INDIAN INSTITUTE OF SCIENCE

STOCHASTIC HYDROLOGY

Lecture -13

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

Summary of the previous lecture

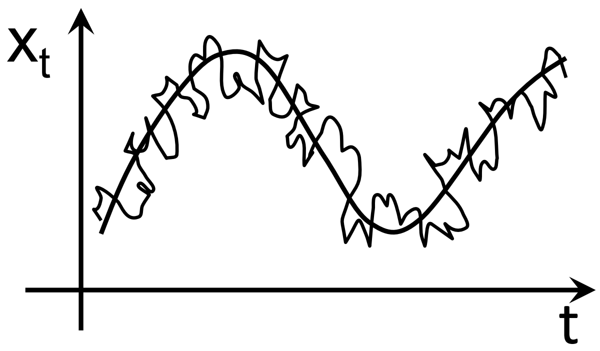
- Data Generation – Serially Correlated Data
  - First order Markov Model
    - Annual flow generation
  - First order Markov model with non-stationarity
    - Thoma Fiering model for monthly and seasonal flow generation

FREQUENCY DOMAIN ANALYSIS

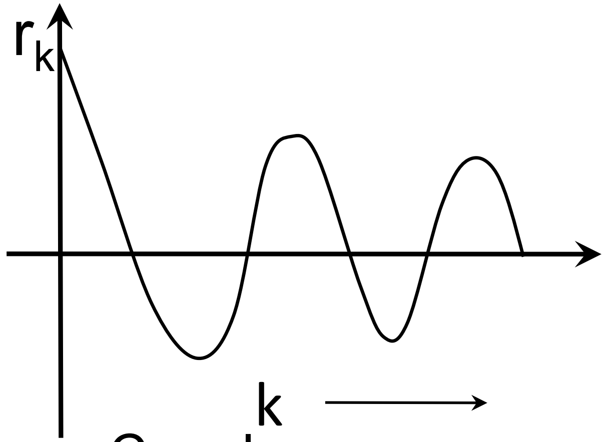
Frequency Domain Analysis

- Auto correlation function or correlogram is used for analyzing the time series in the time domain.
- Time domain analysis

$$X_t = d_t + \varepsilon_t$$



Periodic process with noise



Correlogram

Frequency Domain Analysis

- Periodicities in data can best be determined by analyzing the time series in frequency domain.
- Spectral analysis or the frequency domain analysis: the time series is represented in the frequency domain instead of the time domain
- The observed time series is a random sample of a process over time and is made up of oscillations of all possible frequencies.

Frequency Domain Analysis

- Spectral analysis is widely used in electrical engineering, physics, meteorology and hydrology
- Hydrologic applications of spectral analysis include:
  - Dry and wet run analysis
  - Synthetic data generation
  - Hydrologic forecasting
  - Climate change impact studies

Frequency Domain Analysis

N odd N even

$$X_t = \alpha_0 + \sum_{k=1}^{\frac{N-1}{2}, \frac{N}{2}} \left[\alpha_k \cos\left(2\pi f_k t\right) + \beta_k \sin\left(2\pi f_k t\right) \right] + \varepsilon_t$$

t = 1, 2, …. N

$$f_k = \frac{k}{N} \quad ;$$

k^{th} harmonic of the fundamental frequency (1/N)

N is the no. of observations

Periodicity (P): $P = \dfrac{1}{f_k}$

Frequency Domain Analysis

$$\alpha_0 = \overline{x}$$

$$\alpha_k = \frac{2}{N}\sum_{t=1}^{N} x_t \cos\left(2\pi f_k t\right) \qquad \text{k = 1, 2, …. M}$$

$$\beta_k = \frac{2}{N}\sum_{t=1}^{N} x_t \sin\left(2\pi f_k t\right) \qquad \text{k = 1, 2, …. M}$$

M is maximum lag (typically considered up to 0.25N)

The above equations for α_k and β_k are valid up to k=N/2

Frequency Domain Analysis

When 'N' is odd, the expressions are true until

$$k = \frac{N}{2} - 1$$

$$\alpha_{N/2} = \frac{1}{N} \sum_{t=1}^{n} (-1)^t x_t$$

$$\beta_{N/2} = 0$$

Frequency Domain Analysis

- A variance spectrum divides the variance into no. of intervals or bands of frequency.
- Spectral density (I_k) is the amount of variance per interval of frequency.

$$I(k) = \frac{N}{2}\left[\alpha_k^2 + \beta_k^2\right] \qquad k = 1, 2, \ldots.\ M$$

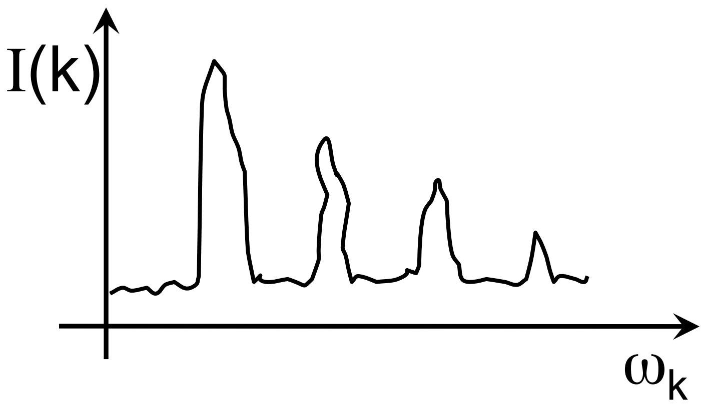
- Angular frequency

$$\omega_k = \frac{2\pi k}{N} \qquad k = 1, 2, \ldots.\ M$$

Frequency Domain Analysis

$$\omega_k = \frac{2\pi}{P}$$

A plot of ω_k vs $\mathrm{I}(k)$ is called spectrum



- Total area under the spectrum is equal to the variance of the process
- A peak in the spectrum indicates an important contribution to variance at frequencies close to the peak
- Prominent spikes indicate periodicity
- Several expressions for spectrum exist in literature

Example-1

Obtain ω_k and $\mathrm{I}(k)$ for k=1

t	X_t	$\cos\left(2\pi f_k t\right)$	$\sin\left(2\pi f_k t\right)$	$X_t \cos\left(2\pi f_k t\right)$	$X_t \sin\left(2\pi f_k t\right)$
1	105	0.809	0.5878	84.945	61.719
2	115	0.309	0.9511	35.535	109.3765
3	103	-0.309	0.9511	-31.827	97.9633
4	94	-0.809	0.5878	-76.046	55.2532
5	95	-1	0	-95	0
6	104	-0.809	-0.5878	-84.136	-61.1312
7	120	-0.309	-0.9511	-37.08	-114.132
8	121	0.309	-0.9511	37.389	-115.083
9	127	0.809	-0.5878	102.743	-74.6506
10	79	1	0	79	0
Σ				15.523	-40.6849

Example-1 (contd.)

$$f_k = \frac{k}{N}$$

$$= \frac{1}{10} = 0.1$$

$$\alpha_k = \frac{2}{N}\sum_{t=1}^{N} x_t \cos\left(2\pi f_k t\right)$$

$$= \frac{2}{10} \times \left(15.523\right)$$

$$= 3.1046$$

$$\beta_k = \frac{2}{N}\sum_{t=1}^{N} x_t \sin\left(2\pi f_k t\right)$$

$$= \frac{2}{10} \times \left(-40.6849\right)$$

$$= -8.13698$$

Example-1 (contd.)

