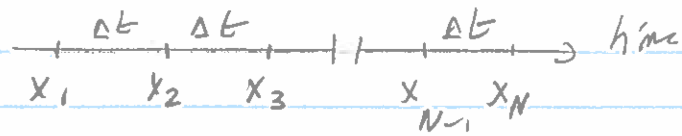
A (VERY!) SHORT INTRODUCTION TO SPECTRAL ANALYSIS

- GOAL: DECOMPOSE A TIME SERIES INTO FREQUENCY CONSTITUENTS

- TIME SERIES: COLLECTION OF OBS. MADE SEQUENTIALLY IN TIME

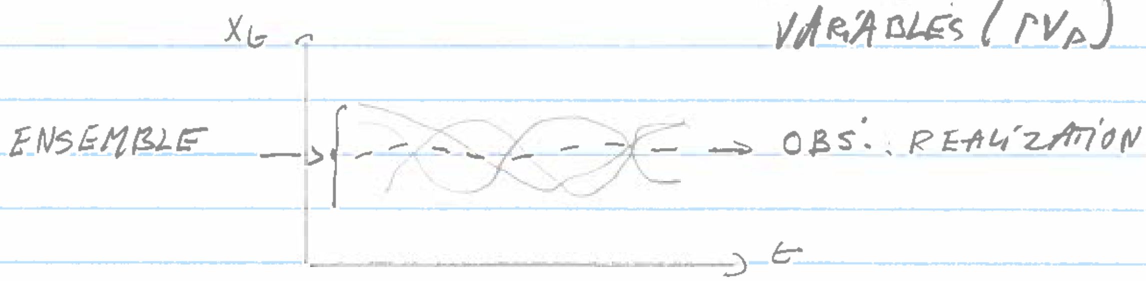
- DISCRETE TIME SERIES: $x_1, x_2, \ldots, x_N$: OBS. AVAILABLE ONLY AT DISCRETE TIMES



WE WILL ASSUME Δt = CONST. ($\Delta t = 1$)

- DETERMINISTIC VS. INDETERMINISTIC TIME SERIES

- STOCHASTIC (RANDOM) PROCESS: $\{x_t\}$: ORDERED SET OF RANDOM VARIABLES (RVs)



1. INTRODUCTION: A FEW STATISTICAL CONCEPTS

X_t IS A RV, SO IT HAS A PROBABILITY DENSITY FUNCTION WITH MEAN, VARIANCE, etc.

MEAN: $E[X_t] = \mu_t$ (1a)

VARIANCE: $VAR[X_t] = E[(X_t - \mu_t)^2]$ (1b)

AUTOCOVARIANCE FUNCTION (acvf)

$$COV[X_t, X_{t+k}] = E[(X_t - \mu_t)(X_{t+k} - \mu_{t+k})] \quad (1c)$$

K: LAG

IF $K=0$: $COV[X_t, X_t] = E[(X_t - \mu_t)^2] = VAR[X_t]$ (1d)

STATIONARY STOCHASTIC PROCESS: (2nd STATIONARY):

$$E[X_t] = \mu \quad (2a)$$

$$VAR[X_t] = \sigma_X^2 \quad (2b)$$

$$COV[X_t, X_{t+k}] = \gamma(k) \quad (2c)$$

MOST OF THEORY ASSUMES 2nd ORDER STATIONARITY

AUTOCORRELATION FUNCTION (acf)

$$\rho(k) = \gamma(k)/\gamma(0) \quad \text{(def. FOR STATIONARY STOCH. PROCESS)} \quad (2d)$$

THEORETICAL ACVF: $\gamma(k) = E\left[(X_t - \mu)(X_{t+k} - \mu)\right]$ (3a)

SAMPLE ACVF: $c_k = \frac{1}{N} \sum_{t=1}^{N-k} (x_t - \bar{x})(x_{t+k} - \bar{x})$ (3b)

WHERE

$$\bar{x} = \frac{1}{N} \sum_{t=1}^{N} x_t \quad (4)$$

OF COURSE,

$c_0 = \frac{1}{N} \sum_{t=1}^{N} (x_t - \bar{x})^2$ IS THE VARIANCE OF THE TIME SERIES

THEORETICAL ACF: $\rho(k) = \gamma(k) / \gamma(0)$ (5a)

SAMPLE ACF: $r_k = c_k / c_0$ (5b)

2. ESTIMATION OF THE SPECTRUM OF A TIME SERIES

FOR A DISCRETE STATIONARY STOCHASTIC PROCESS:

$$\text{VARIANCE} = VAR[X_t] = \sigma_x^2 = \int_0^{\pi} f(\omega)\, d\omega \qquad (6)$$

$f(\omega)$: SPECTRAL DENSITY FUNCTION OR SPECTRUM

ω: (ANGULAR) FREQUENCY

$\omega = \pi$ IS THE NYQUIST FREQUENCY

$(\Delta t = 1)$

$T = 2, \quad \omega = \frac{2\pi}{T} = \pi \qquad (T = \text{PERIOD})$

CASE 1: SINGLE SINUSOIDAL COMPONENT

$$x_t = a_0 + R\cos(\omega t + \phi) + Z_t \qquad (7)$$

a_0: CONST.; R: AMPLITUDE; ω: FREQUENCY; ϕ: (INITIAL) PHASE

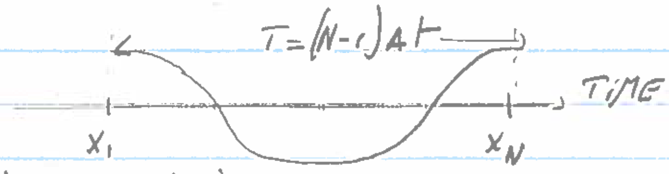
Z_t IS A PURELY RANDOM PROCESS: $E[Z_t] = \text{CONST.}$ (=0 HERE)

$VAR[Z_t] = \text{CONST}$

$COV[Z_t, Z_{t+k}] = 0$ for all $k \neq 0$

CASE 2: MULTIPLE SINUSOIDAL COMPONENTS

$$X_t = a_0 + \sum_{p=1}^{M} R_p \cos(\omega_p t + \phi_p) + z_t \qquad (8)$$



- ω_1: LOWEST FREQUENCY

$$T_{MAX} = (N-1)\Delta t \simeq N\Delta t \quad (N \gg 1)$$

$$\text{so} \quad \omega_1 = \frac{2\pi}{T_{MAX}} = \frac{2\pi}{N} \quad (\Delta t = 1)$$

- $\omega_M = \pi$ (Nyquist frequency)

