INDIAN INSTITUTE OF SCIENCE

STOCHASTIC HYDROLOGY

Lecture -14

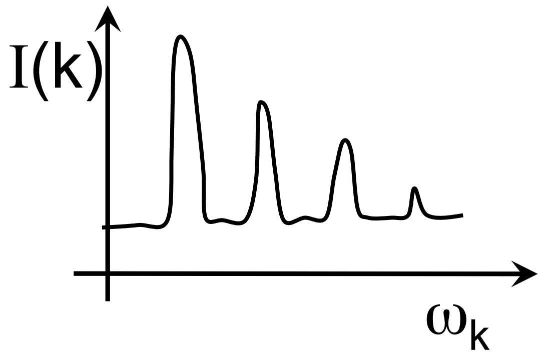
Course Instructor : Prof. P. P. MUJUMDAR
Department of Civil Engg., IISc.

Summary of the previous lecture

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

Frequency Domain Analysis

- Spectral density (I_k) is the amount of variance per interval of frequency
- Spectral analysis helps indentify the significant periodicities themselves
- A peak in the spectrum indicates an important contribution to variance at frequencies close to the peak
- Prominent spikes indicate periodicity
- Line spectrum - inconsistent estimate
- Power spectrum - consistent estimate



Example – 1
(Spectral Analysis)

Monthly Stream flow (in cumec) statistics(1979-2008) for a river is selected for the study. (Part data shown below)

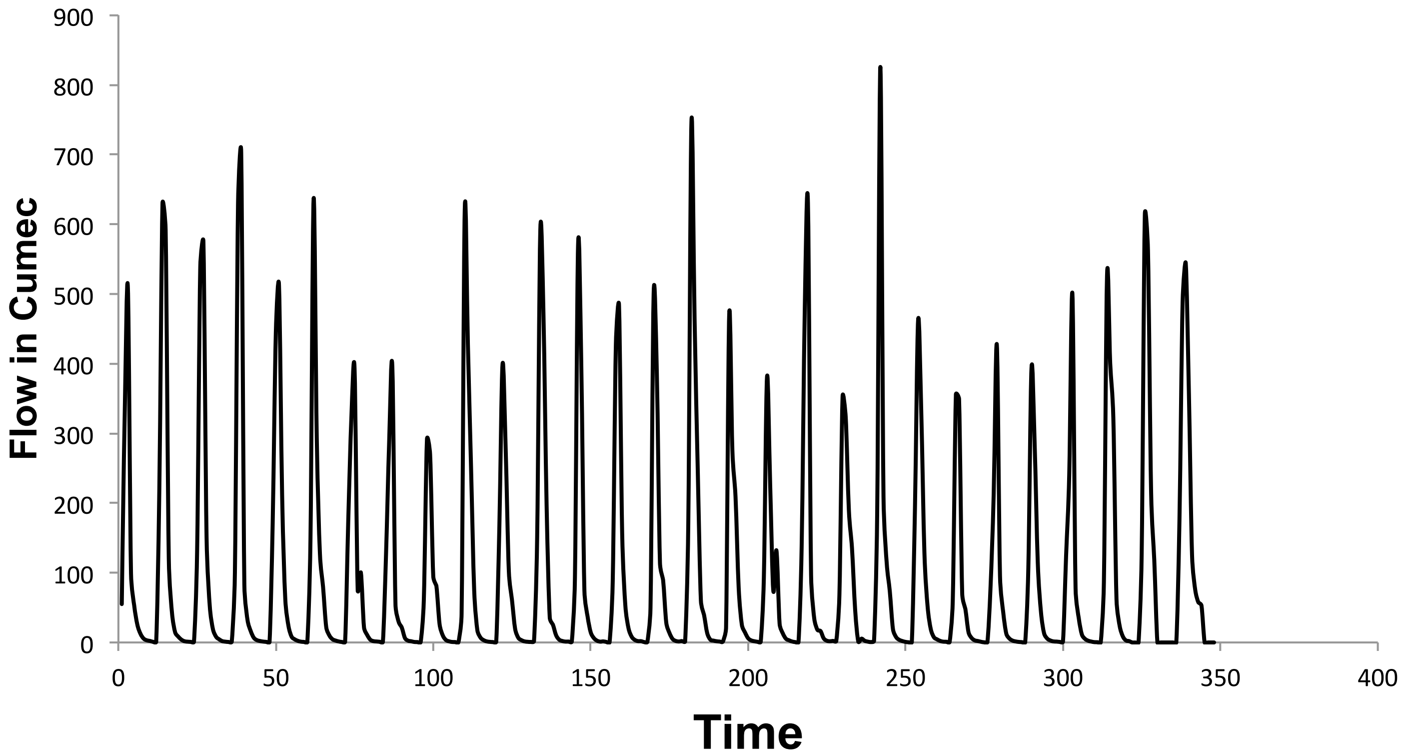
Year	Month	S.No.	Flow
1979	June	1	54.6
	July	2	325.4
	August	3	509.5
	September	4	99.4
	October	5	53.5
	November	6	25.8
	December	7	12.5
1980	January	8	5.6
	February	9	3.1
	March	10	2.2
	April	11	0.9
	May	12	0.81

N = 348

Example – 1 (contd.)

(Spectral Analysis)

The time series plot



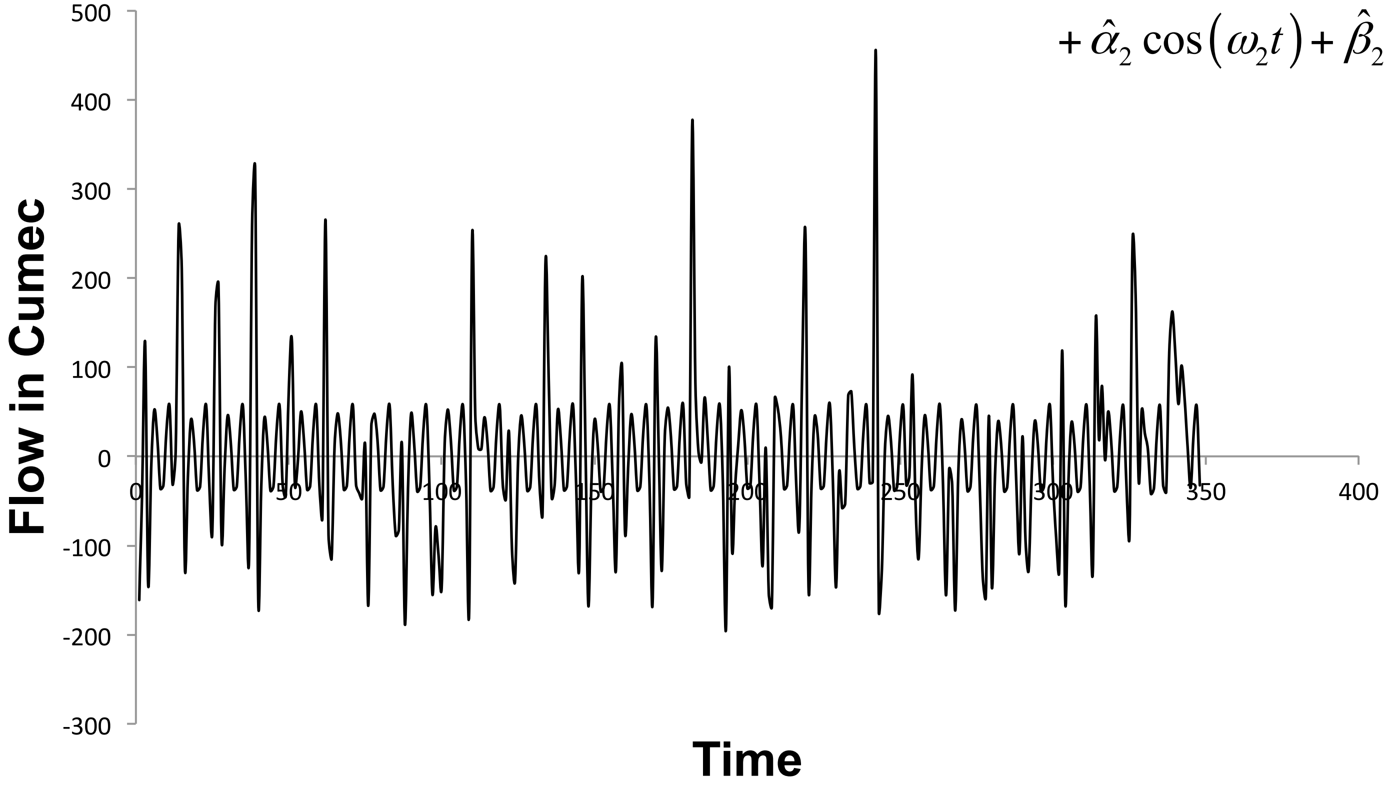
Example – 1 (contd.)