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

$$= \frac{10}{2}\left[(3.1046)^2 + (-8.13698)^2\right]$$

$$= 379.245$$

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

$$= \frac{2 \times \pi \times 1}{10}$$

$$= 0.62832$$

Frequency Domain Analysis

- Spectral density as defined earlier is also called as line spectrum

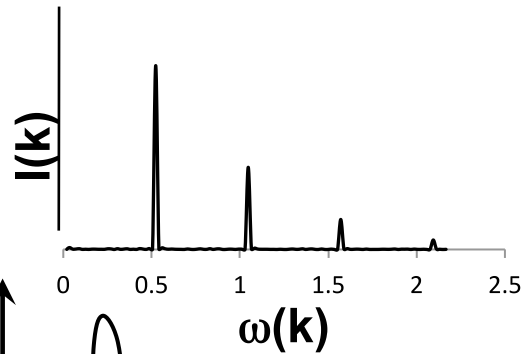
- The line spectrum transforms the information from time domain to the frequency domain

- While the correlogram indicate the presence of periodicities in the data, the spectral analysis helps indentify the significant periodicities themselves

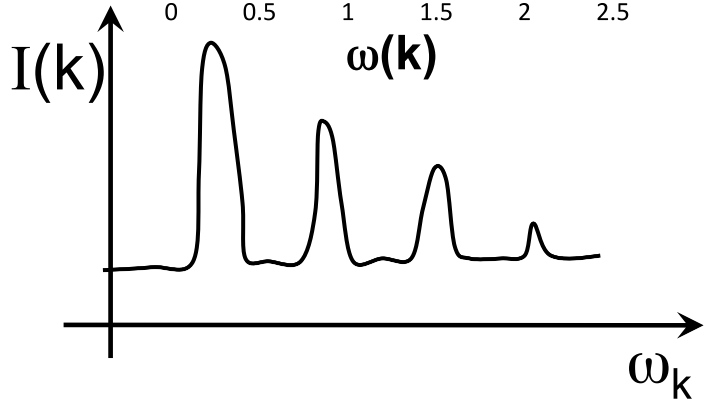
Frequency Domain Analysis

- Line spectrum as defined is an inconsistent estimate

- The plot is not a smooth function



- The smoothened spectrum is called as power spectrum



- Power spectrum is a consistent estimate of spectral density

Frequency Domain Analysis

- Power spectrum – Fourier cosine transform of auto covariance function.

$$I(k) = 2\left[c_0 + 2\sum_{j=1}^{\frac{N-1}{2}} \lambda_j c_j \cos\left(2\pi f_k j\right)\right]$$

c_j = Auto covariance function

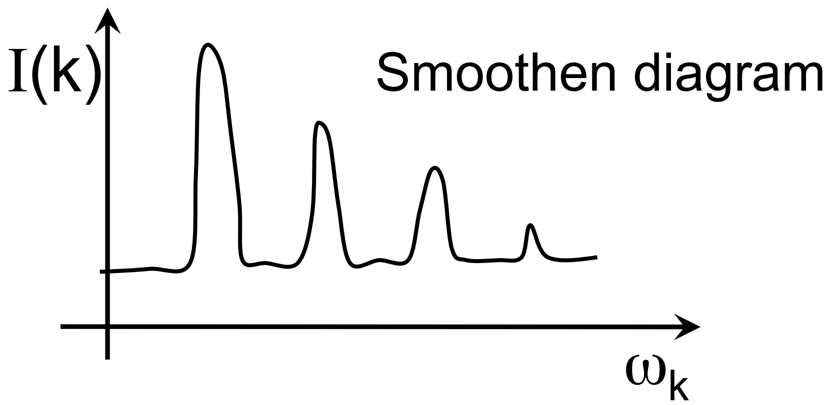
λ_j = lag window (or smoothing window)

Frequency Domain Analysis

Tukey window

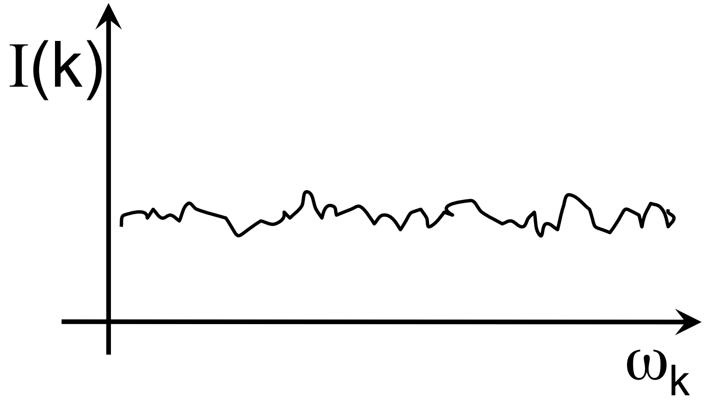
$$\lambda_j = \frac{1}{2}\left[1 + 2\cos\left(\frac{2\pi}{M'}\right)\right]$$

M' = Maximum lag (~ 0.25N)



Frequency Domain Analysis

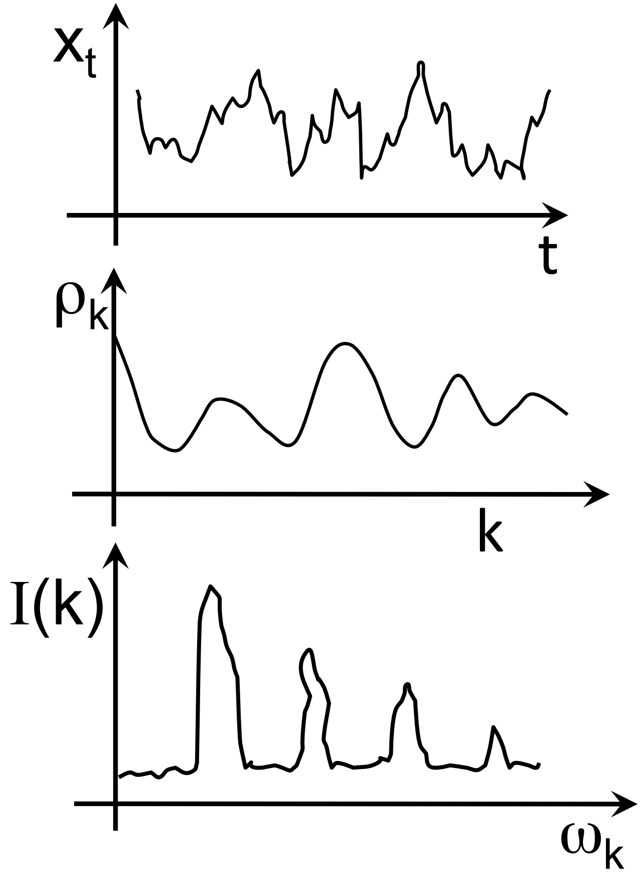
- Information content is extracted from spectrum.
- For a completely random series (e.g., uniformly distributed random numbers), the spectral density function is constant – termed as white noise
- White noise indicates that no frequency interval contains any more variance than any other frequency interval. (auto correlation function $\rho_k = 0$, for $k \neq 0$)



Frequency Domain Analysis

The steps for analyzing the data are as follows

- Plot the time series
- Plot the correlogram
- Plot the spectrum



Frequency Domain Analysis

- The spectrum shows prominent spikes (which represent the periodicities inherent in the data)
- The period corresponding to any value of ω_k may be computed by $2\pi/\ \omega_k$.