- DEFINE $\omega_p = \frac{2\pi}{N} p$, $p = 1, 2, \ldots, \frac{N}{2}$. THEN (8) BECOMES

We assume N is even

$$X_t = a_0 + \sum_{p=1}^{N/2} R_p \cos(\omega_p t + \phi_p) + z_t \qquad (9)$$

Now $R_p \cos(\omega_p t + \phi_p) = R_p \cos\omega_p t \cos\phi_p - R_p \sin\omega_p t \sin\phi_p$

Let $a_p = R_p \cos\phi_p$, $b_p = -R_p \sin\phi_p$ so $R_p = \sqrt{a_p^2 + b_p^2}$, $\tan\phi_p = -b_p/a_p$

THEN, (9) BECOMES

$$X_t = a_0 + \sum_{p=1}^{N/2} (a_p \cos\omega_p t + b_p \sin\omega_p t) + Z_t \qquad (10)$$

$$X_t = a_0 + \sum_{p=1}^{\frac{N}{2}-1} (a_p \cos\omega_p t + b_p \sin\omega_p t) + a_{N/2} \cos\pi t + z_t$$

SINCE $\sin(\omega_{N/2} t) = \sin\left(\frac{2\pi}{N} \frac{N}{2} t\right) = \sin(\pi t) = 0$

LET US MAKE A COUNT: DATA: $N \; (x_t)$

PARAMETERS: $1 \; (a_0)$
$+ \frac{N}{2} - 1 \; (a_p)$
$+ \frac{N}{2} - 1 \; (b_p)$
$+ \; 1 \; a_{N/2}$
N

SO ⑩ is A SYSTEM OF ⑩ EQUATIONS WITH ⑩ UNKNOWS.

AS A RESULT, $z_t = 0$ in ⑨ – ⑩ (PERFECT FIT):

$$x_t = a_0 + \sum_{p=1}^{\frac{N}{2}-1} \left(a_p \cos \omega_p t + b_p \sin \omega_p t \right) + a_{N/2} \cos \pi t \qquad (11)$$

This is A FINITE FOURIER SERIES REPRESENTATION OF x_t

THE FOURIER COEFFICIENTS (a, b) ARE OBTAINED FROM

$$a_0 = \frac{1}{N} \sum_{t=1}^{N} x_t \quad (= \bar{x}) \qquad (12a)$$

$$a_p = \frac{2}{N} \sum_{t=1}^{N} x_t \cos \omega_p t \qquad (12b)$$

$$b_p = \frac{2}{N} \sum_{t=1}^{N} x_t \sin \omega_p t \qquad (12c)$$

$$p = 1, 2, \ldots, \frac{N}{2} - 1 \quad \text{(for 12b, 12c)}$$

$$a_{N/2} = \frac{1}{N} \sum_{t=1}^{N} x_t \cos \pi t \qquad (12d)$$

where

$$\omega_p = \frac{2\pi}{N} p$$

PARSEVAL'S THEOREM

EQ. (11) CAN BE WRITTEN AS

$$x_t = a_0 + \sum_{p=1}^{\frac{N}{2}-1} R_p \cos(\omega_p t + \phi_p) + a_{N/2} \cos \pi t$$

IT CAN BE SHOWN THAT

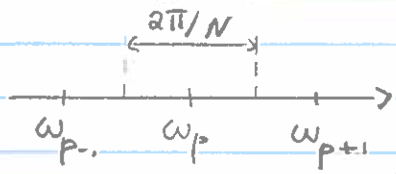
$$\frac{1}{N} \sum_{t=1}^{N} (x_t - \bar{x})^2 = \sum_{p=1}^{\frac{N}{2}-1} \frac{R_p^2}{2} + a_{N/2}^2 \qquad (13)$$

THIS IS PARSEVAL'S THEOREM. IT SHOWS HOW THE VARIANCE OF x_t IS PARTITIONED INTO DIFFERENT FREQUENCY COMPONENTS

Note: R_p : AMPLITUDE OF p^{th} HARMONIC

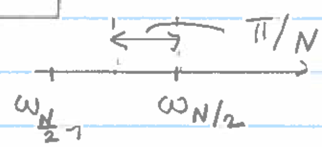
PERIODOGRAM ~ ANALOG OF HISTOGRAM

- $\frac{R_p^2}{2}$ = AREA OF HISTOGRAM RECTANGLE
  = HEIGHT × $\frac{2\pi}{N}$ (HEIGHT = $I(\omega_p)$)



SO $$I(\omega_p) = \frac{N}{4\pi} R_p^2 \quad , p = 1, 2, \ldots, \frac{N}{2}-1 \qquad (14a)$$

- FOR $p = \frac{N}{2}$



$$a_{N/2}^2 = I(\omega_p) \times \frac{\pi}{N} \quad \text{so} \quad I(\omega_p) = \frac{N}{\pi} a_{N/2}^2 \qquad (14b)$$

Note: $$\sum_{p=1}^{\frac{N}{2}-1} I(\omega_p) \frac{2\pi}{N} + I(\pi) \frac{\pi}{N} = \frac{1}{N} \sum_{t=1}^{N} (x_t - \bar{x})^2$$

$I(\omega_p)$ IS THE PERIODOGRAM

$I(\omega_p)$ CAN ALSO BE COMPUTED FROM

$$I(\omega_p) = \frac{1}{\pi}\left(c_0 + 2\sum_{k=1}^{N-1} c_k \cos\omega_p k\right) \quad (15)$$

WHERE

$$c_k = \frac{1}{N}\sum_{t=1}^{N-1}(x_t-\bar{x})(x_{t+k}-\bar{x})$$

ARE THE ACV COEFFICIENTS OF THE TIME SERIES

Note. $I(\omega_p) = \frac{1}{\pi}\sum_{k=-(N-1)}^{N-1} c_k e^{-i\omega_p k}$ with $c_k=0$ for $|k|>N$

PROPERTIES OF PERIODOGRAM

$\lim_{N\to\infty} E[I(\omega)] = f(\omega)$: PERIODOGRAM IS ASYMPTOTICALLY UNBIASED

$\lim_{N\to\infty} VAR[I(\omega)] \neq 0$: PERIODOGRAM IS NOT CONSISTENT ESTIMATOR FOR $f(\omega)$

SPECTRAL ANALYSIS: EXAMPLE OF CONSISTENT ESTIMATION METHOD

BASIC IDEA: WEIGHT AND TRUNCATE ACV COEF. IN (15):

$$\hat{f}(\omega) = \frac{1}{\pi}\left(\lambda_0 c_0 + 2\sum_{k=1}^{M}\lambda_k c_k \cos\omega k\right) \quad (16)$$

