Chapter 3   Filtering and Spectrum analysis

3.1 Fourier analysis

With time series data, an alternative way to view the data is via the frequency representation first developed by J. B. Fourier in the early nineteenth century. Given a function y(t) defined on the interval [0,T], the Fourier series representation for y(t) is

$$\tilde{y}(t) = \frac{a_0}{2} + \sum_{m=1}^{\infty} [a_m \cos(\omega_m t) + b_m \sin(\omega_m t)], \qquad (3.1)$$

With the (angular) frequency ω_m given by

$$\omega_m = \frac{2\pi m}{T}, \qquad m = 1,2, \dots \qquad (3.2)$$

Also the frequency $f = \omega / 2\pi$ and $f_m = m / T$

The Fourier coefficients a_m and b_m given by

$$a_0 = \frac{2}{T} \int_0^T y(t) dt$$

$$a_m = \frac{2}{T} \int_0^T y(t) \cos(\omega_m t) dt \qquad (3.3)$$

$$b_m = \frac{2}{T} \int_0^\pi y(t) \sin(\omega_m t) dt$$

For a discrete time series, $y(t_n) \equiv y_n$, n=1, …, N. With a sample interval Δt, the samples are collected at time $t_n = n\Delta t$

$$y_n = \frac{a_0}{2} + \sum_{m=1}^{M} [a_m \cos(\omega_m t_n) + b_m \sin(\omega_m t_n)], \qquad (3.4)$$

where M is the largest integer <=N/2 and N is the length of sample.

The Fourier coefficients are given by

$$a_0 = \frac{2}{N}\sum_{n=1}^{N} y_n = 2\overline{y}$$

$$a_m = \frac{2}{N}\sum_{n=1}^{N} y_n \cos(\omega_m t_n) = \frac{2}{N}\sum_{n=1}^{N} y_n \cos(\frac{2\pi m}{N} n\Delta t) \qquad (3.5)$$

$$b_m = \frac{2}{N}\sum_{n=1}^{N} y_n \sin(\omega_m t_n) = \frac{2}{N}\sum_{n=1}^{N} y_n \sin(\frac{2\pi m}{N} n\Delta t)$$

The cosine and sine functions have orthogonality properties:

$$\sum_{n=1}^{N}\cos(\omega_l t_n)\cos(\omega_m t_n) = \frac{N}{2}\delta_{lm},$$

$$\sum \sin(\omega_l t_n)\sin(\omega_m t_n) = \frac{N}{2}\delta_{lm}, \qquad (3.6)$$

$$\sum \cos(\omega_l t_n)\sin(\omega_m t_n) = 0$$

where δ_{lm} is the Kronecker delta function (which equals when l=m, and 0, otherwise).

General transform of Fourier:

$$y(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} F(\omega)e^{i\omega t} d\omega \qquad (3.A)$$

$$F(\omega) = \int_{-\infty}^{\infty} x(t)e^{-i\omega t} dt \qquad (3.B)$$

3.2 The discrete spectrum

The variance of the time series y can be written in terms of its Fourier coefficients

$$\text{var}(y) = \frac{1}{N}\sum_{i=1}^{N}(y_n - \overline{y_n})^2 = \frac{1}{N}\sum_{n}(y_n - \frac{a_0}{2})^2$$

$$= \frac{1}{N}\sum_{n}[\sum_{m}(a_m \cos(\omega_m t_n) + b_m \sin(\omega_m t_n))]^2 \qquad (3.7)$$

Using (3.6), (3.7) can be expressed in terms of the Fourier coefficients

$$\text{var}(\boldsymbol{y}) = \frac{1}{2}\sum_{m}({\boldsymbol{a}_m}^2 + {\boldsymbol{b}_m}^2) \tag{3.8}$$

The autospectrum, also called the power spectrum, is defined as

$$S_m = \frac{T}{4\pi}({\boldsymbol{a}_m}^2 + {\boldsymbol{b}_m}^2) = \frac{N\Delta t}{4\pi}({\boldsymbol{a}_m}^2 + {\boldsymbol{b}_m}^2) \tag{3.9}$$

Since $\dfrac{N\Delta t}{4\pi}$ is a constant, the power spectrum is also defined as

$$S_m = \frac{1}{2}({\boldsymbol{a}_m}^2 + {\boldsymbol{b}_m}^2) \tag{3.9.1}$$

Thus (3.8) can be expressed as

$$\text{var}(\boldsymbol{y}) = \sum S_m \Delta\omega \tag{3.10}$$

with $\Delta\omega = \dfrac{2\pi}{T} = \dfrac{2\pi}{N\Delta t}$ $\tag{3.11}$

Hence, the spectrum S_m can be viewed as the variance or 'energy' in the ω_m frequency band (with bandwidth $\Delta\omega$), and the total variance var(y) can be computed by integrating S_m over all frequency bands. When S_m is plotted as a function of the frequency, peaks in S_m reveal the frequencies where the energy is relatively high.