- To test the significance of a periodicity, the periodicities which are earlier tested to be significant, are removed from the original series to get a new series $\{Z_t\}$, where

$$Z_t = X_t - Y_t\ ,$$

Frequency Domain Analysis

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

where d is no. of periodicities removed (which are known to be significant)

- The spectrum of new series Z_t is plotted and the spikes are observed.
- A wrong conclusion may be made that these spikes are significant. However they need to be analyzed for their statistical significance

Frequency Domain Analysis

Statistical significance of the periodicities:

The periodicities are tested for significance by defining a statistic '∩' as follows (Kashyap and Rao 1976)

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

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

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

Ref: Kashyap R L and Ramachandra Rao A 'Dynamic stochastic models from empirical data', Academic press, New York, 1976

Frequency Domain Analysis

The periodicity corresponding to ω_k is significant at level α only if

$$\mathrm{I} \quad \geq F\left(2, N-2\right)$$

Where ‘F’ denotes F distribution

•This test examines one periodicity at a time and should be carried out on a series from which all periodicities (previously found significant) are removed.

Frequency Domain Analysis

- A necessary condition in stochastic models is that the series being modeled must be free from any significant periodicities.
- One way of removing the periodicities from the time series is to simply transform the series into a standardized one.
- One method of standardizing the series $\{X_t\}$ is by expressing $\{X_t\}$ as the new series $\{Z_t\}$ where,

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

Frequency Domain Analysis

- In a monthly series, for example, X_i is the estimate of mean of month 'i' to which period 't' belongs
- S_i is the estimate of the standard deviation of month 'i'
- The series has zero mean and unit variance.

- The series without periodicities is then obtained for which a stochastic (e.g., ARMA – Auto Regressive Moving Average) model is created.