- THE WEIGHTS $\{\lambda_k\}$ ARE A SET CALLED THE LAG WINDOW
- M IS THE TRUNCATION POINT $(M < N-1)$

THE TWO BEST-KNOWN LAG WINDOWS ARE:

1. TUKEY WINDOW

$$\lambda_k = \frac{1}{2}\left(1 + \cos\frac{k\pi}{M}\right), \quad k = 0, 1, \ldots, M \qquad (17)$$

THIS WINDOW IS ALSO KNOWN AS TUKEY–HANNING OR BLACKMAN–TUKEY WINDOW

2. PARZEN WINDOW

$$\lambda_k = \begin{cases} 1 - 6\left(k/M\right)^2 + 6\left(k/M\right)^3, & 0 \le k \le M/2 \quad (18a) \\ 2\left(1 - k/M\right)^3, & M/2 \le k \le M \quad (18b) \end{cases}$$

USEFUL ROUGH GUIDE: $M = 2\sqrt{N}$ (19a)

ω USUALLY TAKEN AS $\omega_j = \pi \frac{j}{M}$ (19b)

CONFIDENCE INTERVAL FOR $f(\omega)$

As $N \to \infty$, $\frac{\nu \hat{f}(\omega)}{f(\omega)}$ is APPROXIMATELY DISTRIBUTED AS χ^2_ν

WHERE ν IS THE NUMBER OF DEGREES OF FREEDOM:

$$\nu(\text{TUKEY}) = 2.67\frac{N}{M}, \quad \nu(\text{PARZEN}) = 3.71\frac{N}{M} \qquad (20a,b)$$

THUS, $\text{prob.}\left(\chi^2_{\nu,1-\alpha/2} < \nu \hat{f}(\omega)/f(\omega) < \chi^2_{\nu,\alpha/2}\right) = 1-\alpha$

SO $100(1-\alpha)\%$ CONFIDENCE INTERVAL FOR $f(\omega)$ IS

$$\left| \frac{\nu \hat{f}(\omega)}{\chi^2_{\nu,\alpha/2}} \text{ to } \frac{\nu \hat{f}(\omega)}{\chi^2_{\nu,1-\alpha/2}} \right| \qquad (21)$$

BANDWIDTH

DEFINE FIRST SPECTRAL WINDOW AS:

$$k(\omega) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \lambda_k e^{-i\omega k} \quad \text{(FOURIER TRANSFORM)} \tag{22a}$$

WHERE

$$\lambda_k = \lambda_{-k} \text{ AND } \lambda_k = 0 \text{ FOR } k > M$$

THEN

$$\lambda_k = \int_{-\pi}^{\pi} k(\omega) e^{i\omega k} d\omega \quad \text{(INVERSE FOURIER TRANSFORM)} \tag{22b}$$

IT CAN THEN BE SHOWN THAT

$$\hat{f}(\omega_0) = \int_{-\pi}^{\pi} k(\omega) I(\omega_0 - \omega) d\omega \tag{23a}$$

SO, AS $N \to \infty$,

$$E[\hat{f}(\omega_0)] = \int_{-\pi}^{\pi} k(\omega) f(\omega_0 - \omega) d\omega \tag{23b}$$

THE NAME "WINDOW" ARISES FROM THE FACT THAT $k(\omega)$ DETERMINES THE PART OF PERIODOGRAM "SEEN" BY THE ESTIMATOR

Note: $\lambda_0 = 1 = \int_{-\pi}^{\pi} k(\omega) d\omega$ AS DESIRED (24)

THE BANDWIDTH IS ROUGHLY SPEAKING THE WIDTH OF THE SPECTRAL WINDOW

BANDWIDTH (TUKEY) $= \frac{8\pi}{3M}$, BANDWIDTH (PARZEN) $= \frac{2\pi(1.86)}{M}$

OTHER ASPECTS

1. Use (16) to get $\hat{f}(\omega)$ AND MULTIPLY $\hat{f}(\omega)$ by Δt OF TIME SERIES TO GET SPECTRAL ESTIMATES WITH UNITS OF

$$[X_t]^2 [T]$$

$[T]$ — means unit of time

$[\ ]$ — means "units of"

2. $\Delta t\, \hat{f}(\omega) = \Delta t\, \hat{f}(2\pi/T)$ COULD BE PLOTTED AGAINST $1/T$ (number of cycles per unit time) RATHER ω

3. IF LOG SCALE IS USED FOR $\hat{f}(\omega)$ (OR $\Delta t\, \hat{f}(\omega)$), THEN CONFIDENCE INTERVALS FOR $f(\omega)$ COULD BE SHOWN BY A SINGLE VERTICAL LINE

4. ALWAYS SPECIFY LAG WINDOW USED, TRUNCATION POINT USED, AND CONFIDENCE INTERVALS (THIS ON LOG SCALE, see POINT 3 ABOVE)

A (VERY) SHORT INTRODUCTION TO SPECTRAL ANALYSIS

Figures

Main reference

Chatfield, C., *The analysis of time series – An Introduction*, Chapman & Hall, 1996

See also:

Jenkins G. & Watts D., *Spectral analysis and its applications*, Holden-Day, 1968

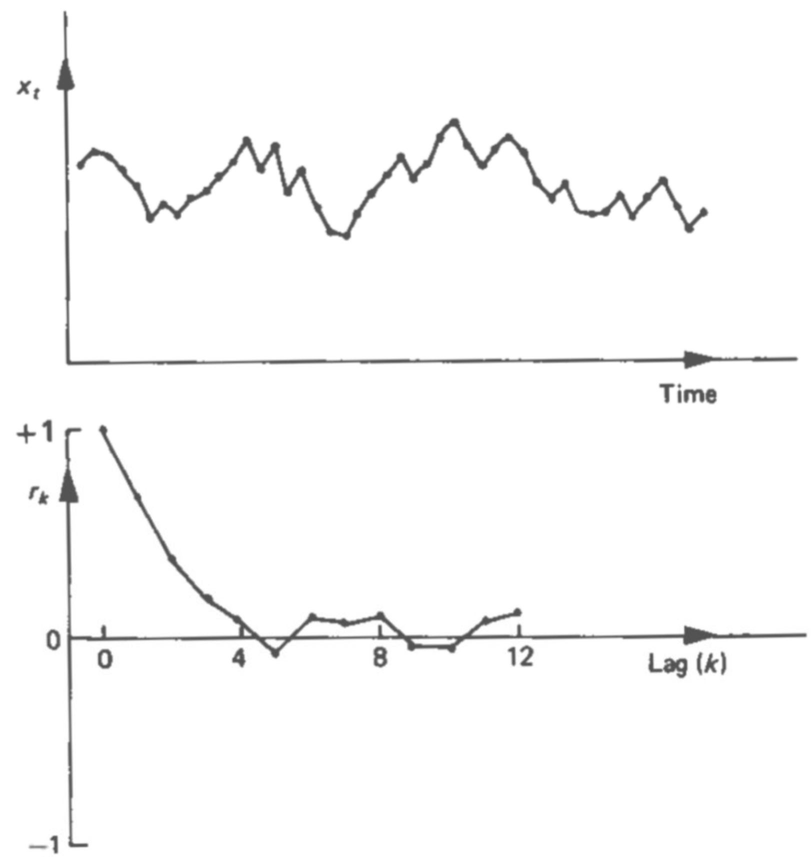


Figure 2.1 A time series showing short-term correlation together with its correlogram.