From (3.2), the lowest frequency in the spectrum, known as the fundamental frequency, is

$$\omega_1 = \frac{2\pi}{T}, \quad \text{or } f_1 = \frac{1}{T} \tag{3.12}$$

Often a time series displays a *trend,* i.e., a positive or negative slope in the data over the time record. Fort instance, the Canadian prairie yield shows a

positive trend with time, largely due to the gradual improvement in agricultural technology. The frequency associated with a trend is lower than the fundamental frequency, thus energy from the trend will leak to other low frequency spectral bands, thereby distorting the low frequency part of the spectrum. By subtracting the linear regression line from the data time series, trends can easily be removed prior to spectral analysis.
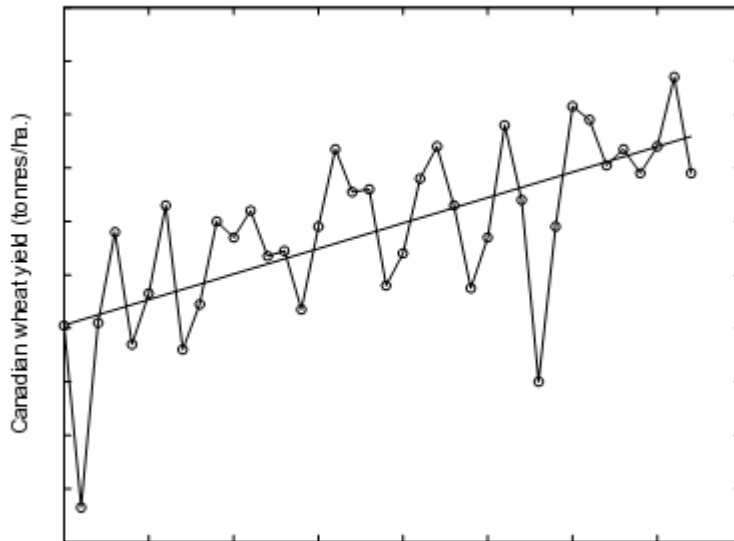

Fig.3.1

The highest resolvable frequency from (3.2) is $\omega = 2\pi M / T$, but with $M \approx N / 2$, we have $M / T \approx 1/(2\Delta t)$. Hence the highest resolvable frequency, called the Nyquist frequency, is

$$\omega_N = \frac{\pi}{\Delta T}, \text{ or } f_N = \frac{1}{2\Delta T} \qquad\qquad (3.13)$$

To resolve a wave of period of period τ , we need at least two data points to cover the period τ , i.e. $\tau = 2\Delta t = \dfrac{1}{f_N}$. Aliasing arises when Δt is too large to resolve the highest frequency oscillations in the data. Fig.3.2 illustrates a signal sampled too infrequently, resulting in an incorrect inference aliased signal of much low frequency.
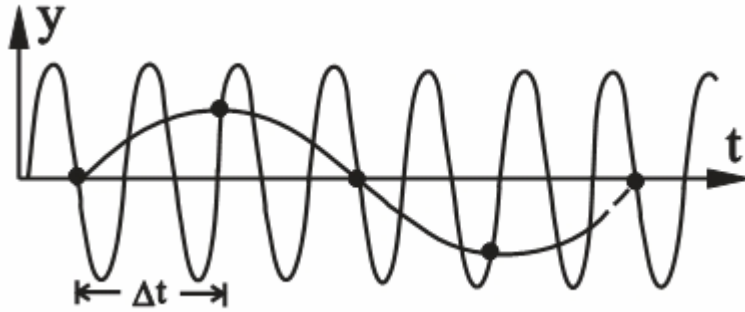
Fig3.2

From (3.2), the frequency $\Delta\omega$ between adjacent frequency bands is

$$\Delta\omega = \frac{2\pi}{T}$$

The ability to resolve neighboring spectral peaks is controlled by $\Delta\omega$, which is proportional to 1/T. Hence a longer record T will yield sharper spectral peaks, thereby allowing the resolution of two signals with close-by frequencies as distinct peaks in the spectrum. A shorter record will blur the two signals into a single spectral peak.