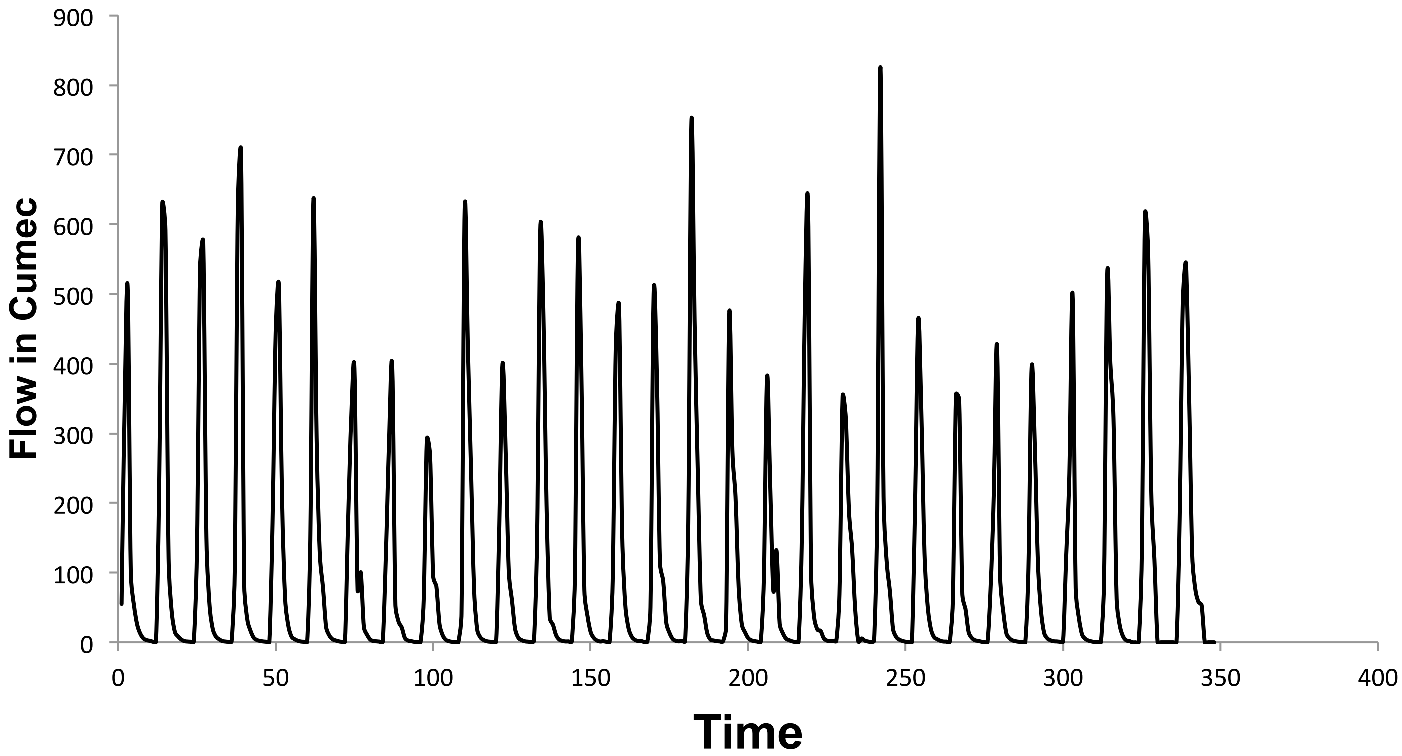
Example – 2

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

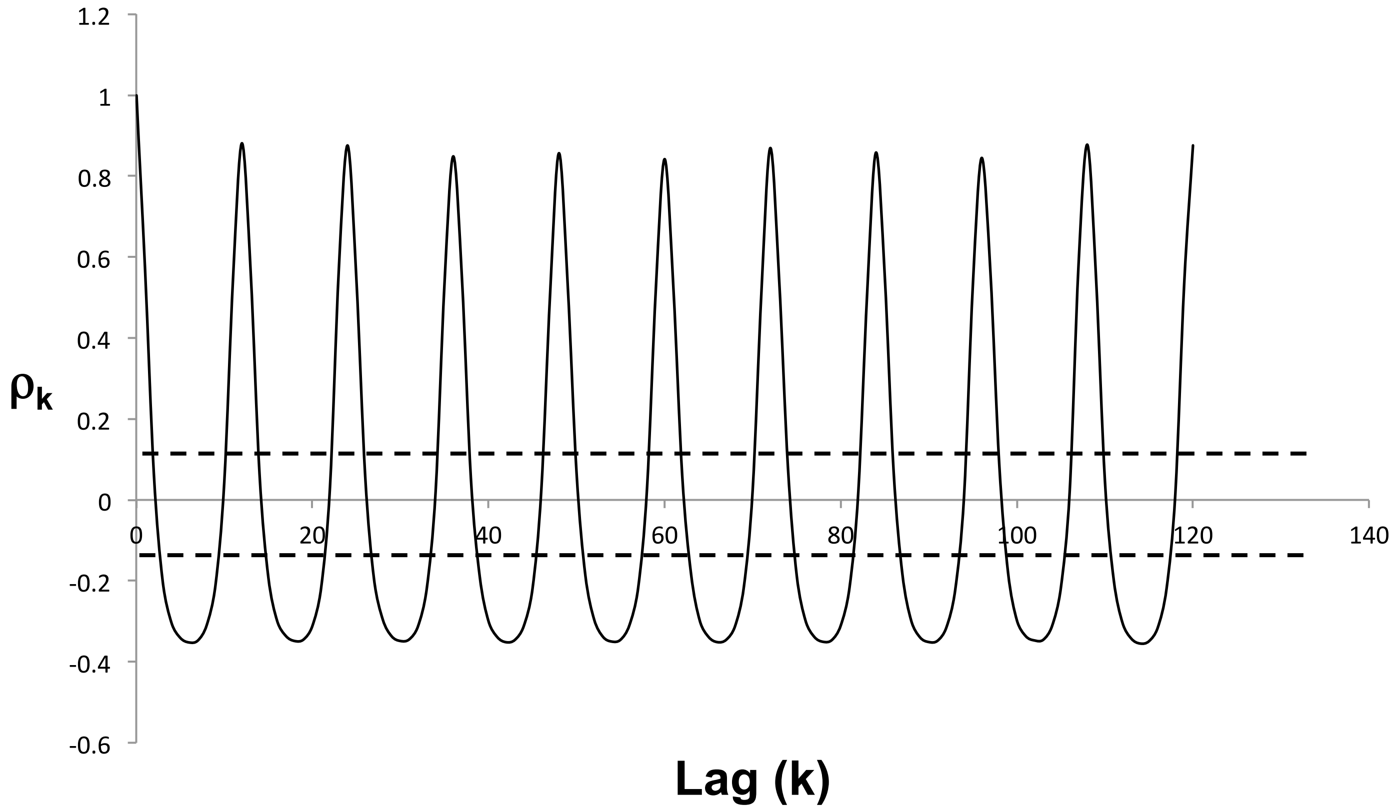
Example – 2 (contd.)

The time series plot



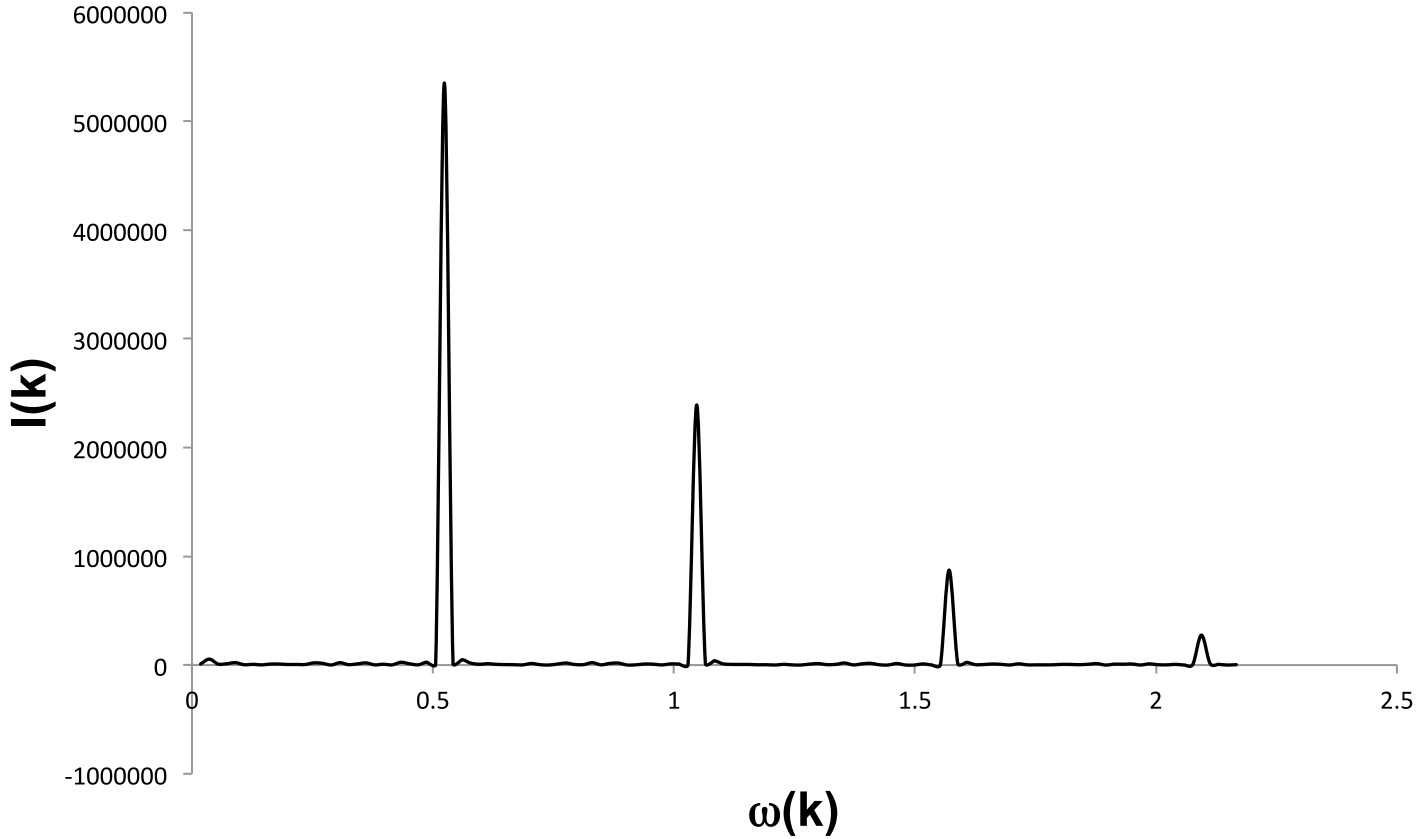
Example – 2 (contd.)