(Spectral Analysis)

$Z_t = X_t - Y_t$

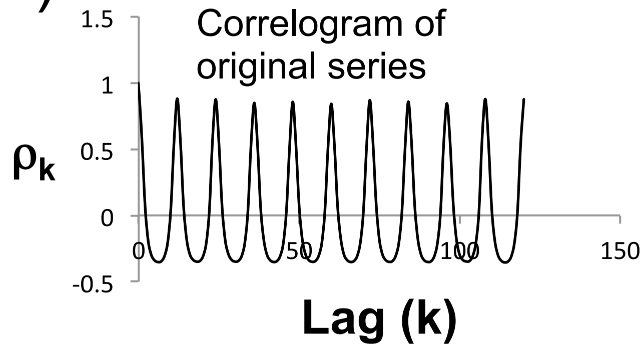
$$Y_t = \mu + \hat{\alpha}_1 \cos(\omega_1 t) + \hat{\beta}_1 \sin(\omega_1 t) + \hat{\alpha}_2 \cos(\omega_2 t) + \hat{\beta}_2 \sin(\omega_2 t)$$

Time series plot of Z_t,

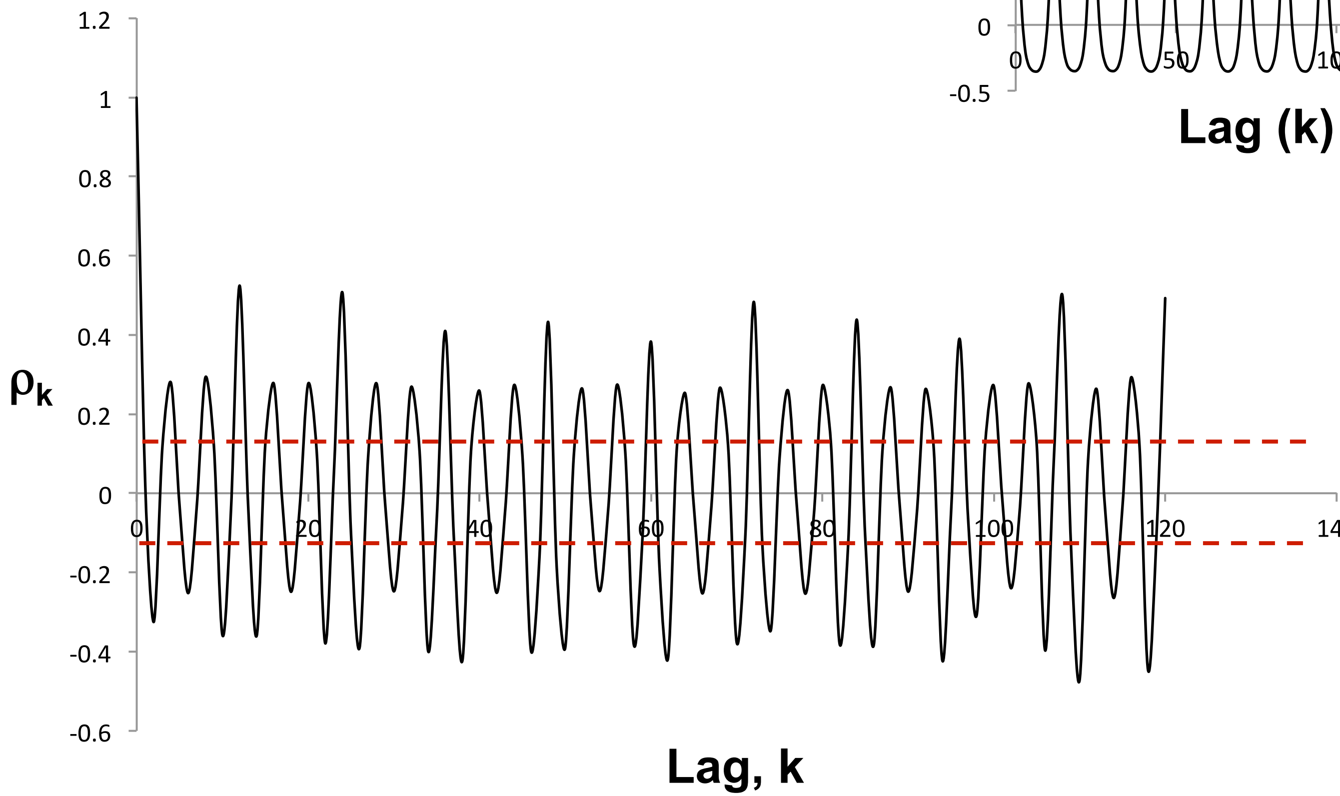


Example – 1 (contd.)

(Spectral Analysis)



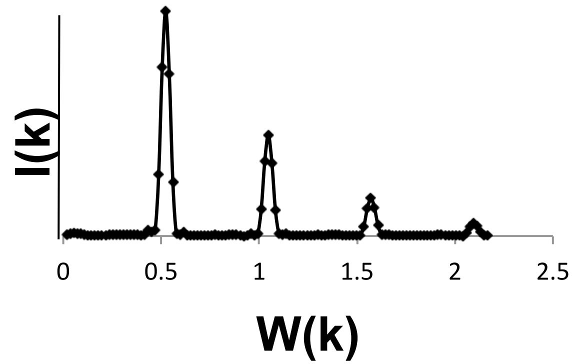
Correlogram of Z_t,



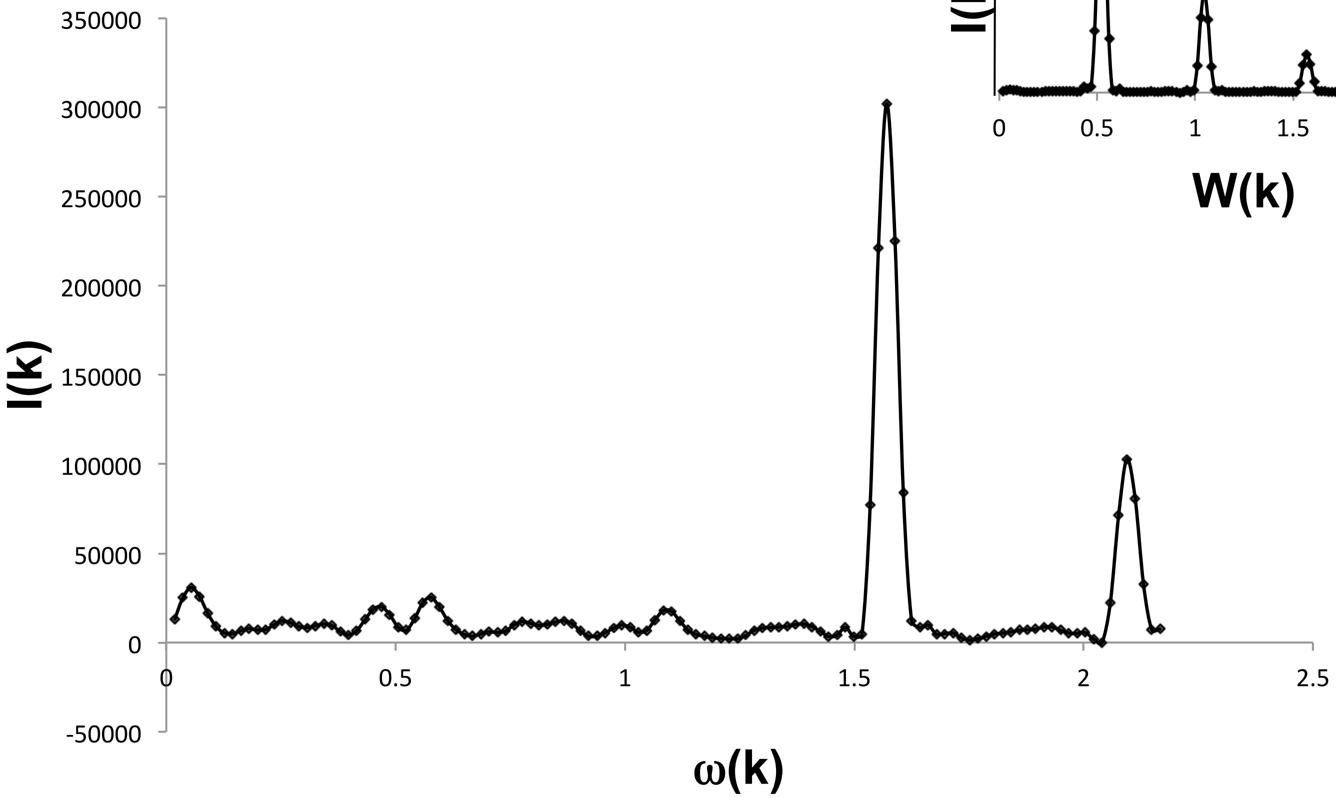
Example – 1 (contd.)

(Spectral Analysis)

Power Spectrum of original series



Power Spectrum of Z_t,



Example – 1 (contd.)

(Spectral Analysis)

- Significance test:

$$\mathrm{I} = \frac{\gamma^2 (N-2)}{4\hat{\rho}_1}$$

Where $\gamma^2 = \alpha^2 + \beta^2$ and

$$\hat{\rho}_1 = \frac{1}{N}\left[\sum_{t=1}^{N}\left\{x_t - \hat{\alpha}\cos(\omega_k t) - \hat{\beta}\sin(\omega_k t)\right\}\right]$$

For first peak, $\omega_1 = 0.5236$, $\alpha_1 = 29.28$, $\beta_1 = 172.93$

Therefore $\gamma^2 = 29.28^2 + 172.93^2$

$= 30762$

Example – 1 (contd.)

(Spectral Analysis)

$$\hat{\rho}_1 = \frac{1}{N}\left[\sum_{t=1}^{N}\left\{x_t - \alpha_1\cos\left(\omega_1 t\right) - \beta_1\sin\left(\omega_1 t\right)\right\}\right]$$

$$= \frac{1}{348}\times 36810.56$$

$$= 105.78$$

$$\mathrm{I} = \frac{\gamma^2\left(N-2\right)}{4\hat{\rho}_1} = \frac{30762\left(348-2\right)}{4\times 105.78} = 25155$$

From ‘F’ distribution table at 95% significance level,

F(2, 346) = 3.0

Example – 1 (contd.)

(Spectral Analysis)

$$\mathbf{I} \quad > F(2,346)$$

Therefore the periodicity is significant.

The values for other periodicities are as follows

ω_k	Statistic	F(2, N-2)
0.5236	25154	3.0
1.0472	11242	3.0
1.5708	4104	3.0
2.0944	1295	3.0

Example – 1 (contd.)

(Spectral Analysis)

- The periodicities from the time series is removed by transforming the series into a standardized one.
- The series $\{X_t\}$ is expressed as the new series $\{Z'_t\}$ where,

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

The mean and standard deviation for each month is tabulated.

Month	Mean	Stdev.
Jun	117.49	52.24
Jul	474.50	150.18
Aug	421.39	126.53
Sep	145.94	77.65
Oct	66.61	30.67
Nov	22.99	13.26
Dec	10.30	9.82
Jan	5.55	9.16
Feb	1.91	0.74
Mar	1.09	0.54
Apr	0.76	0.51
May	0.80	0.60

Example – 1 (contd.)