The raw spectrum S_m calculated from (3.9) is often very noisy in appearance. There are two common methods for smoothing the spectrum: (a) band-average (or running mean) is applied to the raw spectrum

$$\tilde{S}_m = \frac{1}{(2K+1)} \sum_{k=-K}^{K} S_{m+k} \qquad (3.14)$$

where $\tilde{S}_m$ is the smoothed spectrum resulting from averaging the raw spectrum over 2K+1 frequency bands.

In method (b), the data record is divided into J blocks of equal length L=T/J. Compute the spectrum for each block to get $S_m^{(j)}$ (j=1,…, J). The spectrum S_m is then the ensemble average over the J blocks:

$$S_m = \frac{1}{J} \sum_{j=1}^{J} S_m^{(j)} \qquad (3.15)$$

Method (b) has an advantage over (a) when there are data gaps --- in (b), the data gaps do not pose a serious problem since the data record is to be chopped into J blocks anyway, whereas in (a), the data gaps may have to be filled with interpolated values or zeros. The disadvantage of (b) is that the lowest resolvable frequency is $f_1 = 1/L = J/T$, hence there is a loss of low frequency information when the record is chopped up.

There is a trade off between the variance of the spectrum S and the band width. Increasing the band width (by increasing K or J) leads to a less noisy S, but spectrum peaks are broadened, so that nearby spectral peaks may merge together, resulting in a loss of resolution.

3.3 The continuous spectrum

The variance of the continuous time series y can be written by

$$\text{var}(y) = \int_{-\infty}^{\infty} [y(t)]^2 dt = \int_{-\infty}^{\infty} y(t) [\frac{1}{2\pi} \int_{-\infty}^{\infty} F(\omega) e^{i\omega t} d\omega] dt$$

$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\omega) [\int_{-\infty}^{\infty} y(t) e^{i\omega t} dt] d\omega = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\omega) F(-\omega) d\omega$$

(3.16)

Where (3.A) and (3.B) have been invoked.

$$F(-\omega) = F^*(\omega) \tag{3.17}$$

Asterisk denotes complex conjugation. So,

$$\text{var}(y) = \frac{1}{2\pi} \int_{-\infty}^{\infty} |F(\omega)|^2 d\omega \tag{3.18}$$

$|F(\omega)|^2$ is called the density of power spectrum, i.e.,

$$S(\omega) = F(\omega) F(-\omega) = |F(\omega)|^2, \text{ and } F(\omega) = \int_{-\infty}^{\infty} y(t) e^{-i\omega t} dt \tag{3.19}$$

The continuous spectrum could be calculated using (3.19), but it is usually

got by the Fourier transform of autocorrelation function. Assuming *y(t)* is a continuous time series with zero mean and unity variance, its autocorrelation function can be expressed

$$\rho(\tau) = \int_{-\infty}^{\infty} y(t)y(t+\tau)dt = \int_{-\infty}^{\infty} y(t)[\frac{1}{2\pi}\int_{-\infty}^{\infty} F(\omega)e^{i\omega(t+\tau)}d\omega]dt$$

$$= \int_{-\infty}^{\infty} y(t)[\frac{1}{2\pi}\int_{-\infty}^{\infty} F(\omega)e^{i\omega t}e^{i\omega\tau}d\omega]dt = \int_{-\infty}^{\infty} F(\omega)[\frac{1}{2\pi}\int_{-\infty}^{\infty} y(t)e^{i\omega t}dt]e^{i\omega\tau}d\omega$$

$$= \frac{1}{2\pi}\int_{-\infty}^{\infty} F(\omega)F(-\omega)e^{i\omega\tau}d\omega$$

From (3.17) and (3.19), we have

$$\rho(\tau) = \frac{1}{2\pi}\int_{-\infty}^{\infty} S(\omega)e^{i\omega\tau}d\tau \qquad (3.20)$$