- Correlogram



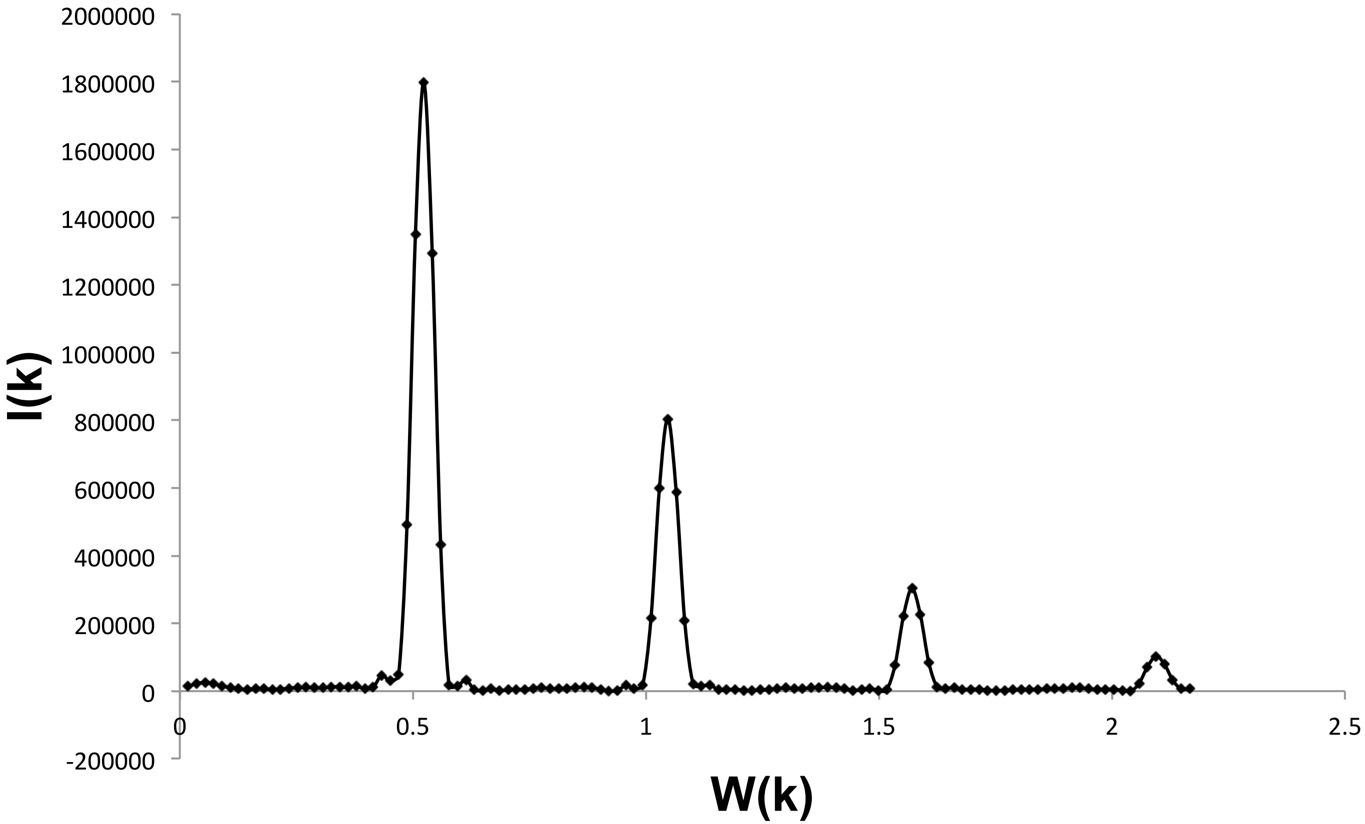
Example – 2 (contd.)

- Line Spectrum



Example – 2 (contd.)

- Power Spectrum



Example – 2 (contd.)

- Peaks represent the periodicities inherent in the data.

$$P = 2\pi / \omega_k$$

- The peaks correspond to w(k) = 0.5236 indicating a periodicity of 12 months,
- w(k) = 1.0472 : periodicity of 6 months,
- w(k) = 1.571 : periodicity of 4 months,
- w(k) = 2.094 : periodicity of 3 months

Example – 2 (contd.)

- Considering the first two periodicities are significant, the two periodicities are removed from the original series to get a new series $\{Z_t\}$, where

$$Z_t = X_t - Y_t\ ,$$

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

Mean of the series μ_{Xt} = 105.78

ω_1= 0.5236 and corresponding α_1 = 29.28, β_1 = 172.93

ω_2= 1.0472 and corresponding α_2 = -102.6, β_2 = 56.79

Example – 2 (contd.)

For t=1,

$$Y_1 = 105.78 + 29.28\cos\left(0.5236\times 1\right) + 172.93\sin\left(0.5236\times 1\right)$$
$$+\left(-102.6\right)\cos\left(1.0472\times 1\right) + 56.79\sin\left(1.0472\times 1\right)$$
$$= 215.5$$