Chatfield (1996)

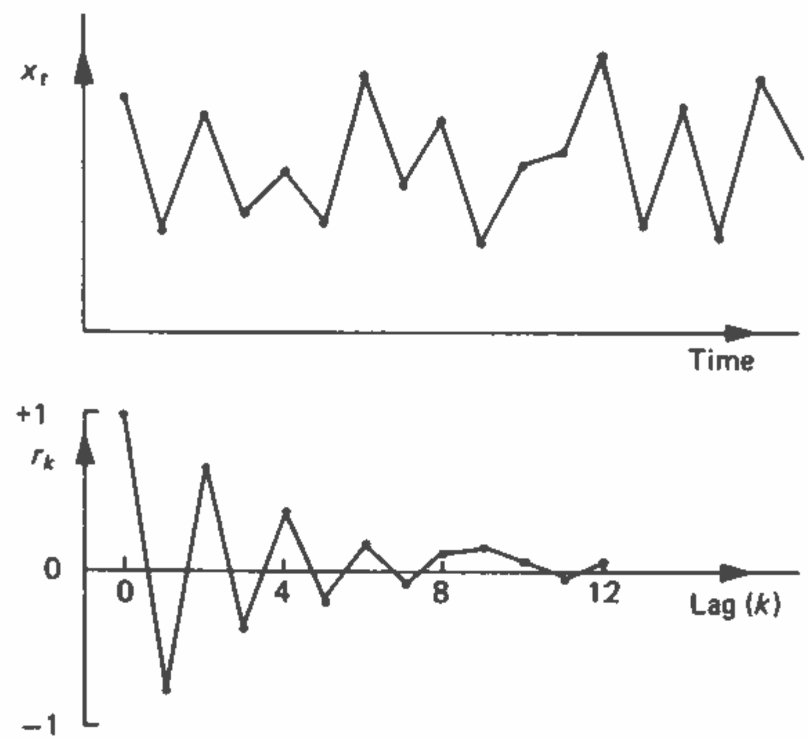


Figure 2.2 An alternating time series together with its correlogram.

Chatfield (1996)

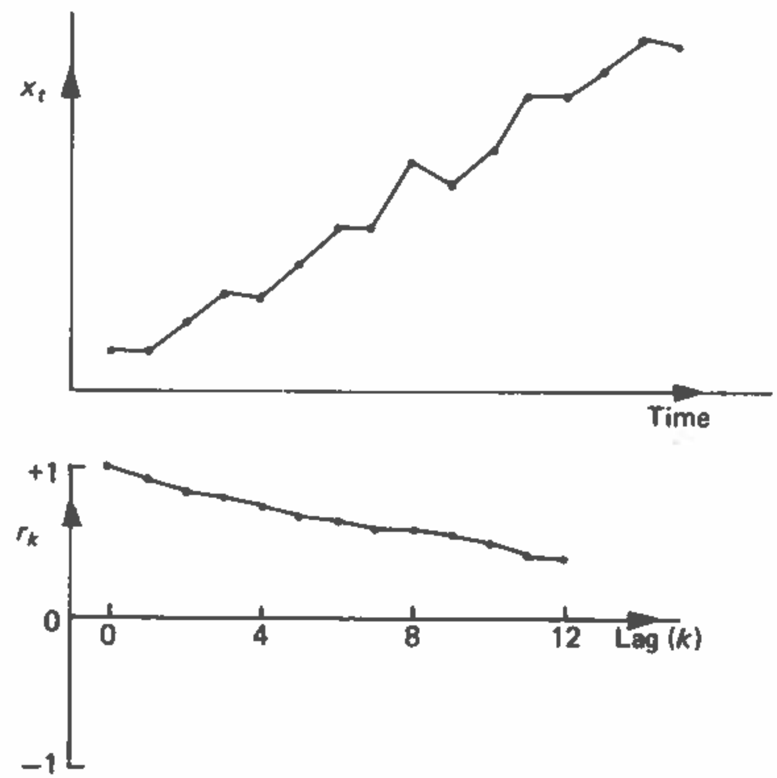


Figure 2.3 A non-stationary time series together with its correlogram.

Chatfield (1996)

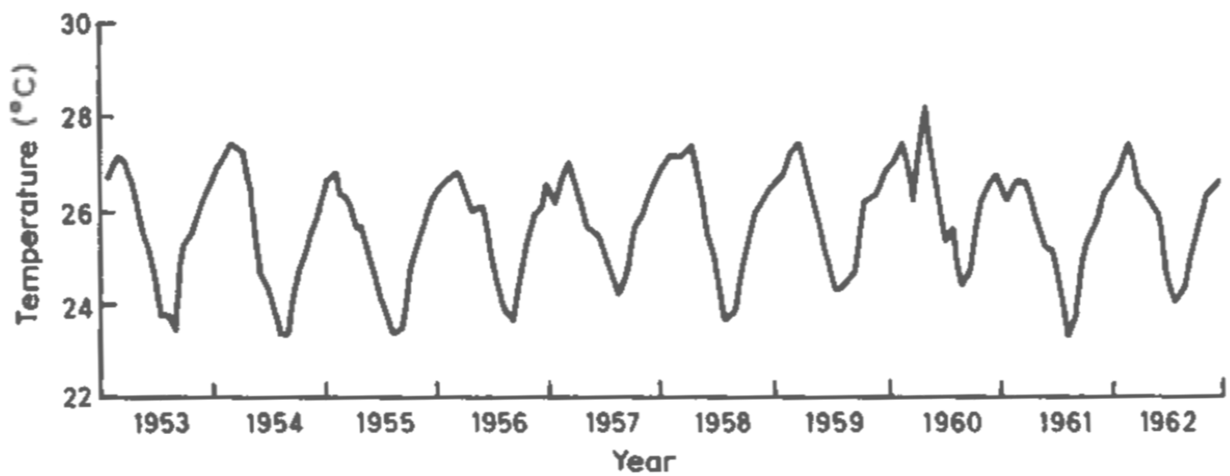


Figure 1.2 Average air temperature at Recife, Brazil, in successive months.