(3.20) can also be written

$$S(\omega) = \int_{-\infty}^{\infty} \rho(\tau)e^{i\omega\tau}d\tau = 2\int_{0}^{\infty} \rho(\rho)\cos(\omega\tau)d\tau \qquad (3.21)$$

So, the spectrum is related to the autocorrelation function by a Fourier transform. The procedure to estimate continuous spectrum using (3.21) is as below

(1) calculating autocorrelation

$$\rho(\tau) = \frac{\sum_{i=1}^{N-\tau}(y_i - \overline{y})(y_{i+\tau} - \overline{y})}{\sum_{i=1}^{N}(y_i - \overline{y})^2} \qquad (3.22)$$

Where $\tau = 1,...,M$, and M is the maximum lag, usually set to ½ - 1/5 of N.

(2) using the trapezoidal rule to approximate (3.2)

$$\tilde{S}_k = 2\{\frac{1}{M}[\frac{1}{2}\rho(0) + \sum_{\tau=1}^{M}\rho(\tau)\cos(\omega_k\tau) + \frac{1}{2}\rho(M)\cos(\omega_k M)]\} \qquad (3.23)$$

where $\omega_k = \dfrac{2\pi k}{M}$

(3) smoothing raw spectrum to remove noise

$$S_0 = \frac{1}{2}\tilde{S}_0 + \frac{1}{2}\tilde{S}_k$$

$$S_l = \frac{1}{4}\tilde{S}_{k-1} + \frac{1}{2}\tilde{S}_k + \frac{1}{4}\tilde{S}_{k+1} \qquad (1 \le l \le M\text{-}1)$$

$$S_M = \frac{1}{2}\tilde{S}_{M-1} + \frac{1}{2}\tilde{S}_M$$

$$(3.24)$$

(4) plotting S_l as a function of wave number (k), and analyzing the period using (3.2).

(5) statistical test of spectrum

   (a) discrete spectrum test

$$F = \frac{\frac{1}{2}(a_k{}^2 + b_k^2)/2}{(var(y) - \frac{1}{2}a_k^2 - \frac{1}{2}b_k^2)/(n-2-1)} \sim F(2, n\text{-}2\text{-}1) \qquad (3.25)$$

   (b) continuous spectrum test

$$S_{ok} = \overline{S}[\frac{1 - \rho(1)^2}{1 + \rho(1)^2 - 2\rho(1)\cos(2\pi k / M)}] \qquad (3.26)$$

$$S_{ok}{}^* / S_{ok} = \chi_\upsilon^2 / \upsilon, \quad \upsilon = (2N - M/2)/M \qquad (3.27)$$

When S_k from (3.24) is greater than S^*_{ok}, the criteria of significance, the S_k is statistically significant; otherwise it is statistically insignificant.

3.4 Cross-spectrum

Now let us consider two time series, $x(t)$ and $y(t)$, their covariance:

$$\int_{-\infty}^{\infty} x(t)y(t)dt = \int_{-\infty}^{\infty} x(t)[\frac{1}{2\pi}\int_{-\infty}^{\infty} F_y(\omega)e^{i\omega t}d\omega]dt$$

$$= \frac{1}{2\pi}\int_{-\infty}^{\infty} F_y(\omega)d\omega \int_{-\infty}^{\infty} x(t)e^{i\omega t}dt = \frac{1}{2\pi}\int_{-\infty}^{\infty} F_y(\omega)F_x(-\omega)d\omega$$

(3.28)

Similar to (3.19), $S_{xy}(\omega) = F_y(\omega)F_x(-\omega)$ is defined the cross-spectrum. It describes the relationship between x and y in frequency domain. When $x=y$, $S_{xy}(\omega)$ is equal to the spectrum as shown in (3.19).

$S_{xy}(\omega)$ is a complex, and can be split into a real part and an imaginary part,

$$S_{xy}(\omega) = R_{xy}(\omega) - iI_{xy}(\omega)$$

(3.29)

where $R_{xy}(\omega)$ is the co-spectrum and $I_{xy}(\omega)$ is the quadrature spectrum. $S_{xy}(\omega)$ can also be expressed in polar form,

$$S_{xy}(\omega) = A_{xy}e^{i\theta_{xy}}$$

(3.30)

where A_{xy} is the amplitude spectrum and θ_{xy}, the phase spectrum, with

$$A_{xy} = \left|R_{xy}^2 + I_{xy}^2\right|^{\frac{1}{2}} \text{ and } \theta_{xy} = \tan^{-1}(I_{xy}/R_{xy}).$$