$Z_1 = X_1 – Y_1$

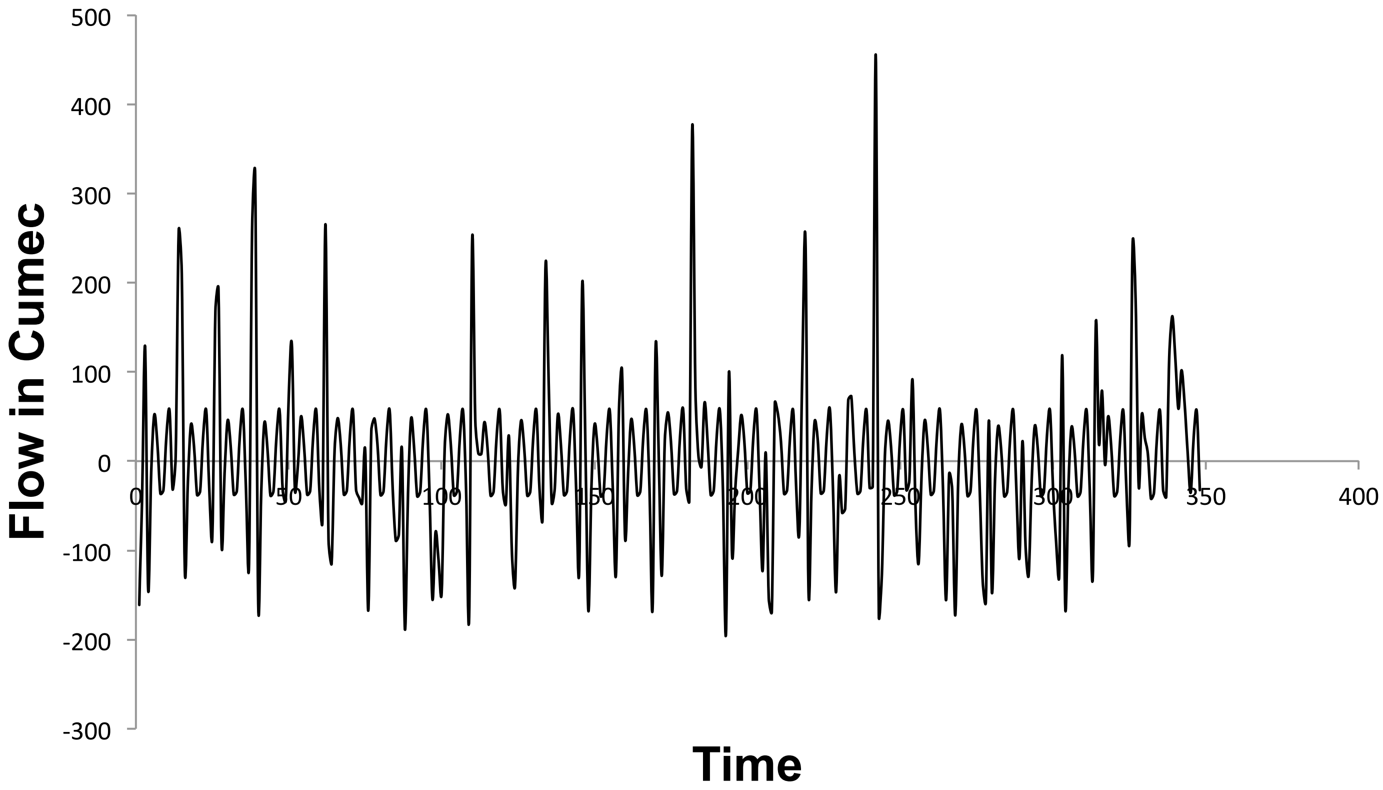
$= 54.6 – 215.5$

$= -160.9$

And so on….

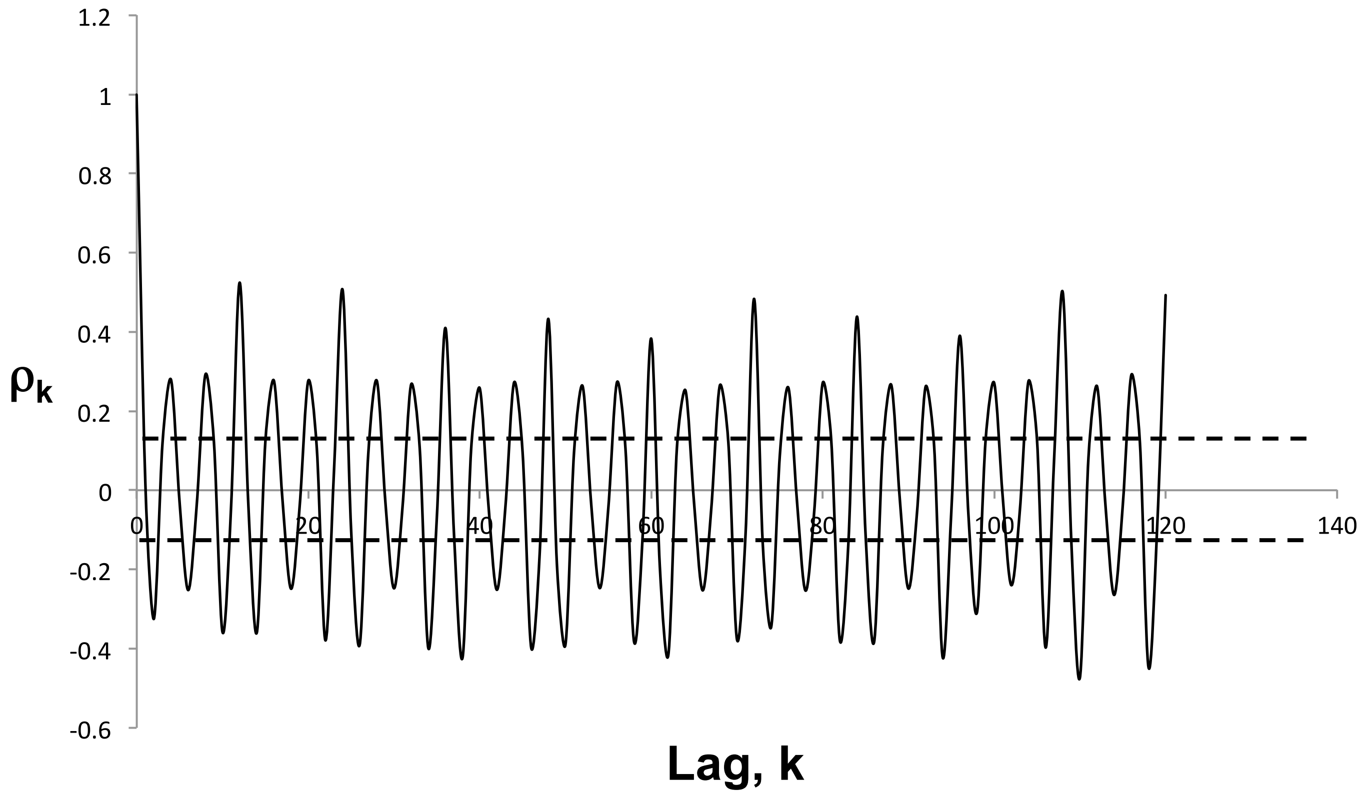
Example – 2 (contd.)

Time series plot of Z_t,



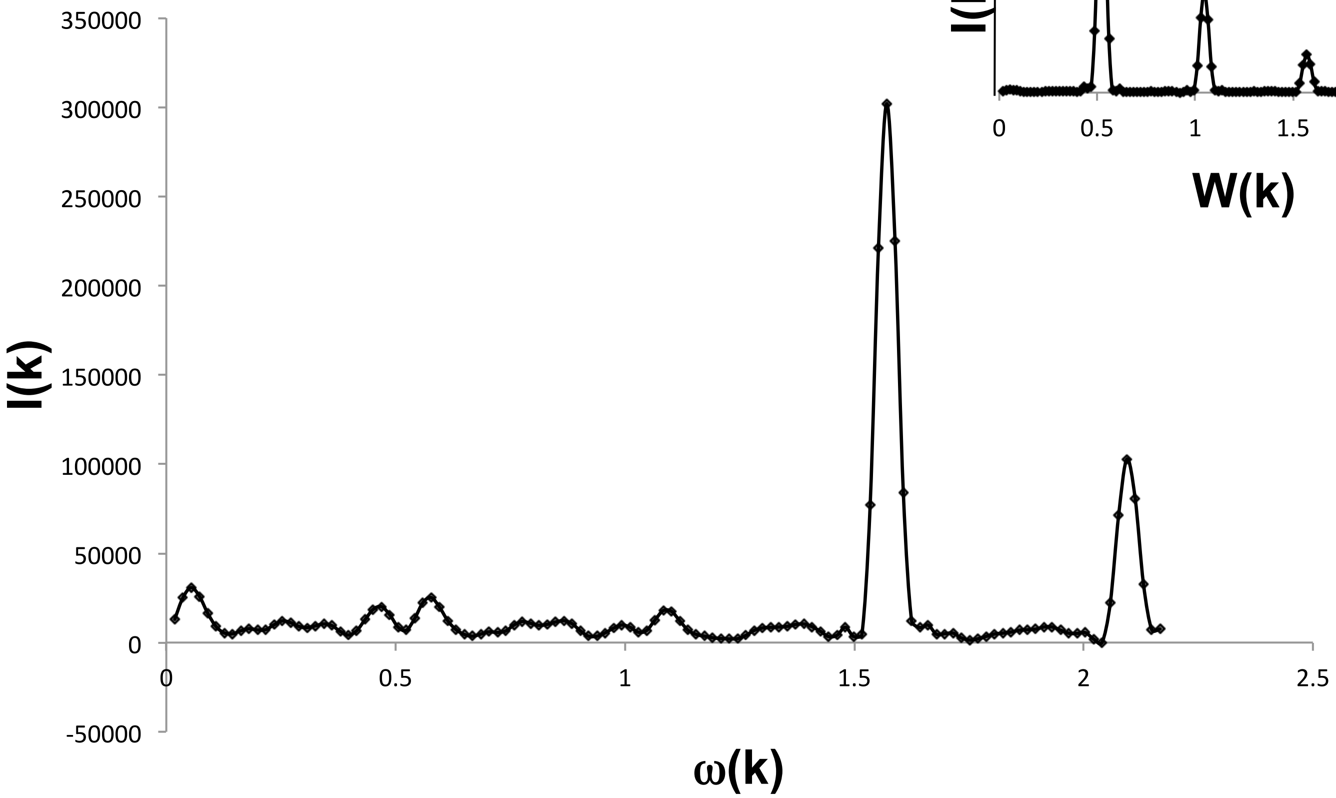
Example – 2 (contd.)

