation function and Power spectrum
  - Standardization to remove periodicities
  - Candidate ARMA models : (contiguous and non-contiguous)
    – Log Likelihood
    – Mean square error



Monthly stream flow data



Power spectrum

Summary of the Course (contd.)

Markov chain

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

Transition Probability Matrix(TPM):

$$\sum_{j=1}^{m} P_{ij} = 1 \quad \forall \text{ i} \qquad\qquad \hat{P}_{ij} = \frac{n_{ij}}{\sum_{j=1}^{m} n_{ij}}$$

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

– Steady state Markov chains

Summary of the Course (contd.)

- Introduction to frequency analysis
  - Extreme events
- Recurrence interval
- Return period
  - Expected value of recurrence interval
- $P[X \geq x_T] = p = 1/T$
- Probability that a T year return period event will occur at least once in N years is

$$1-\left(1-\frac{1}{T}\right)^{N}$$

Summary of the Course (contd.)

- Frequency factors $x_T = \bar{x} + K_T s$
  - Normal distribution
  - Gumbel's Extreme Value Type-I distribution

$$K_T = -\frac{\sqrt{6}}{\pi}\left\{0.5772 + \ln\left[\ln\left(\frac{T}{T-1}\right)\right]\right\}$$

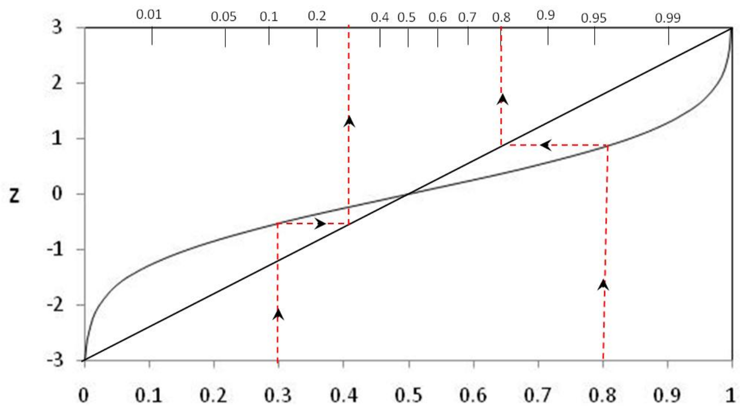
  - Log Pearson Type-III distribution

- Probability plotting

Method to check whether a probability distribution fits a set of data or not.

Summary of the Course (contd.)

- Probability paper construction
  - Mathematical construction
  - Graphical construction
    - Normal distribution



- Plotting position
  - Weibull's formula

$$p\left(X \geq x_m\right) = \frac{m}{n+1}$$

Summary of the Course (contd.)

- Goodness of fit
  - Chi-square test

$$\chi^2_{data} = \sum_{i=1}^{k} \frac{\left(N_i - E_i\right)^2}{E_i} \qquad \chi^2_{data} < \chi^2_{1-\alpha,k-p-1}$$

k - no. of class intervals
p – no. of parameters

  - Kolmogorov-Smirnov test

$$\Delta = maximum \left|P\left(x_i\right) - F\left(x_i\right)\right|$$

$\Delta < \Delta_o$

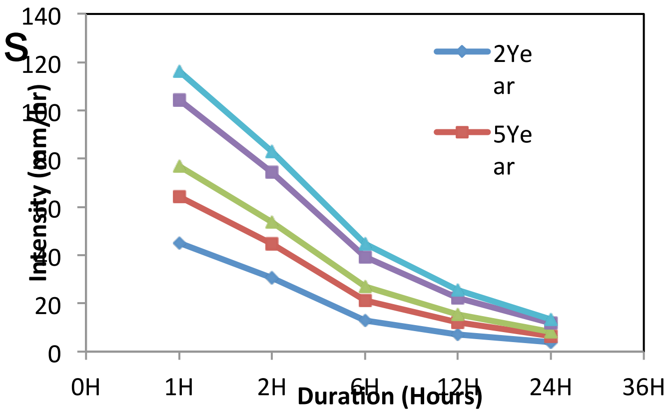
Summary of the Course (contd.)

- IDF relationship
  - Procedure for creating IDF curves

Step 1: Preparation of annual maximum data series
Step 2: Fitting the probability distribution
Step 3: Determining the rainfall depths

$$x_T = \overline{x} + K_T s$$

$$x_T = F^{-1}\left(\frac{T-1}{T}\right)$$



  - Empirical equations for IDF relationships

Summary of the Course (contd.)