(3.31)

Cross-spectrum can also expressed in terms of auto cross-correlation $\rho_{xy}(\tau)$.

$$\rho_{xy}(\tau). = \int_{-\infty}^{\infty} x(t)y(t+\tau)dt \tag{3.32}$$

$$\rho_{yx}(\tau). = \int_{-\infty}^{\infty} y(t)x(t+\tau)dt \tag{3.33}$$

Since $\rho_{xy}(-\tau). = \int_{-\infty}^{\infty} x(t)y(t-\tau)dt = \int_{-\infty}^{\infty} x(t+\tau)y(t)dt = \rho_{yx}(\tau)$, we have

$$\rho_{xy}(\tau). = \rho_{yx}(-\tau)$$
$$\rho_{xy}(-\tau) = \rho_{yx}(\tau) \tag{3.34}$$

$$\rho_{xy}(\tau). = \int_{-\infty}^{\infty} x(t)y(t+\tau)dt = \int_{-\infty}^{\infty} x(t)[\frac{1}{2}\int_{-\infty}^{\infty} F_y(\omega)e^{i\omega(t+\tau)}d\omega]dt$$

$$= \int_{-\infty}^{\infty} F_y(\omega)[\frac{1}{2\pi}\int_{-\infty}^{\infty} x(t)e^{i\omega t}dt]e^{i\omega\tau}d\omega = \frac{1}{2\pi}\int_{-\infty}^{\infty} F_y(\omega)F_x(-\omega)e^{i\omega\tau}d\omega \tag{3.35}$$

$$= \frac{1}{2\pi}\int_{-\infty}^{\infty} S_{xy}(\omega)e^{i\omega\tau}d\omega$$

So,

$$S_{xy}(\omega) = \int_{-\infty}^{\infty} \rho_{xy}(\tau)e^{i\omega\tau}d\tau = \int_{-\infty}^{\infty} \rho_{xy}(\tau)\cos(\omega\tau)d\tau - i\int_{-\infty}^{\infty} \rho_{xy}(\tau)\sin(\omega\tau)d\tau$$

$$\tag{3.36}$$

$$R_{xy}(\omega) = \int_{-\infty}^{\infty} \rho_{xy}(\tau)\cos(\omega\tau)d\tau \tag{3.37}$$

$$I_{xy}(\omega) = \int_{-\infty}^{\infty} \rho_{xy}(\tau)\sin(\omega\tau)d\tau \tag{3.38}$$

Using (3.34), we can further write

$$R_{xy}(\omega) = \int_{-\infty}^{0} \rho_{xy}(\tau)\cos(\omega\tau)d\tau + \int_{0}^{\infty} \rho_{xy}(\tau)\cos(\omega\tau)d\tau$$

$$= -\int_{0}^{\infty} \rho_{xy}(-\tau)\cos\omega(-\tau)d(-\tau) + \int_{0}^{\infty} \rho_{xy}(\tau)\cos(\omega\tau)d\tau \qquad (3.39)$$

$$= \int_{0}^{\infty} [\rho_{yx}(\tau) + \rho_{xy}(\tau)]\cos(\omega\tau)d\tau$$

Similarly, $\quad I_{xy}(\omega) = \int_{0}^{\infty} [\rho_{xy}(\tau) - \rho_{yx}(\tau)]\sin(\omega\tau)d\tau \qquad (3.40)$

We can use (3.39) and (3.40) to calculate the cross-spectrum, the detailed procedure is as below

(1) calculating cross-correlation $\rho_{xy}(\tau)$ and $\rho_{yx}(\tau)$ using the discrete form of (3.32) and (3.33),

$$\rho_{xy}(\tau) = \frac{\sum_{i=1}^{N-\tau}(x_i - \overline{x})(y_{i+\tau} - \overline{y})}{[\sum_{i=1}^{N}(x_i - \overline{x})^2]^{\frac{1}{2}}[\sum_{i=1}^{N}(y_i - \overline{y})^2]^{\frac{1}{2}}} \qquad (3.41)$$

$$\rho_{yx}(\tau) = \frac{\sum_{i=1}^{N-\tau}(y_i - \overline{y})(x_{i+\tau} - \overline{x})}{[\sum_{i=1}^{N}(x_i - \overline{x})^2]^{\frac{1}{2}}[\sum_{i=1}^{N}(y_i - \overline{y})^2]^{\frac{1}{2}}} \qquad (3.42)$$

Where $\tau = 1,...M$, and M is the maximum lag, usually set to 1/3-1/5 of the original sample length N.