Correlogram of Z_t,

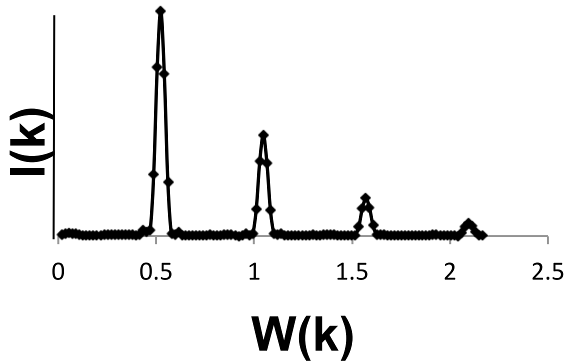


Example – 2 (contd.)

Power Spectrum of Z_t,



Power Spectrum of original series



Example – 2 (contd.)

- 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 – 2 (contd.)

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

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

$$= 105.78$$

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

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

F(2, 346) = 3.0

Example – 2 (contd.)

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

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 – 2 (contd.)

- 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 – 2 (contd.)

For the first value (June month),

$$Z_1 = \frac{(54.6 - 117.49)}{52.24} = -1.204$$

Second value (July month)

$$Z_2 = \frac{(325.4 - 474.5)}{150.18} = -0.993$$

Third value (August month)

$$Z_3 = \frac{(509.5 - 421.39)}{126.53} = 0.696$$

And so on….

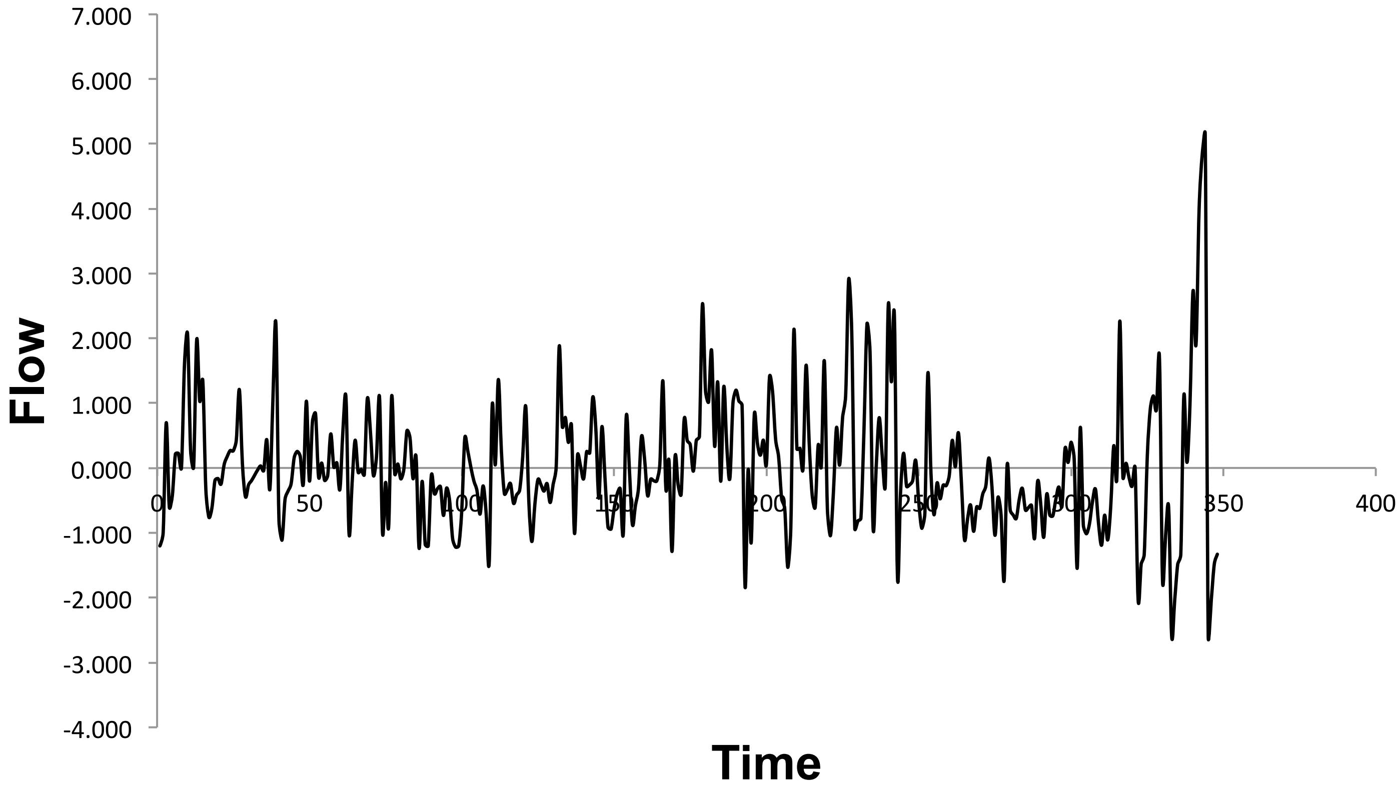
Example – 2 (contd.)