(Spectral Analysis)

$$Z'_1 = \frac{(54.6 - 117.49)}{52.24} = -1.204 \quad \text{(June)}$$

$$Z'_2 = \frac{(325.4 - 474.5)}{150.18} = -0.993 \quad \text{(July)}$$

$$Z'_3 = \frac{(509.5 - 421.39)}{126.53} = 0.696 \quad \text{(August)}$$

And so on.

Example – 1 (contd.)

(Spectral Analysis)

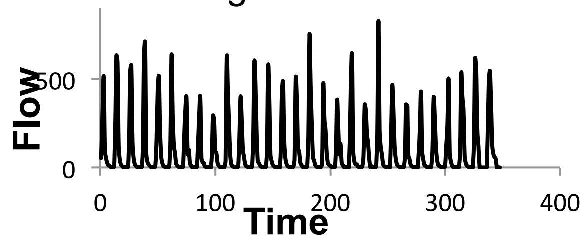
- Series of Z'_t (part data shown)

Year	Month	S.No.	X_t	Z'_t
1979	June	1	54.6	-1.204
	July	2	325.4	-0.993
	August	3	509.5	0.696
	September	4	99.4	-0.599
	October	5	53.5	-0.428
	November	6	25.8	0.212
	December	7	12.5	0.224
1980	January	8	5.6	0.006
	February	9	3.1	1.609
	March	10	2.2	2.063
	April	11	0.9	0.272
	May	12	0.81	0.019

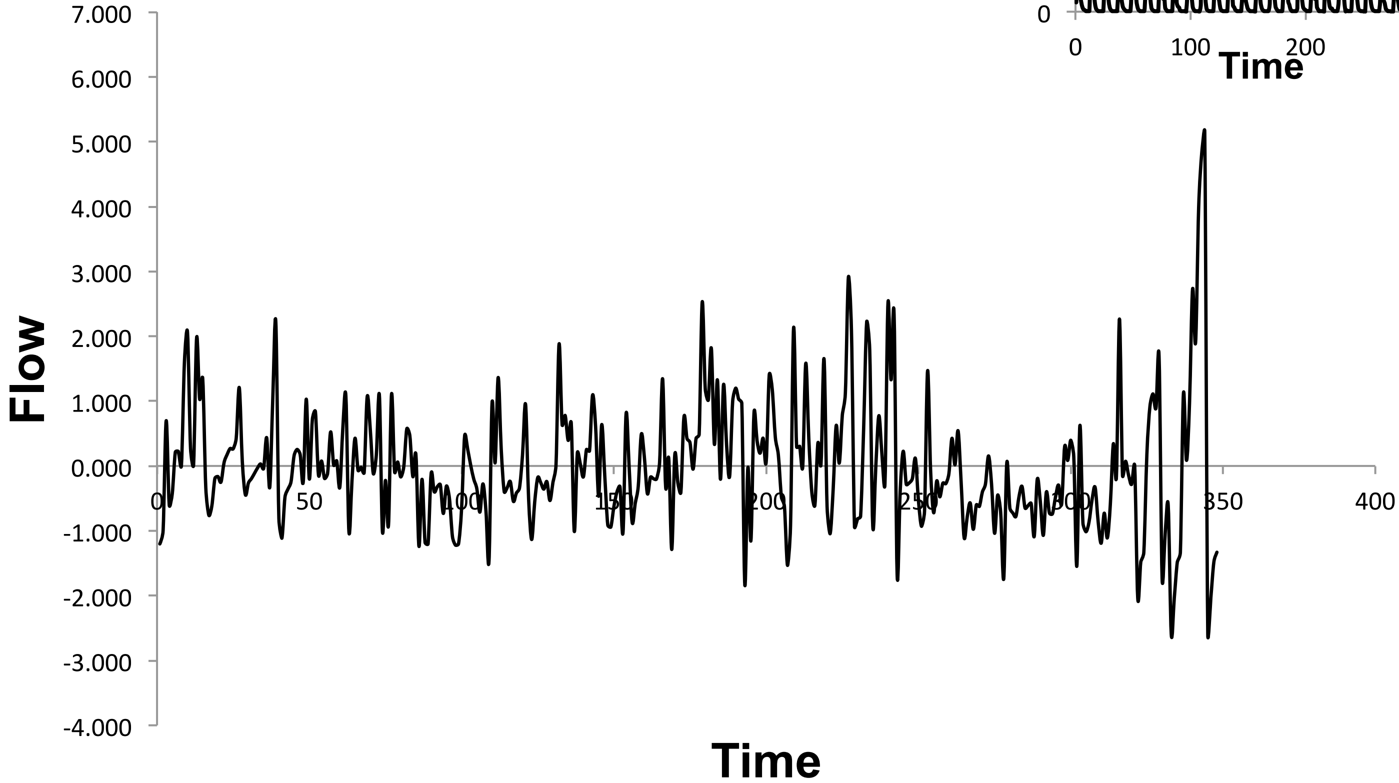
Example – 1 (contd.)

(Spectral Analysis)

Timeseries of original series



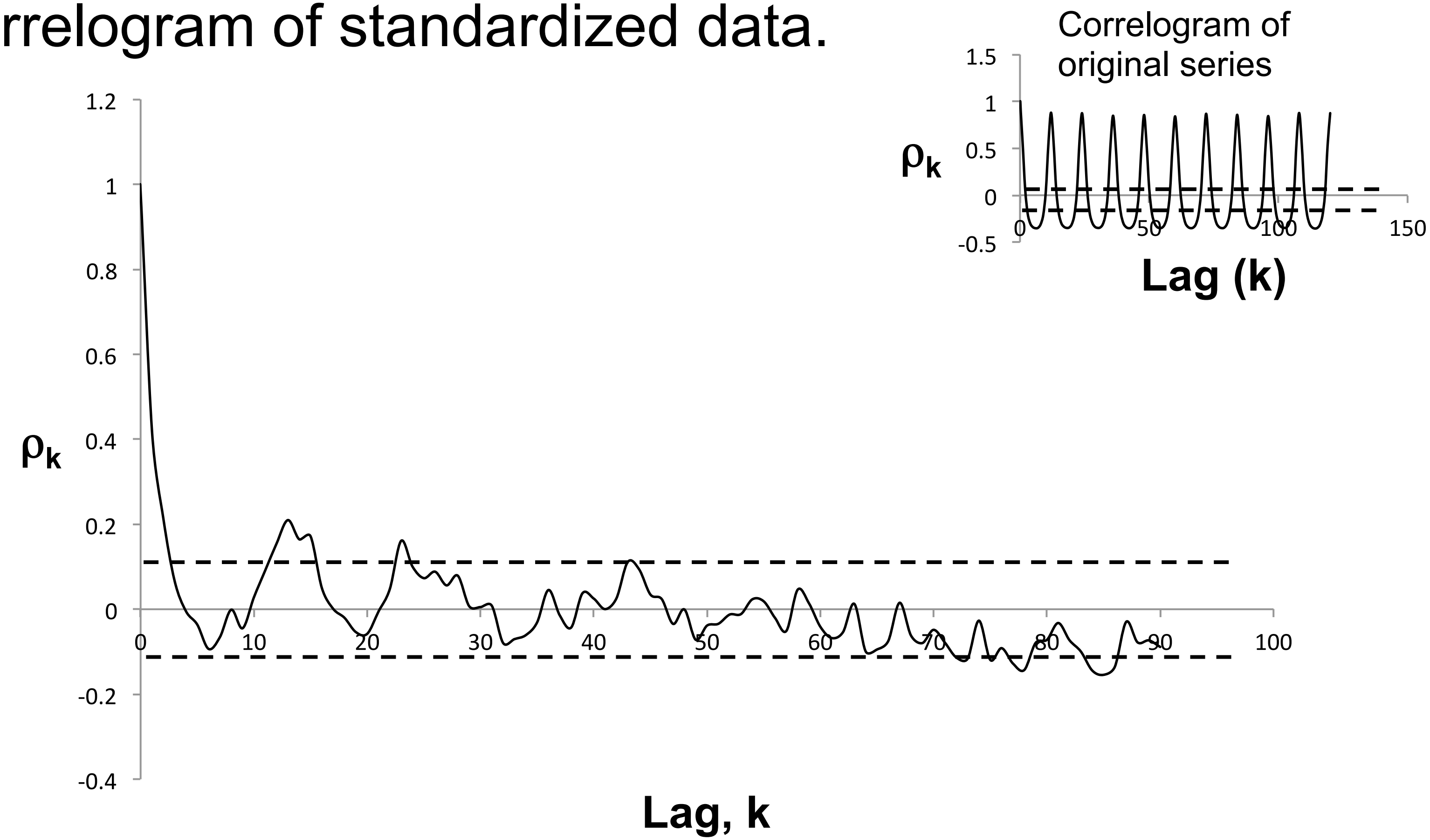
Time series of standardized data.



Example – 1 (contd.)