(2) using the trapezoidal rule to approximate (3.39) and (3.40), we have

$$R_{xy}(\omega_k) = \frac{1}{M}[\rho(0) + \sum_{\tau=1}^{M}(\rho_{xy}(\tau) + \rho_{yx}(\tau))\cos(\omega_k\tau) + \rho_{xy}(M)\cos(\omega_k M)]\}$$

$$(3.43)$$

$$I_{xy}(\omega_k) = \frac{1}{M}\sum_{\tau=1}^{M}[\rho(\tau)_{xy} - \rho_{yx}(\tau)]\sin(\omega_k\tau) \qquad (3.44)$$

Where $\omega_k = \dfrac{2\pi k}{M}$.

(3) smoothing raw spectrum to remove noise

$$R_{xy}(k) = \frac{B_k}{M}[\rho_{xy}(0) + \frac{1}{2}\sum_{\tau=1}^{M}(\rho_{xy}(\tau) + \rho_{yx}(\tau))(1 + \cos(\frac{\pi\tau}{M})\cos(\frac{2k\pi\tau}{M})] \qquad (3.45)$$

$$I_{xy}(k) = \frac{B_k}{M}\sum_{\tau=1}^{M}[\rho(\tau)_{xy} - \rho_{yx}(\tau)][\frac{1}{2}(1 + \cos(\frac{\pi\tau}{M})\sin(\frac{2k\pi\tau}{M})] \qquad (3.46)$$

Where
$B_k = 1 \qquad (k \neq 0, M/2)$
$B_k = 1/2 \qquad$ (k=0, M/2)

A useful quantity is the squared coherency spectrum (where the word "squared" is often omitted for brevity):

$$P_{xy}(\omega) = \frac{R_{xy}^{2}(\omega) + I_{xy}^{2}(\omega)}{R_{xx}(\omega)R_{yy}(\omega)} \qquad (3.47)$$

Physically, it indicates the representation of the relation (coherency) between x and y at the frequency of ω.

In practice, (3.47) is used to examine the relation of two time series at different frequencies. A diagram is often plotted to show the relation of P and ω. When a highly coherency frequency is found (corresponding the maximum of P), the lag phase between the two time series responsible for the most coherency frequency can be obtained by θ_{xy} in (3.31), or by

$$L(\omega) = \frac{\theta_{xy}(\omega)T}{2\pi} \qquad (3.48)$$

Where T is the period of ω.

$$F = \frac{(\upsilon-1)P_{xy}^2}{1-P_{xy}^2} \quad \sim \quad F(2, 2(\upsilon-1)) \tag{3.49}$$

(3.49) can be used for significant test, where υ is determined by (3.27).

3.5 Filters

One often would like to perform digital filtering on the raw data. For instance, one may want a smoother data field, or want to concentrate on the low-frequency or high-frequency signals in the time series. Let us express a time series $x(t)$ in terms of its complex Fourier components $X(\omega)$:

$$x(t) = \sum_{\omega} X(\omega)e^{i\omega t} \tag{3.50}$$

Where it is understood that ω and t denote the discrete variables ω_m and t_n. A filtered time series is given by

$$\tilde{x}(t) = \sum_{\omega} f(\omega)X(\omega)e^{i\omega t} \tag{3.51}$$

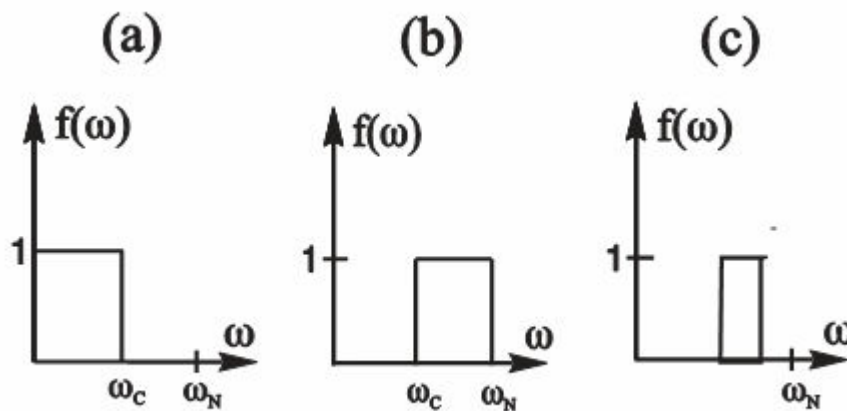where $f(\omega)$ is the filter "response" function.