- Multiple linear regression

$y = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \ldots\ldots.. + \beta_p x_p$

y is dependent variable,
$x_1, x_2, x_3, \ldots\ldots, x_p$ are independent variables and
$\beta_1, \beta_2, \beta_3, \ldots\ldots, \beta_p$ are unknown parameters

$$y_i = \sum_{j=1}^{p} \beta_j x_{i,j} \qquad Y_{(n\times 1)} = X_{(n\times p)} \times \mathrm{B}_{(p\times 1)}$$

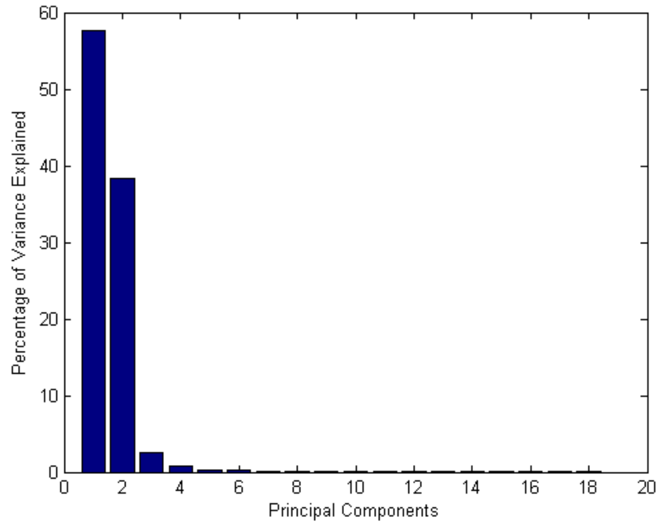
Summary of the Course (contd.)

Principal Component Analysis

PCA is a way of identifying patterns in the data

$$Z = X\,A$$

X is nxp matrix of n observations on p variables
Z is nxp matrix of n values for each of p components
A is pxp matrix of coefficients defining the linear transformation



Summary of the Course (contd.)

- Multivariate stochastic models
  - Cross correlation

$$r_{j,h}(k) = \frac{\sum_{i=1}^{n}\left(x_{j,i} - \overline{x}_j\right)\left(x_{h,i+k} - \overline{x}_h\right)}{(n-k)s_j s_h}$$

  - Two site Markov model

$$X_{h,t} = \overline{x}_h + r_{j,h}(0)\frac{s_h}{s_j}\left(X_{j,t} - \overline{x}_j\right) + u_t s_h \sqrt{1 - r_{j,h}^2(0)}$$

Summary of the Course (contd.)

- Multivariate stochastic models
  - Multisite Markov model

$$X_{t+1} = EX_t + G\varepsilon$$

where

X_t is a p x 1 vector of standardized values of the variable generated at time t,

E is a p x p diagonal matrix whose jth diagonal element is $\rho_j(1)$,

G is a p x p diagonal matrix whose jth diagonal element is $\sqrt{1-\rho_j^2(1)}$

ε is a p x 1 vector of random variates

Summary of the Course (contd.)

- Multivariate stochastic models
  - Matalas model

$$X_{t+1} = AX_t + B\varepsilon_{t+1}$$

where

X_t and X_{t+1} are p x 1 vectors representing standardized data corresponding to p sites at time steps t and $t+1$ resp

ε_{t+1} is N(0,1); p x 1 vector with ε_{t+1} independent of X_t.

A and B are coefficient matrices of size p x p. B is assumed to be lower triangular matrix

$$A = M_1 M_0^{-1}$$

$$BB' = M_0 - M_1 M_0^{-1} M_1'$$

M_0 is the cross-correlation matrix (size pxp) of lag zero

M_1 is the cross-correlation matrix (size pxp) of lag one

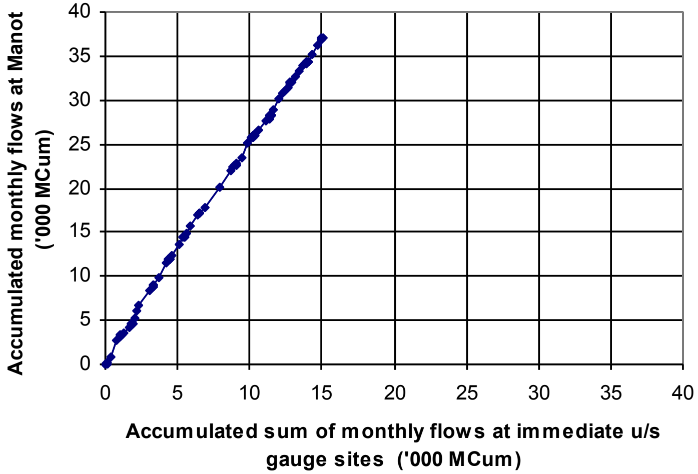
Summary of the Course (contd.)