Chatfield (1996)

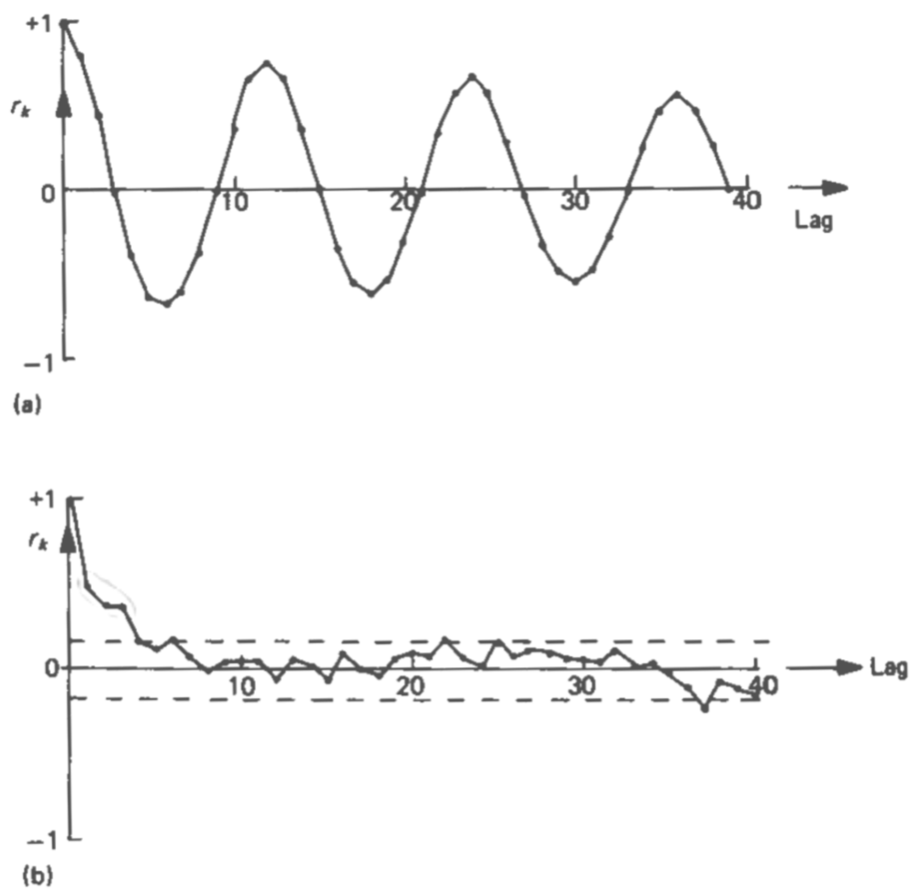


Figure 2.4 The correlogram of monthly observations on air temperature at Recife: (a) for the raw data; (b) for the seasonally adjusted data. The dotted lines in (b) are at $\pm 2/\sqrt{N}$. Values outside these lines are significantly different from zero.

Chatfield (1996)

(a) *Tukey window*

$$\lambda_k = \frac{1}{2}\left(1 + \cos\frac{\pi k}{M}\right) \qquad k = 0, 1, \ldots, M$$

This window is also called the Tukey-Hanning or Blackman-Tukey window.

(b) *Parzen window*

$$\lambda_k = \begin{cases} 1 - 6\left(\frac{k}{M}\right)^2 + 6\left(\frac{k}{M}\right)^3 & 0 \leqslant k \leqslant M/2 \\ 2(1 - k/M)^3 & M/2 \leqslant k \leqslant M \end{cases}$$

These two windows are illustrated in Figure 7.1 with $M = 20$.

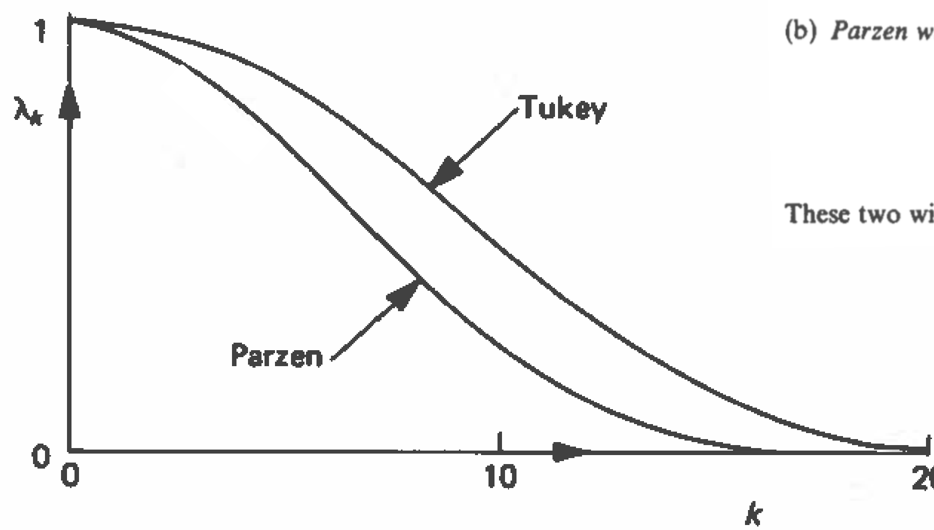


Figure 7.1 The Tukey and Parzen lag windows with $M = 20$.