(Spectral Analysis)

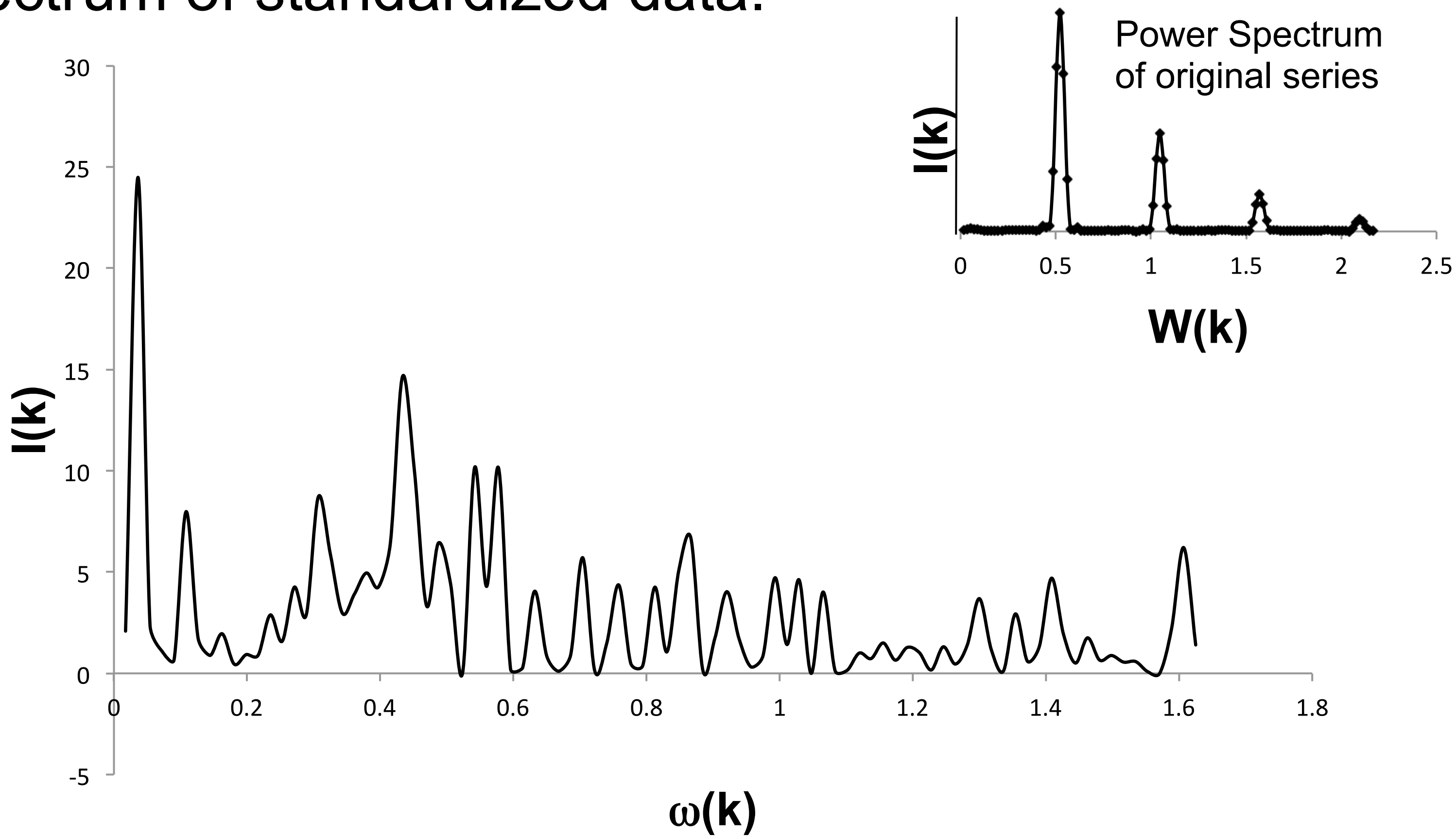
Correlogram of standardized data.



Example – 1 (contd.)

(Spectral Analysis)

Spectrum of standardized data.



Example – 1 (contd.)

(Spectral Analysis)

Test for significance for standardized data:

ω_k	Statistic	F(2, N-2)
0.5236	-4.7E-12	3.0
1.0472	-3.2E-12	3.0
1.5708	-3.5E-11	3.0

$$\mathbf{I} < F(2,346)$$

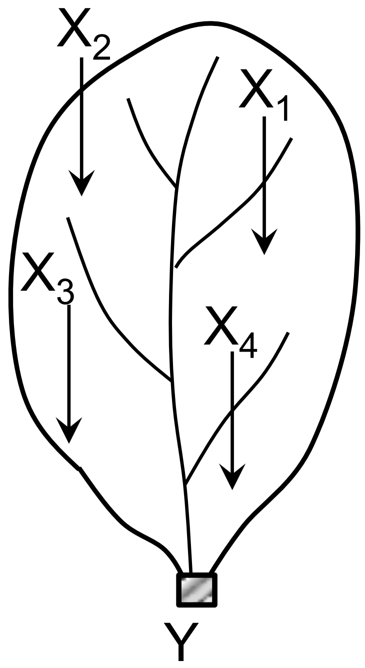
The periodicities are insignificant

ARIMA MODELS

ARIMA Models

Regression:

$Y = f(X_1, X_2, X_3, X_4, \ldots)$



Auto Regression:

$X_t = f(X_{t-1}, X_{t-2}, X_{t-3}, \ldots)$

e.g., AR(1), model

$X_t = \phi_1 X_{t-1} + \varepsilon_t$

(Error, random component, noise, residual)

ARIMA Models

AR(2), model

$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \varepsilon_t$

AR(p) model

$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \ldots\ldots\ldots\ldots\ldots \phi_p X_{t-p} + \varepsilon_t$

$$X_t = \sum_{j=1}^{p} \phi_j X_{t-j} + \varepsilon_t$$

$\{\phi_j\}$ are AR Parameters

Partial Auto Correlation

Partial Auto Correlation (PAC):

Indicates the dependence of X_t on X_{t-k} when the dependence on all other variables X_{t-1}, X_{t-2},…,X_{t-k-1} are removed.

e.g.,Y is regressed upon X_1 and X_2, then it is of interest to ask how much explanatory power X_1 has if the effect of X_2 are partialled out.

This means regressing Y on X_2, getting the residual errors from this analysis and regressing the residuals with X_1.

Partial Auto Correlation

$Y = f(X_1, X_2)$

$Y = f(X_2)$ $\{e_i\}$ get the errors

$X_1 = f(e)$ How much of the relationship is being explained by X_1 alone

For AR(1), model

$X_t = \phi_1 X_{t-1} + \varepsilon_t$ ϕ_1 Partial Auto Correlation (PAC) of order 1

For AR(2), model

$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \varepsilon_t$ ϕ_2 is the PAC of order 2

Partial Auto Correlation

AR(p) model

$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \ldots\ldots\ldots\ldots \phi_p X_{t-p} + \varepsilon_t$

ϕ_p is the PAC of order p

Calculation of Partial Auto Correlations:

(Yule Walker equations)

p^{th} order Yule Walker equations to get ϕ_p

$P_p * \phi_p = \rho_p$

- P_p: Auto Correlation function
- ϕ_p: Partial Auto Correlation
- ρ_p: Auto Correlations

Partial Auto Correlation

Gives partial auto correlation of order 'p'

$$\begin{bmatrix} 1 & \rho_1 & \rho_2 & . & . & \rho_{n-1} \\ \rho_1 & 1 & \rho_1 & . & . & \rho_{n-2} \\ \rho_2 & & & & & \\ . & & & & & \\ . & & & & & \\ \rho_{n-1} & \rho_{n-2} & . & . & . & 1 \end{bmatrix}_{p \times p} \begin{bmatrix} \phi_1 \\ \phi_2 \\ . \\ . \\ . \\ \phi_p \end{bmatrix}_{p \times 1} = \begin{bmatrix} \rho_1 \\ \rho_2 \\ . \\ . \\ . \\ \rho_p \end{bmatrix}_{p \times 1}$$

ρ_{p-1}

Partial Auto Correlation

For PAC of order 1,

$$\begin{bmatrix}1\end{bmatrix}\begin{bmatrix}\phi_1\end{bmatrix} = \begin{bmatrix}\rho_1\end{bmatrix}$$

$$\phi_1 = \rho_1$$

For PAC of order 2, $\begin{bmatrix}1 & \rho_1 \\ \rho_1 & 1\end{bmatrix}\begin{bmatrix}\phi_1 \\ \phi_2\end{bmatrix} = \begin{bmatrix}\rho_1 \\ \rho_2\end{bmatrix}$

$$\phi_1 + \rho_1\phi_2 = \rho_1$$

$$\rho_1\phi_1 + \phi_2 = \rho_2$$

Partial Auto Correlation

$$\phi_1 + \rho_1\left(\rho_2 - \rho_1\phi_1\right) = \rho_1$$

$$\phi_1 + \rho_1\rho_2 - \rho_1^2\phi_1 = \rho_1$$

$$\phi_1 = \frac{\rho_1\left(1 - \rho_2\right)}{1 - \rho_1^2}$$

$$\phi_2 = \rho_2 - \frac{\rho_1^2\left(1 - \rho_2\right)}{1 - \rho_1^2}$$

$$= \frac{\rho_2 - \rho_2\rho_1^2 - \rho_1^2 + \rho_2\rho_1^2}{1 - \rho_1^2}$$

$$= \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2}$$

ϕ_2 is PAC of order 2

Example – 2

Obtain the ϕ_1 and ϕ_2 for

$r_1 = 0.57, r_2 = 0.07$

Since $\phi_1 = r_1$

$\phi_1 = 0.57$

$$\phi_2 = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2}$$

$$= \frac{0.07 - 0.57^2}{1 - 0.57^2}$$

$$= -0.38$$

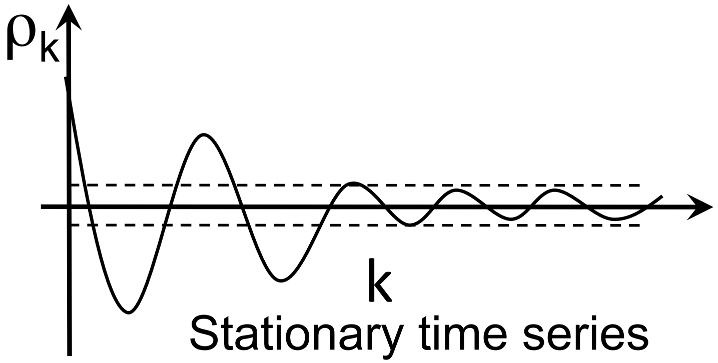
Sample estimates

$-1 \le \phi_p \le +1 \;\forall p$

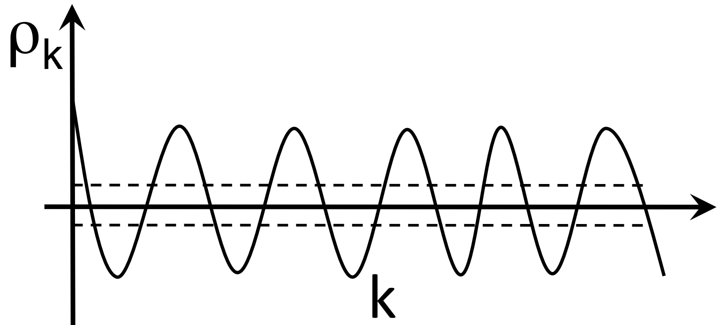
ARIMA Models

Box Jenkins Time series models:

- For stationary time series
- If the time series is stationary, the correlogram dies down fairly quickly (e.g., within 4 or 5 lags, in most hydrologic applications)
- If the time series is non stationary, the decay is very slow



Stationary time series



Non-stationary time series

ARIMA Models

- If the time series is non stationary, convert it to a stationary time series
- One way is by standardizing the time series described in spectral analysis
- Another way is by simply differencing the time series.