- 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

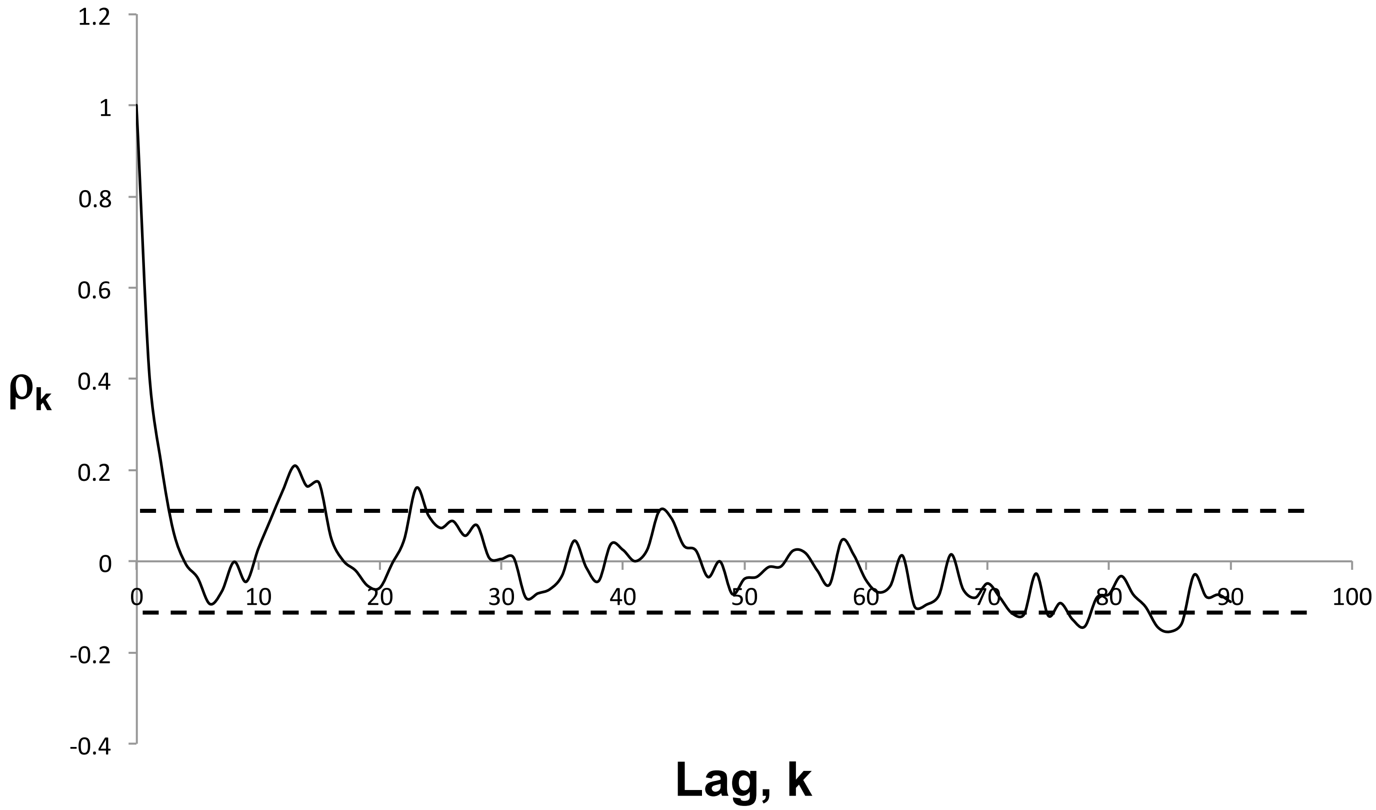
Example – 2 (contd.)

Time series of standardized data.



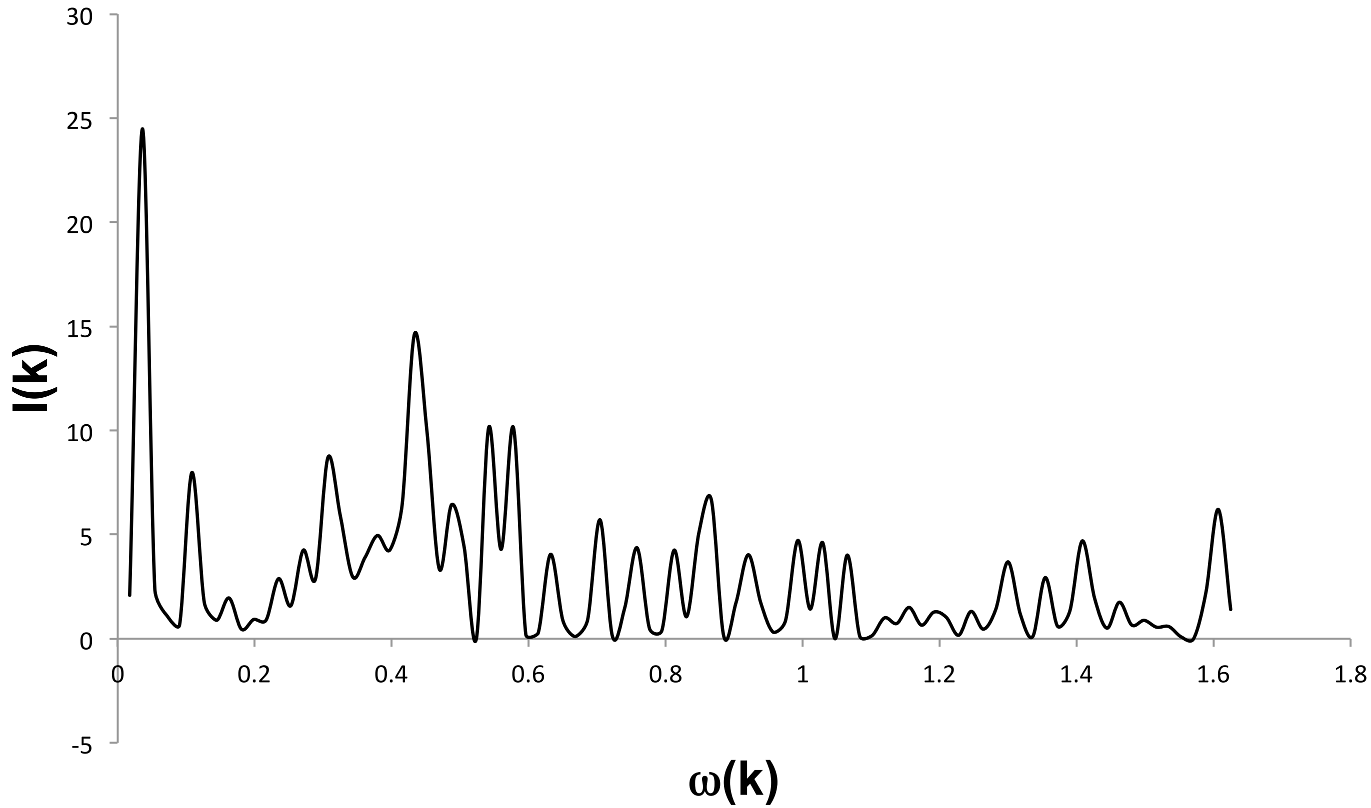
Example – 2 (contd.)

Correlogram of standardized data.



Example – 2 (contd.)

Spectrum of standardized data.



Example – 2 (contd.)

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

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

The periodicities are insignificant and the time series is purely stochastic.