Chatfield (1996)

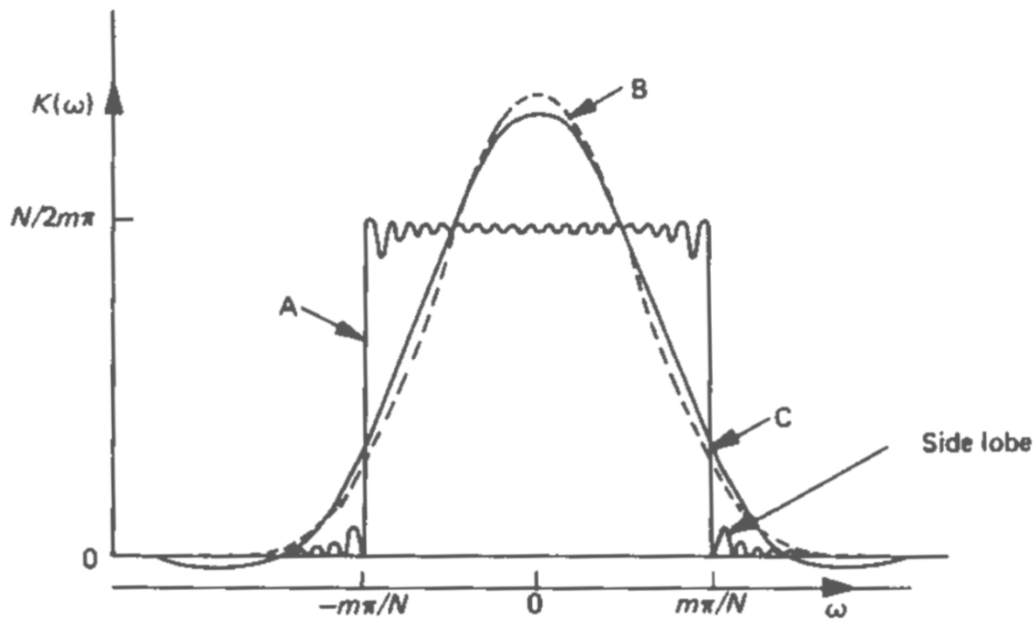


Figure 7.2 The spectral windows for three common methods of spectral analysis: A, smoothed periodogram ($m=20$); B, Parzen ($M=93$); C, Tukey ($M=67$); all with $N=1000$.

Chatfield (1996)

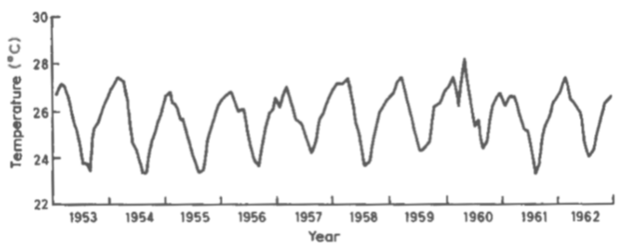


Figure 1.2 Average air temperature at Recife, Brazil, in successive months.

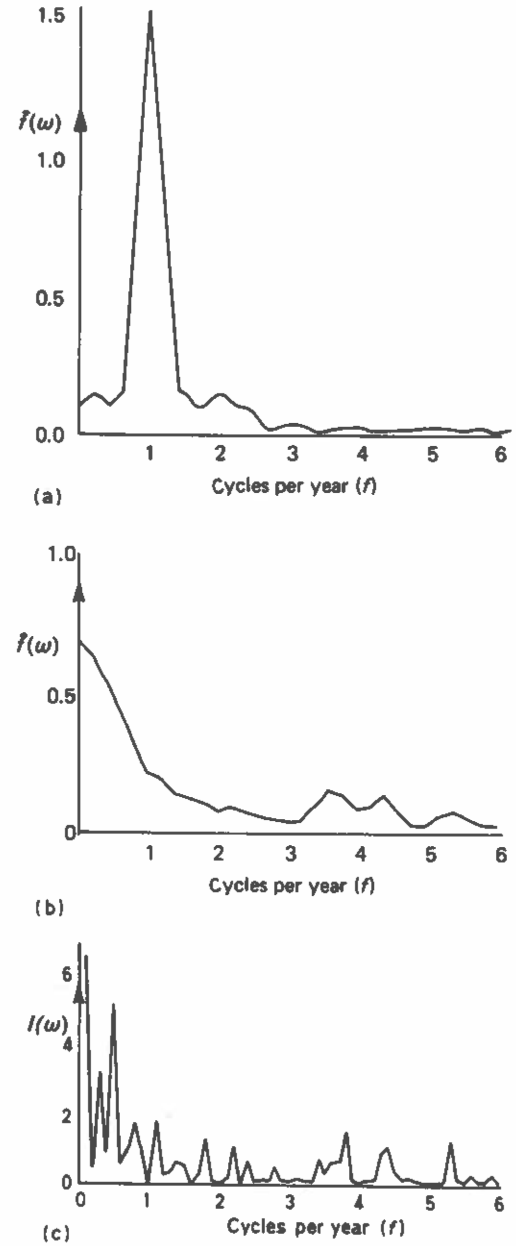


Figure 7.5 Spectra for average monthly air temperature readings at Recife, (a) for the raw data; (b) for the seasonally adjusted data using the Tukey window with $M=24$; (c) the periodogram of the seasonally adjusted data is shown for comparison.

Chatfield (1996)