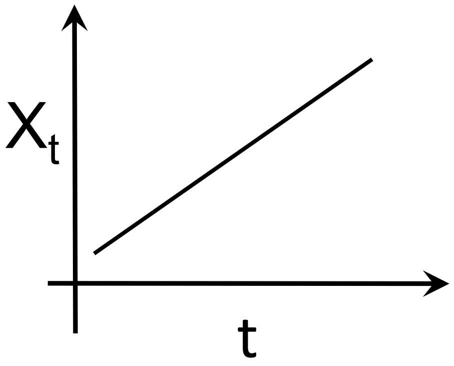
ARIMA Models

- Differencing:

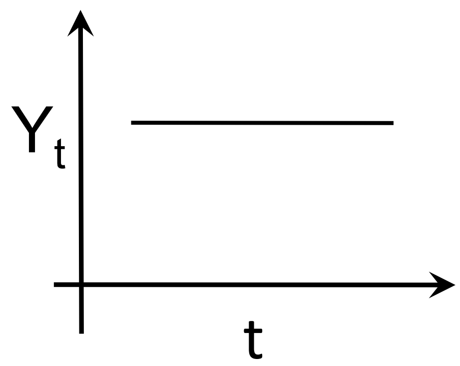
$$Y_t = X_t' = X_t - X_{t-1}$$

X_t' is First order differencing

$\{X_t\} = 2, 4, 6, 8, 10, \ldots\ldots$



$\{Y_t\} = 2, 2, 2, \ldots\ldots$



ARIMA Models

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

X_t'' is Second order differencing

$$\begin{aligned} X_t'' &= X_t' - X_{t-1}' \\ &= \left(X_t - X_{t-1}\right) - \left(X_{t-1} - X_{t-2}\right) \\ &= X_t - 2X_{t-1} + X_{t-2} \end{aligned}$$

Example – 3
(Differencing)

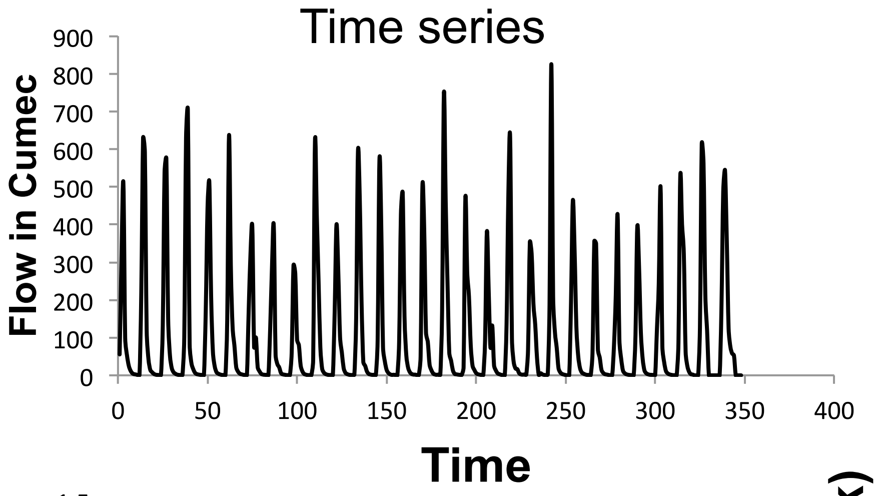
Period,t	X_t	X_t'	X_t''
1	54.6	-	-
2	325.4	-270.8	-
3	509.5	-184.1	-86.7
4	99.4	410.1	-594.2
5	53.5	45.9	364.2
6	25.8	27.7	18.2
7	12.5	13.3	14.4
8	5.6	6.9	6.4
9	3.1	2.5	4.4
10	2.2	0.9	1.6
11	0.9	1.3	-0.4
12	0.81	0.09	1.21

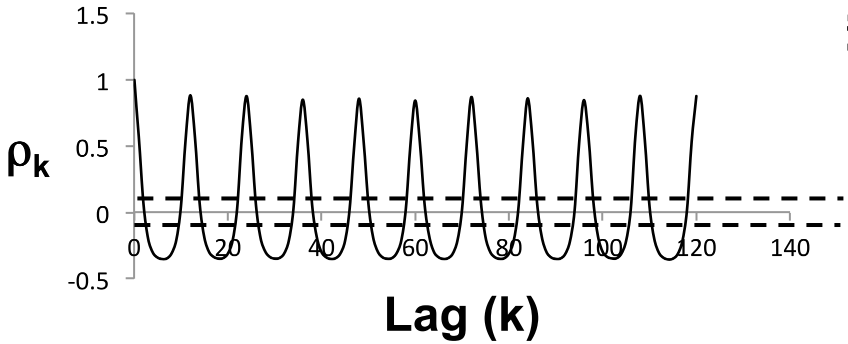
Example – 4

Monthly Stream flow (in cumec) statistics(1979-2008) for a river is selected for the study. (Part data shown below)

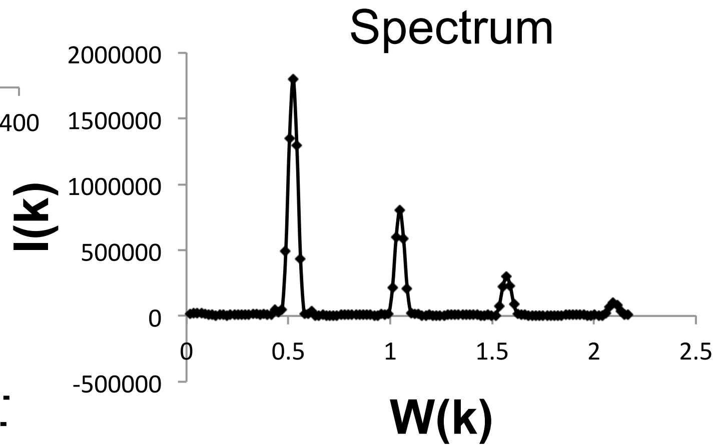
Year	Month	S.No.	Flow
1979	June	1	54.6
	July	2	325.4
	August	3	509.5
	September	4	99.4
	October	5	53.5
	November	6	25.8
	December	7	12.5
1980	January	8	5.6
	February	9	3.1
	March	10	2.2
	April	11	0.9
	May	12	0.81

Example – 4 (contd.)



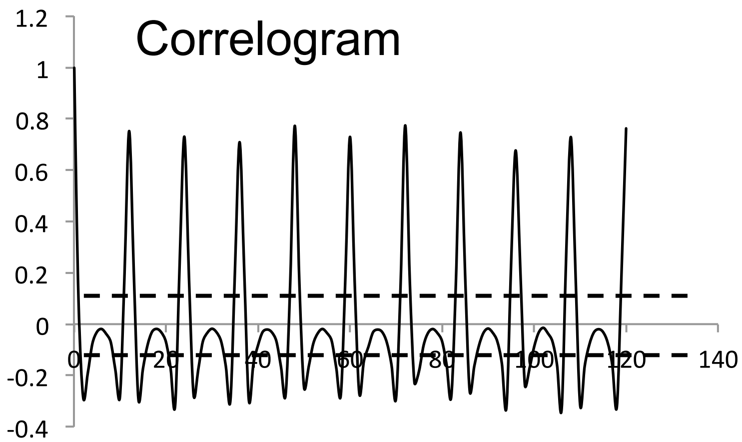


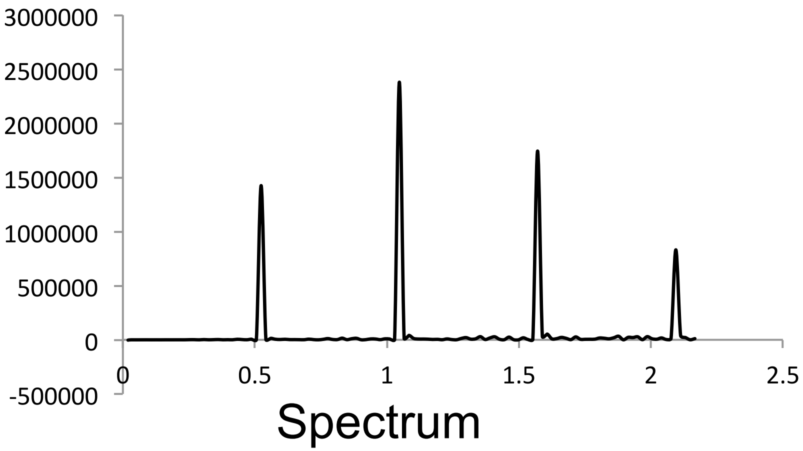
Correlogram



Example – 4 (contd.)