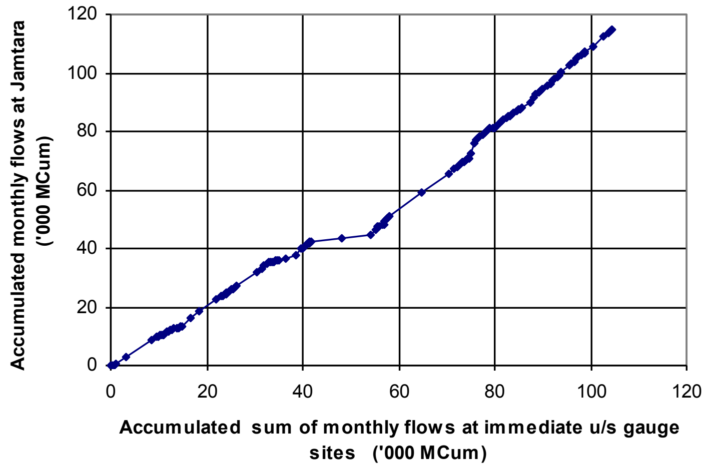
- Data consistency checks
  - Filling the missing data for an example basin
    - Monsoon period
    - Non-monsoon period
  - Statistical analysis of data
  - Specific flow

    The specific flow is expressed as flow volume per unit area of the catchment

Summary of the Course (contd.)

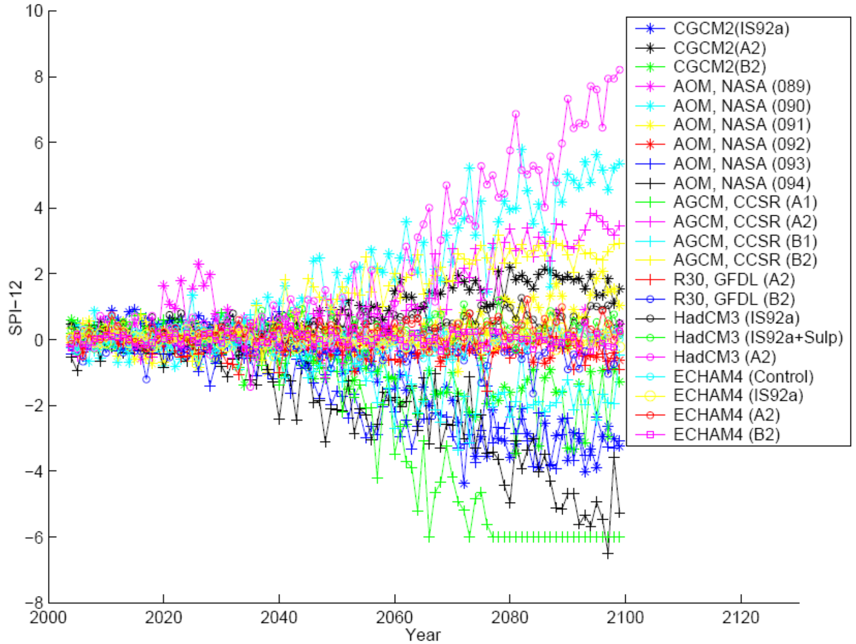
- Data consistency checks
  - Double Mass Curve





Summary of the Course (contd.)

- Data representation through box plots
- Normalisation of flow data
- Assessment of Climate Change Impacts
- Downscaling



Acknowledgment

The Instructor is grateful to:

- NPTEL for creating this opportunity and facilitating the lectures.
- Dr. Subimal Ghosh – for keeping excellent notes of the subject, when he credited the course as a ME student at IISc Bangalore (currently at IIT Bombay, as a faculty member; http://www.civil.iitb.ac.in/~subimal)
- Ms. Chandra Rupa, project staff, IISc., for her invaluable help in preparing the course material, solutions of numerical problems and preparation of presentation slides.
- All my students at IISc, Bangalore – current and past.