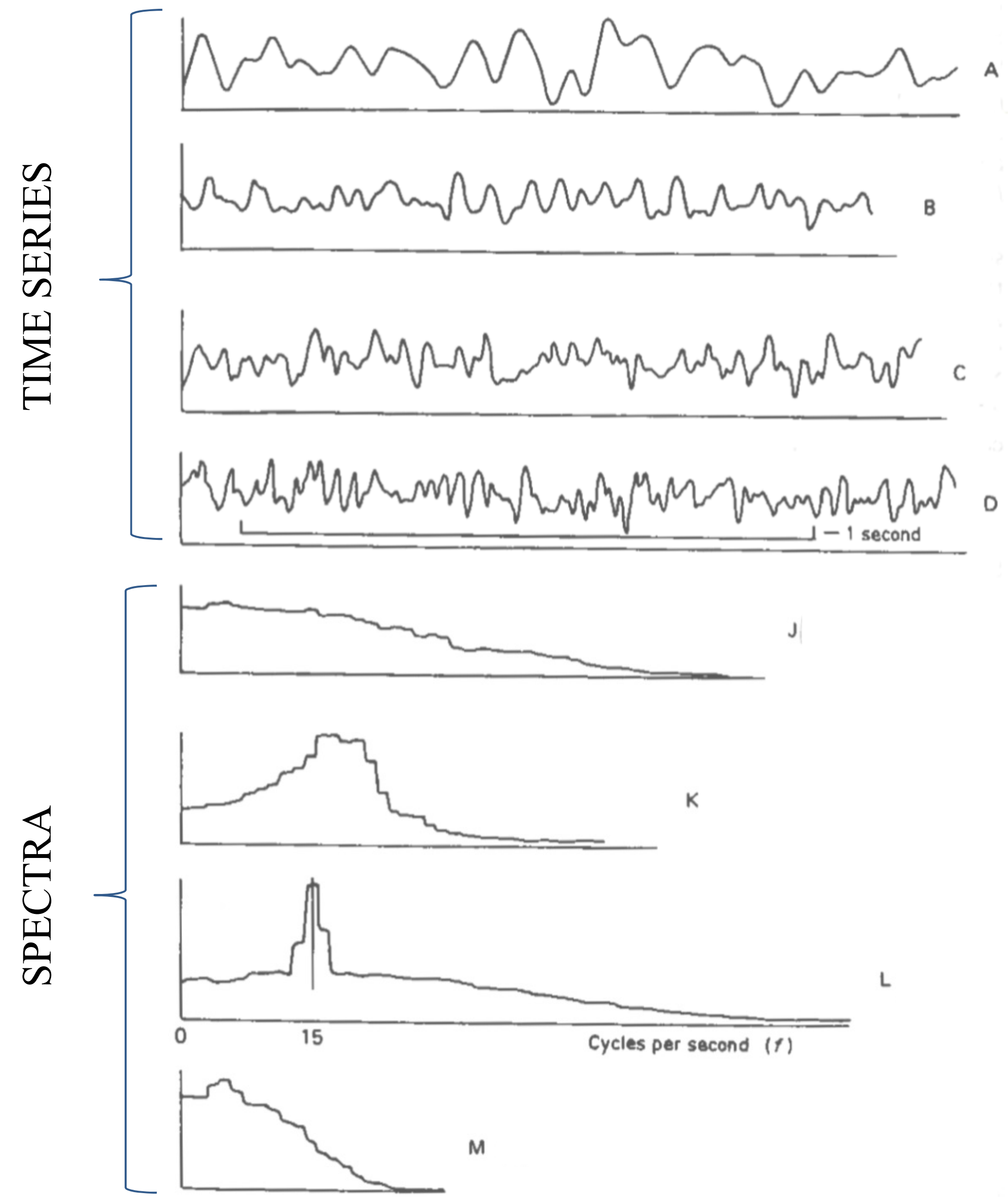


Figure 7.4 Four time series and their spectra. The spectra are given in random order.

Chatfield (1996)

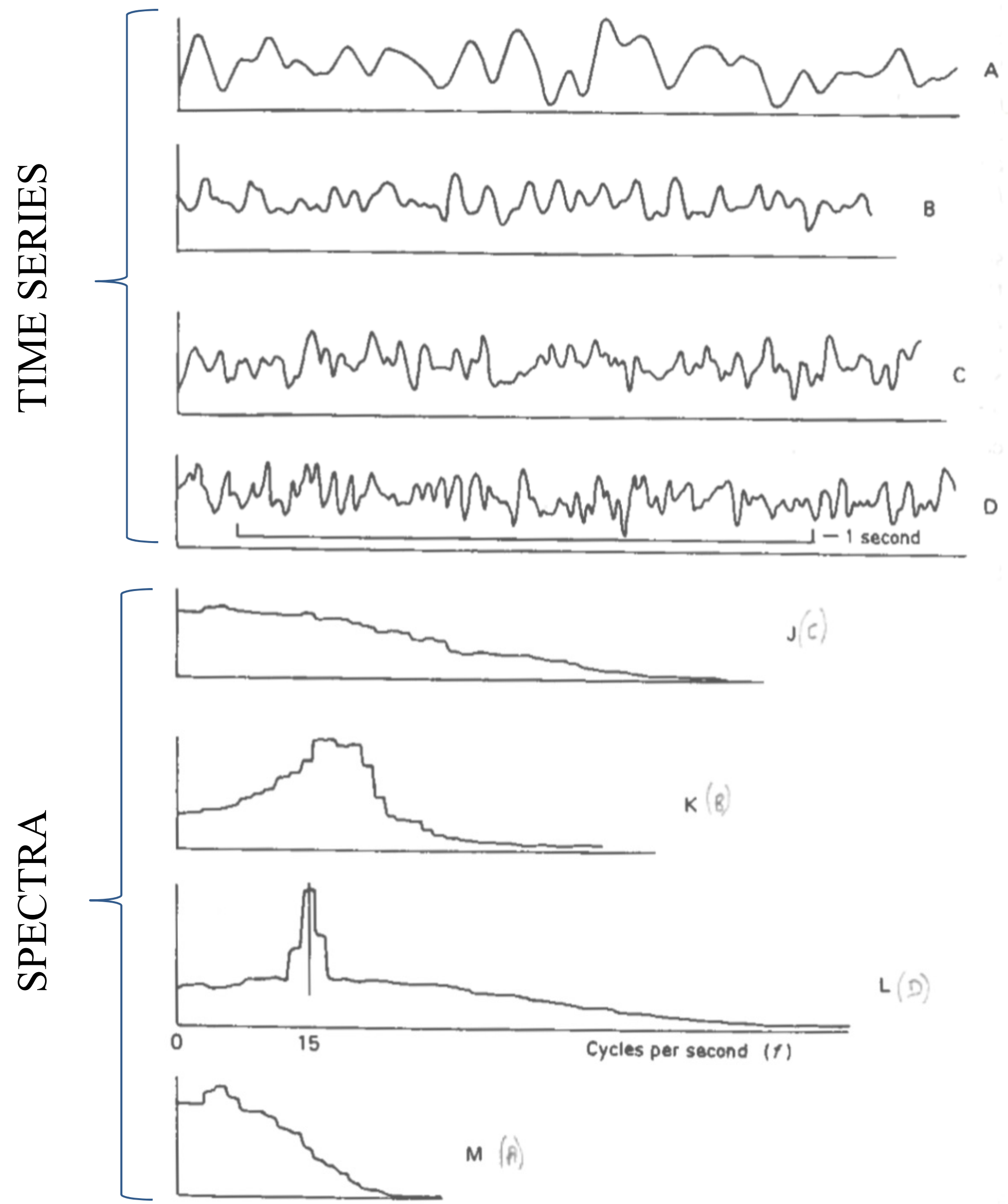


Figure 7.4 Four time series and their spectra. The spectra are given in random order.