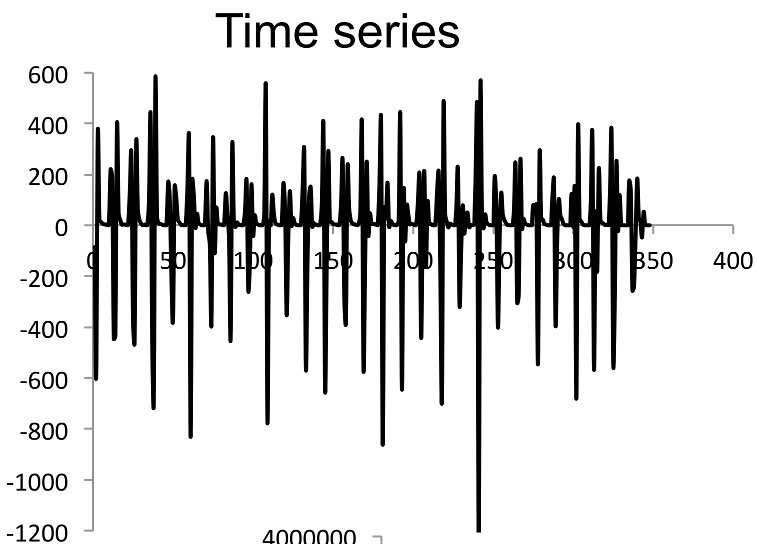
First order differenced data, $X_t' = X_t - X_{t-1}$

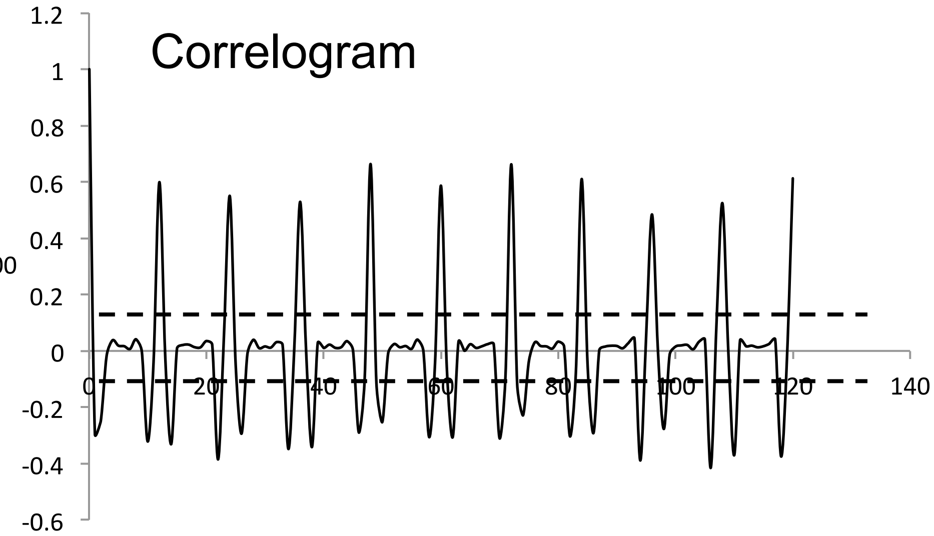


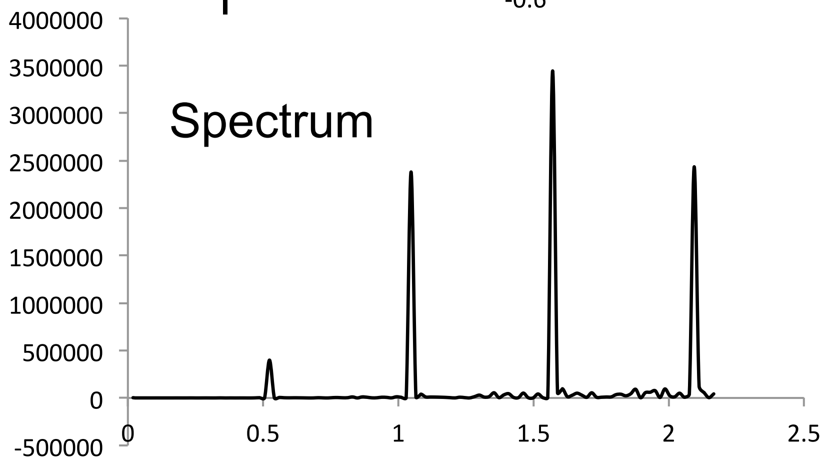


Example – 4 (contd.)

Second order differenced data



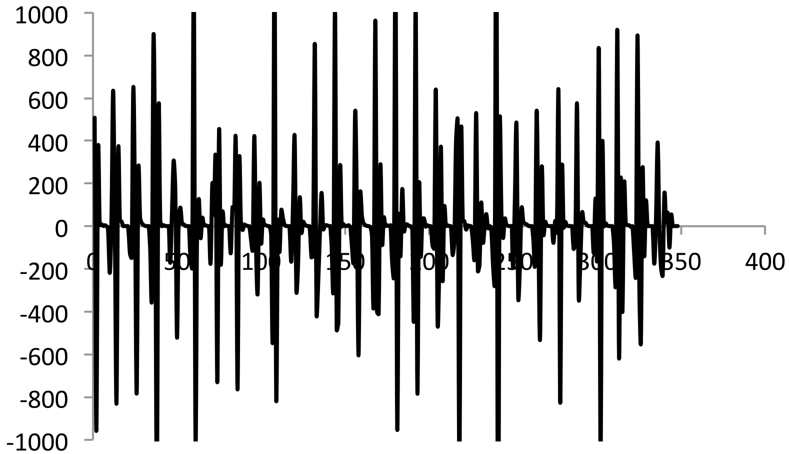


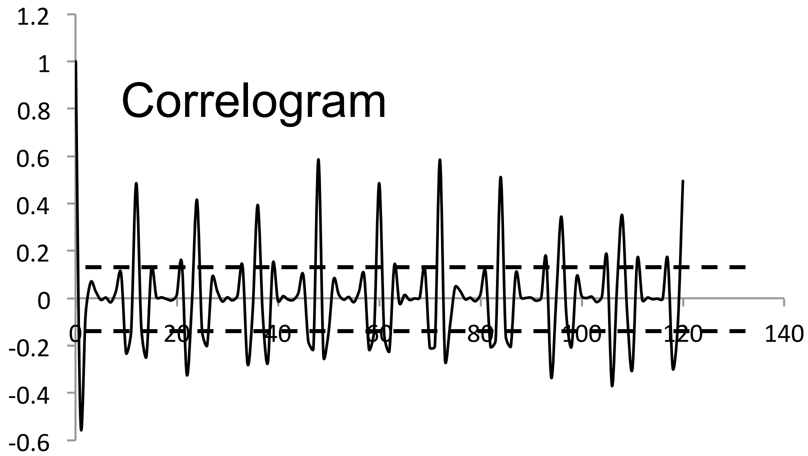


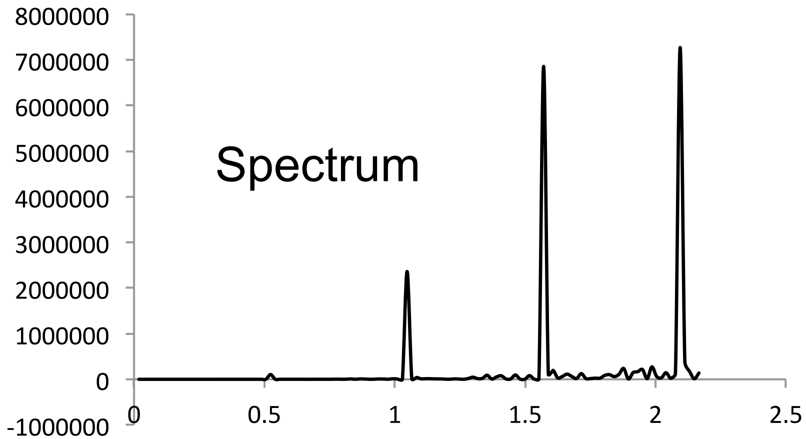
Example – 4 (contd.)

Third order differenced data

Time series

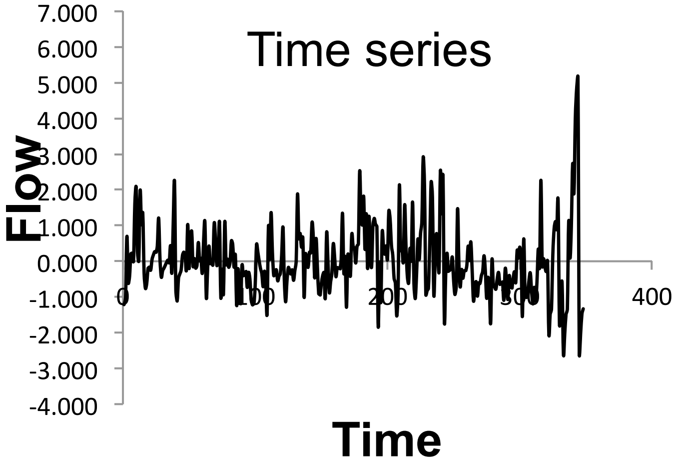


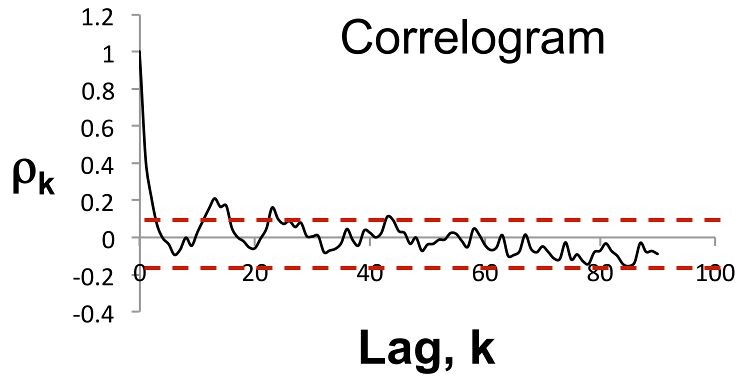


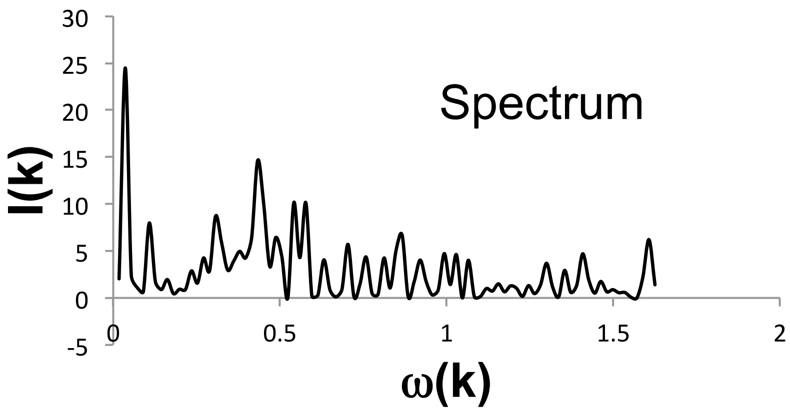


Example – 4 (contd.)

Standardized data







ARIMA Models

- Operator ‘B’:

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

$$BX_t = X_{t-1}$$

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

AR (1) Model: $X_t = \phi_1 X_{t-1} + \varepsilon_t$

$$X_t = \phi_1 BX_t + \varepsilon_t$$

$$\underbrace{X_t(1 - \phi_1 B)}_{\text{AR (1) component}} = \varepsilon_t$$

ARIMA Models

AR (2) Model: $X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \varepsilon_t$

$$X_t = \phi_1 B X_t + \phi_2 B X_{t-1} + \varepsilon_t$$

$$X_t = \phi_1 B X_t + \phi_2 B^2 X + \varepsilon_t$$

$$X_t\underbrace{(1 - \phi_1 B - \phi_2 B^2)}_{\text{AR (2) component}} = \varepsilon_t$$

Generalized form for an AR(p) model is

$$X_t \left(1 - \sum_{i=1}^{p} \phi_i B^i \right) = \varepsilon_t$$

ARIMA Models

Auto Regressive Integrated Moving Average models:

Order of differencing

ARIMA (p, d, q)

No. of Moving average terms

No. of Auto-regressive terms

ARIMA Models

Auto Regressive Moving Average models:

ARMA (p, q)

$$X_t = \underbrace{\phi_1 X_{t-1} + \phi_2 X_{t-2} + \ldots + \phi_p X_{t-p}}_{\text{AR of order 'p'}} + \overbrace{\theta_1 e_{t-1} + \theta_2 e_{t-2} + \ldots + \theta_q e_{t-q}}^{\text{Residuals of order 'q'}} + e_t$$

ARIMA Models

First order differencing:

$X_t – X_{t-1} = e_t$

$X_t – BX_t = e_t$

$X_t (1 – B) = e_t$

Second order differencing: $X_t'' = X_t' - X_{t-1}'$

$$= \left(X_t - X_{t-1}\right) - \left(X_{t-1} - X_{t-2}\right)$$

$$= X_t - 2X_{t-1} + X_{t-2}$$

$$= X_t - 2BX_t + B^2 X_t$$

$$= \left(1 - B\right)^2 X_t$$

ARIMA Models

In general d^{th} order difference is $\left(1-B\right)^{d} X_t$

ARIMA (1, 1, 1)

$Y_t = X_t - X_{t-1}$

$$Y_t = \phi_1 Y_{t-1} + \theta_1 e_{t-1} + e_t$$

$$X_t - X_{t-1} = \phi_1 \left(X_{t-1} - X_{t-2} \right) + \theta_1 e_{t-1} + e_t$$

$$X_t - BX_t = \phi_1 \left(BX_t - B^2 X_t \right) + \theta_1 B e_t + e_t$$

$$X_t \left(1 - B - \phi_1 B + \phi_1 B^2 \right) = e_t \left(1 + \theta_1 B\right)$$

ARIMA Models

Procedure for fitting Box-Jenkins type time series models:

3 steps

1. Identification of the model structure
2. Parameter estimation and calibration
3. Model testing / Validation

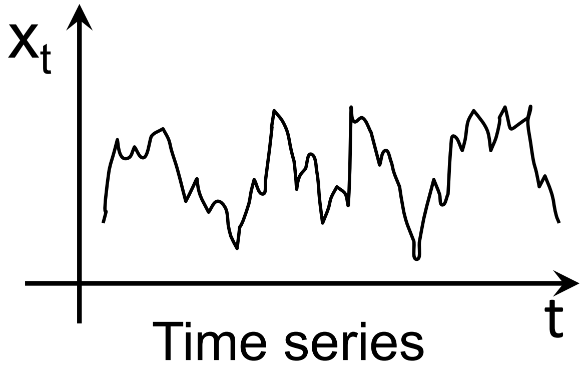
ARIMA Models

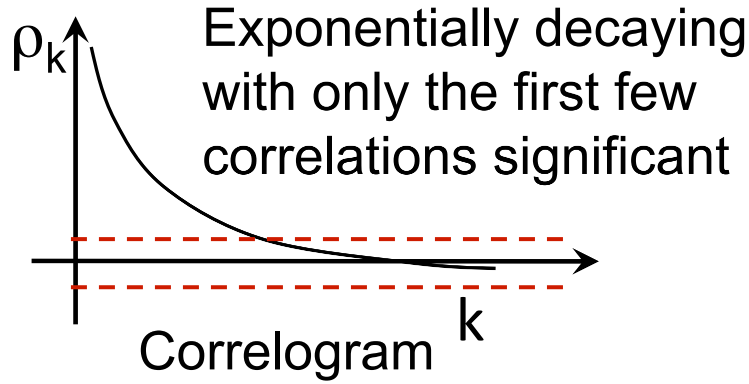
1. Identification of the model structure:

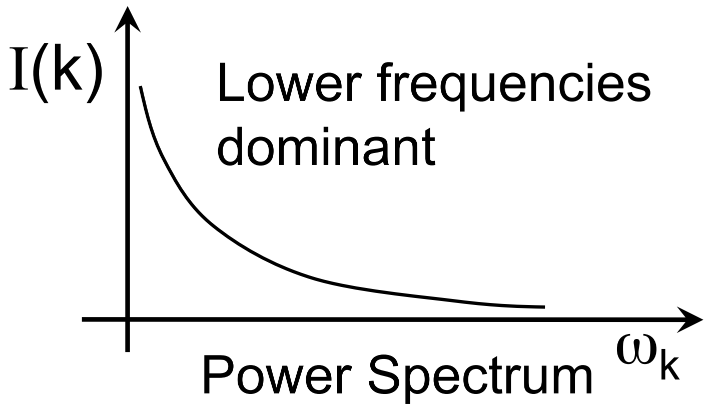
- Identify if the series is stationarity.
  - Plot correlogram (correlogram shows a rapid decay for a stationary series)
- Remove non-stationarity if any by differencing/ standardization.
- Obtain the order of AR and MA components of the model.
- PAC determines the order of the AR process

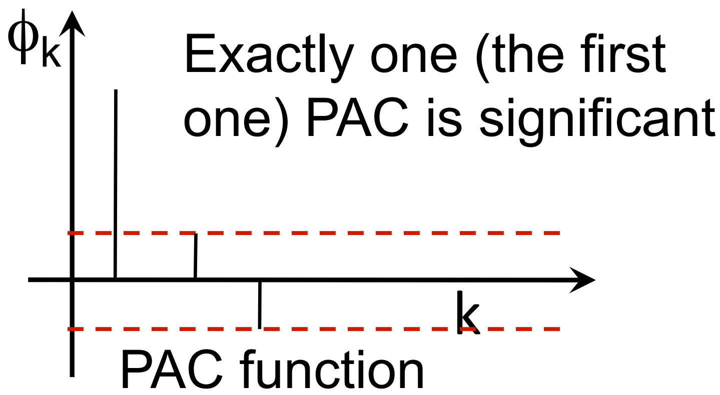
ARIMA Models

For example, AR(1) process:



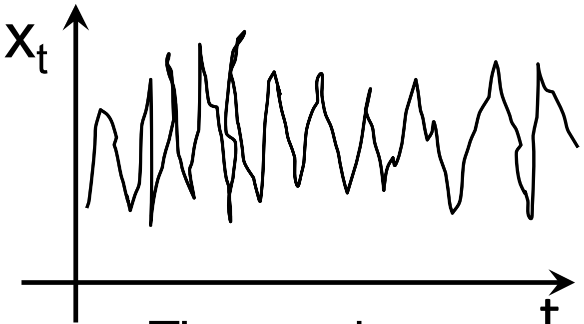




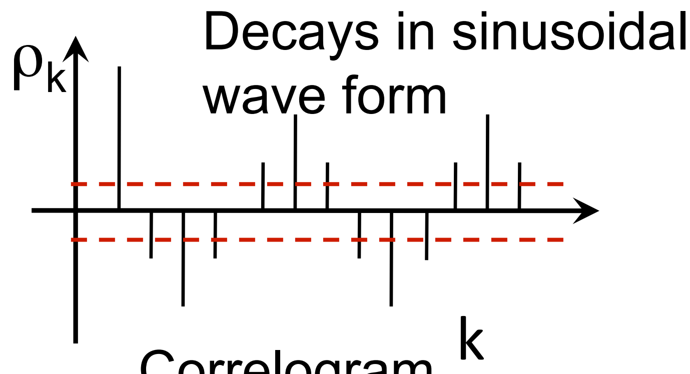


ARIMA Models

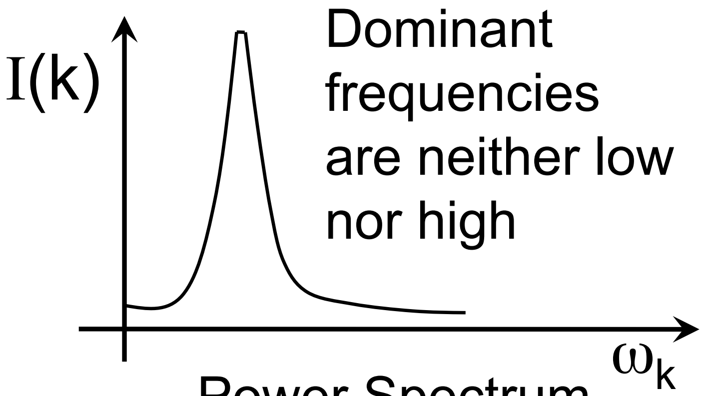
AR(2) process:



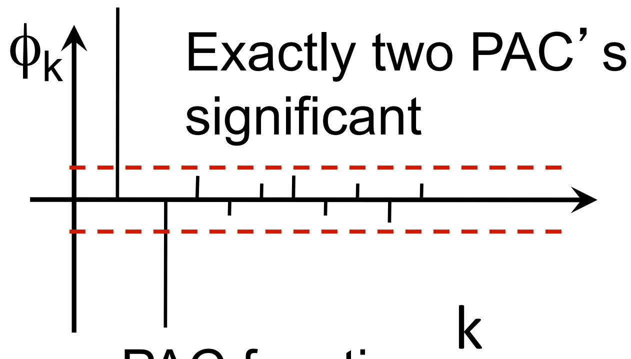
Time series



Correlogram



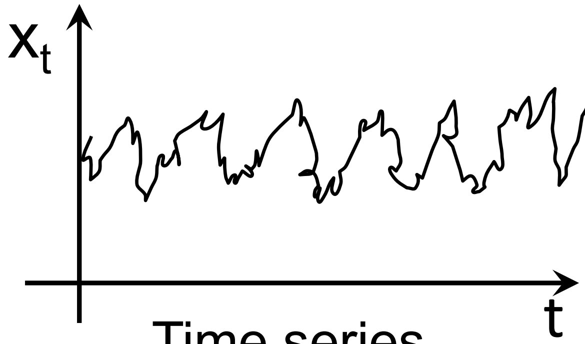
Power Spectrum



PAC function

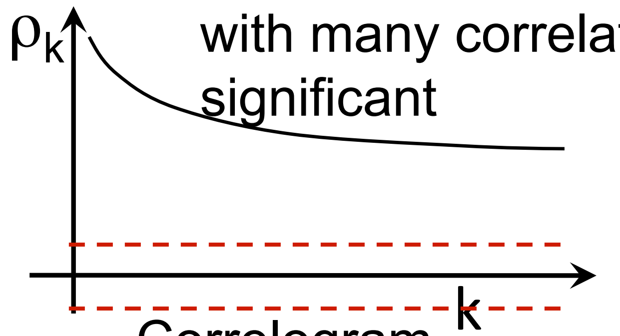
ARIMA Models

Another AR(2) process:



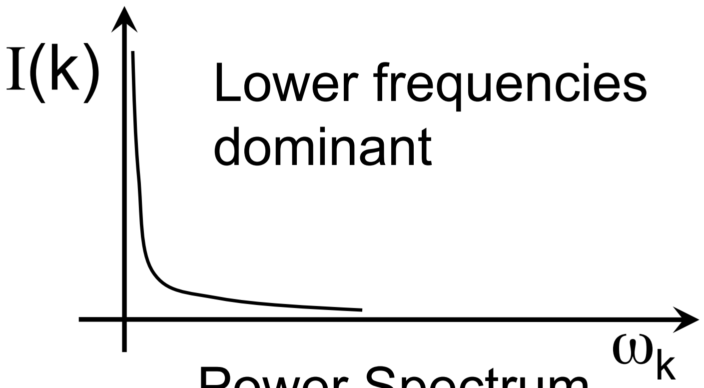
Time series

Exponentially decaying with many correlations significant



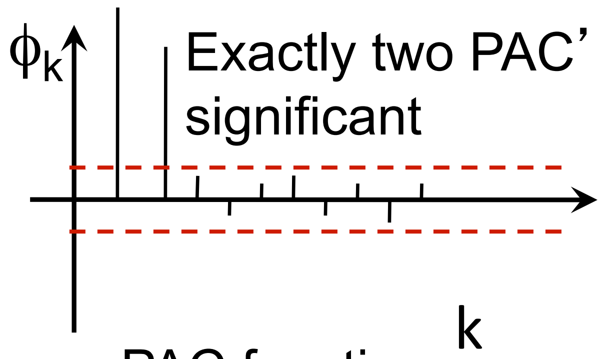
Correlogram

Lower frequencies dominant



Power Spectrum

Exactly two PAC's significant



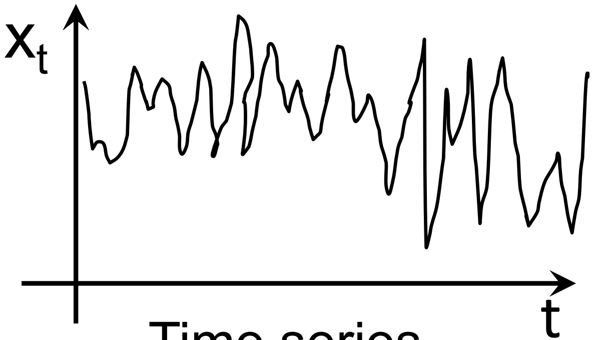
PAC function

ARIMA Models

- Behavior of AR process:
  - Decaying auto correlation function (either exponentially or in a dampened sine wave)
  - Order of AR determined by the significant PAC's

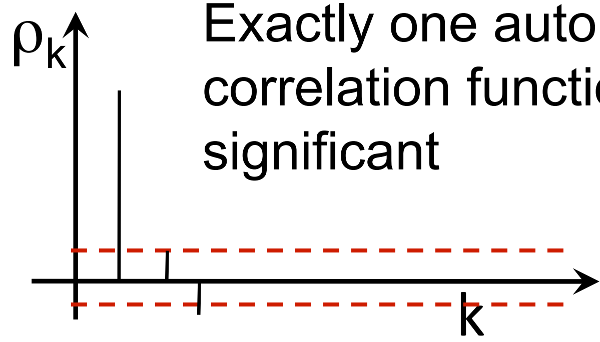
ARIMA Models

MA(1) process:

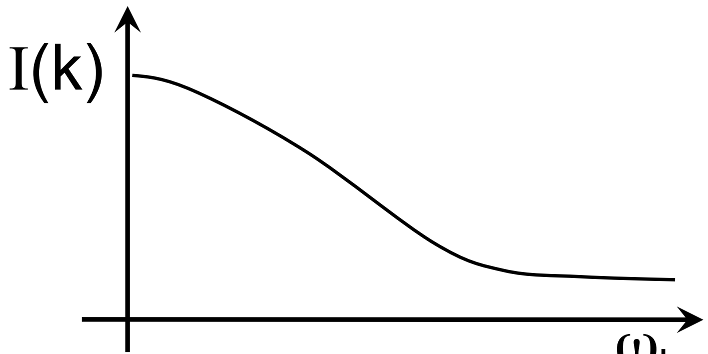


Time series

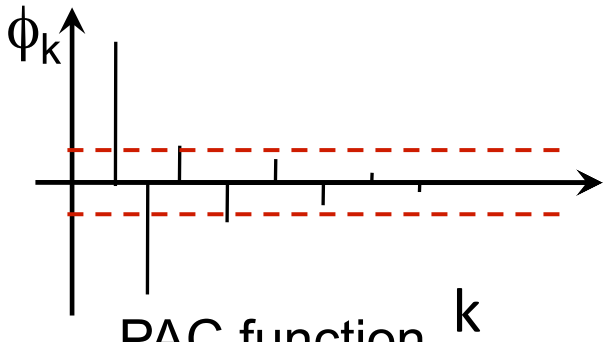
Exactly one auto correlation function is significant



Correlogram



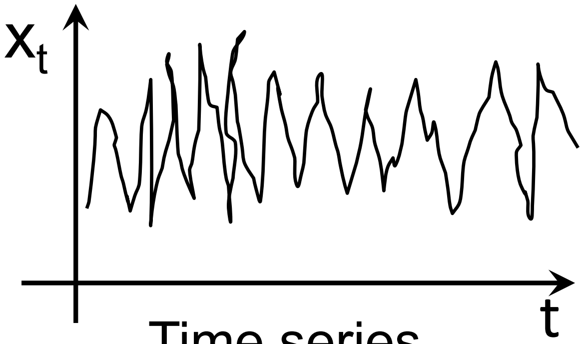
Power Spectrum



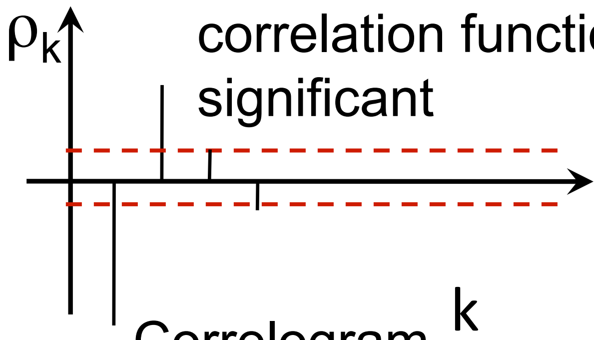
PAC function

ARIMA Models

MA(2) process:

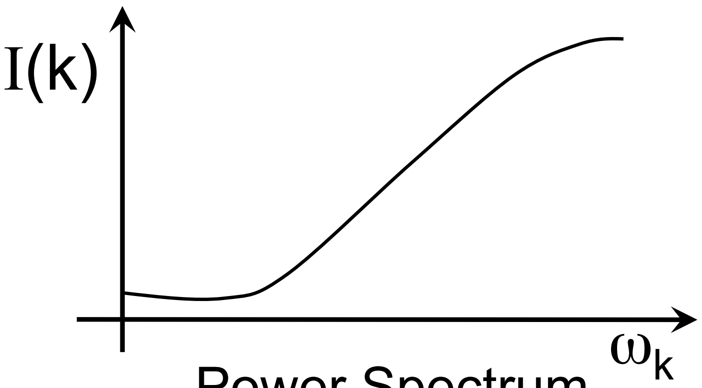


Time series

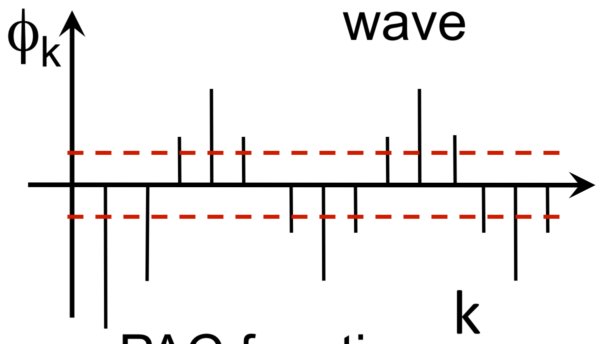


Exactly two auto correlation functions significant

Correlogram



Power Spectrum



Decays in sinusoidal wave

PAC function

ARIMA Models

- Behavior of MA process:
  - The order of MA is determined by the number of significant auto correlations
  - Decaying PAC function (either exponentially or in a dampened sine wave)