Chatfield (1996)

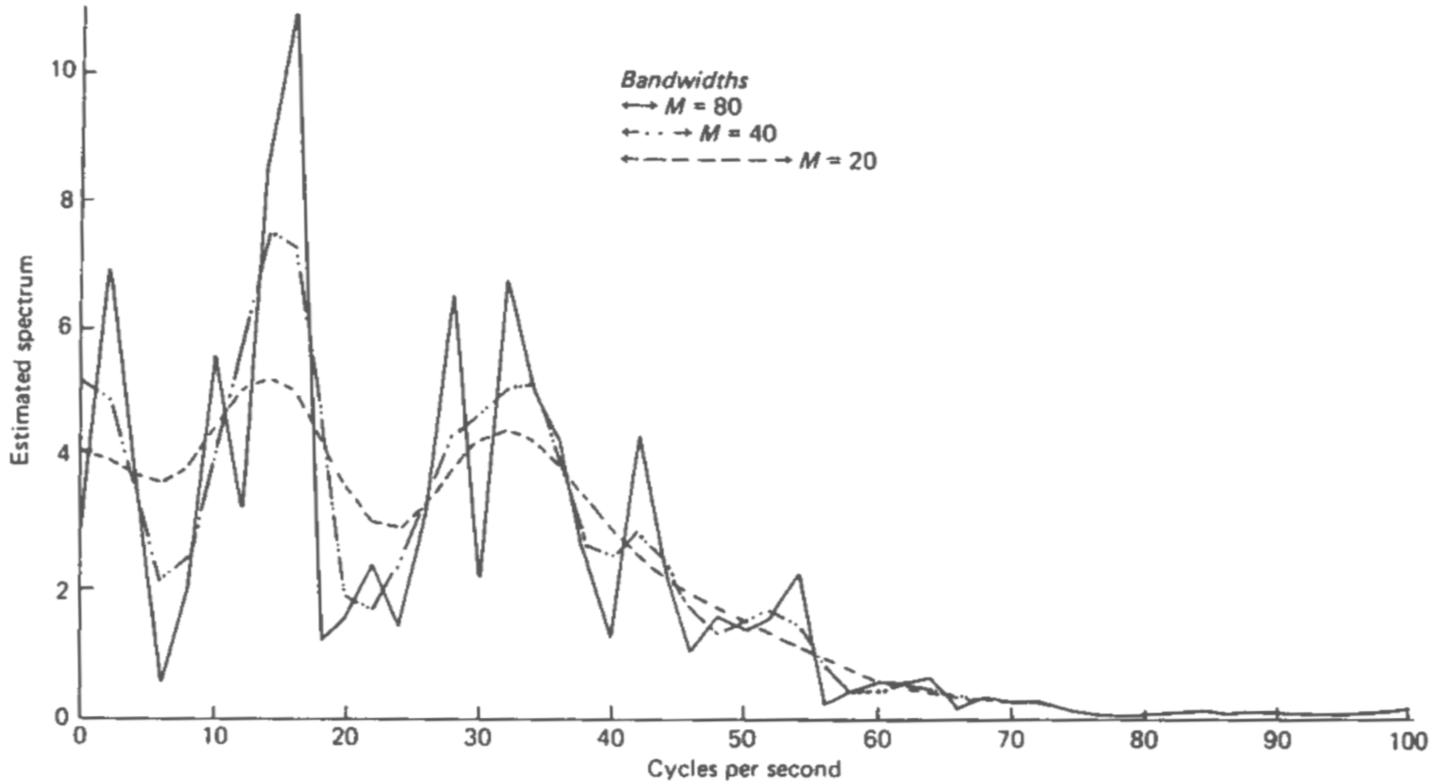


Figure 7.7 Estimated spectra for graph D of Figure 7.4 using the Tukey window with, (a) $M=80$; (b) $M=40$; (c) $M=20$.

Chatfield (1996)

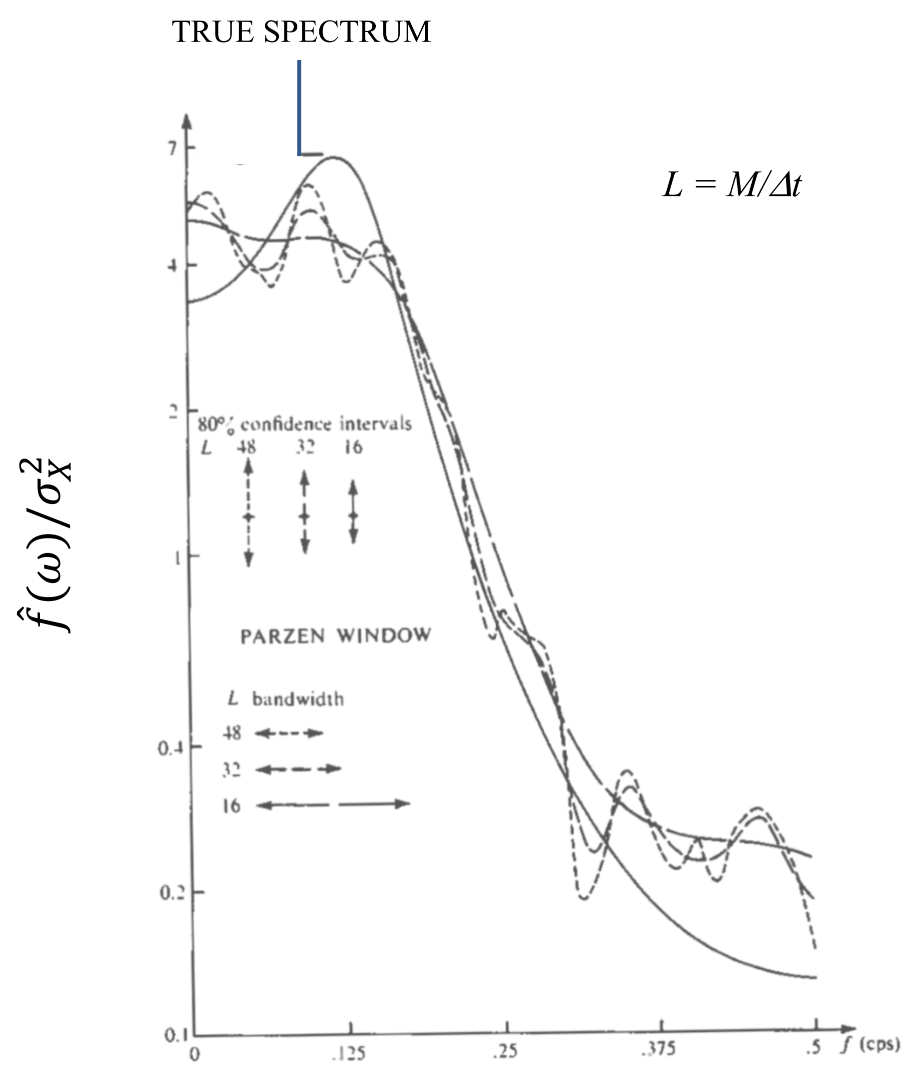


FIG. 7.9: Smoothed spectral density estimates for a second-order ar process ($\alpha_1 = 1.0, \alpha_2 = -0.5; N = 400$)

AR(2): $X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + Z_t$

where

$E(Z_t) = 0$
$Var(Z_t) = \sigma_Z^2$
$Cov(Z_t, Z_{t+k}) = 0$ for all $k \neq 0$

Jenkins & Watts (1968)