Fig.3.3

Fig3.3 illustrates several ideal filters for low-pass, high-pass and band-pass filter. In these ideal filters, the step discontinuity as the cut-off frequency

ω_c produces "ringing" (i.e. oscillations) in the filtered time series (especially at the two ends). This problem of a step discontinuity in the frequency domain leading to ringing in the time domain is like a time series truncated by a rectangular window leading to energy leakage in the frequency domain. Thus in practice, $f(\omega)$ needs to be taped at ω_c to suppress ringing in the filtered time series.

To perform filtering in the frequency domain, the steps are: (1) Fourier transform $x(t)$ to $X(\omega)$; (2) multiply $X(\omega)$ by $f(\omega)$; and (3) inverse transform $f(\omega) X(\omega)$ to get $\tilde{x}(t)$, the filtered time series.

Alternatively, filtering can be performed in the time domain as well. In fact, prior to the invention of last Fourier transform algorithms, filtering in the frequency domain was prohibitively expensive, although nowadays, the tasks are trivial in computing costs, and filtering can be performed in either the frequency or the time domain.

A commonly used time domain filter is the 3-point moving average (or running mean) filter

$$\tilde{x}_n = \frac{1}{3} x_{n-1} + \frac{1}{3} x_n + \frac{1}{3} x_{n+1} \qquad (3.52)$$

i.e., average over the immediate neighbors. More generally, a filtered time series is given by

$$\tilde{x}_n = \sum_{l=-L}^{L} w_l x_{n+l} \qquad (3.53)$$

where w_l are the weights of the filter.

Suppose the filtered time series has the Fourier decomposition

$$\tilde{x}_n = \sum_{l=-L}^{L} \tilde{X}(\omega) e^{i\omega t_n} \qquad (3.54)$$

Comparing with (3.51), one sees that

$$\tilde{X}(\omega) = f(\omega)X(\omega) \tag{3.55}$$

So,

$$f(\omega) = \frac{\tilde{X}(\omega)}{X(\omega)} \tag{3.56}$$

From (3.54) and (3.53)

$$\tilde{X}(\omega) = \sum_{l=-L} \tilde{x}_n e^{-i\omega t_n} = \sum_n \sum_{l=-L}^{L} w_l x_{n+l} e^{-i\omega t_n} = \sum_{l=-L}^{L} w_l e^{i\omega t_l} \sum_n x_{n+l} e^{-i\omega(t_n+t_l)}$$

$$= \sum_{l=-L}^{L} w_l e^{i\omega t_l} X(\omega)$$

So, (3.56) can be expressed

$$f(\omega) = \sum_{l=-L}^{L} w_l e^{i\omega t_l} \tag{3.57}$$

which allows us to calculate the filter response function $f(\omega)$ from the given weights of a time domain filter.

For example, moving average filter have the general form

$$\tilde{x}_n = \sum_{l=-L}^{L} (\frac{1}{2L+1}) x_{n+1} \tag{3.58}$$

Another commonly used filter is the 3-point triangular filter,

$$\tilde{x}_n = \frac{1}{4} x_{n-1} + \frac{1}{2} x_n + \frac{1}{4} x_{n+1} \tag{3.59}$$

One often encounters time series containing strong periodic signals, e.g., the seasonal cycle. While there periodic signals are important, it is often the no-

periodic signals which have the most important impact on humans, as they produce the unexpected events. One often world remove the strong periodic signals from the time series first.

Suppose one has monthly data for a variable *x*, and one would like to extract the seasonal cycle. Average all x values in January to get $\bar{x}_{jan}$, and similarly for the other months. The climatology seasonal cycle is then given by

$$\bar{x}_{seasonal} = [\bar{x}_{jan}, ..., \bar{x}_{dec}]$$

The filtered time series is obtained by subtracting the climatological seasonal cycle from a raw data ---i.e., all January values of x will have $\bar{x}_{jan}$ subtracted, and similarly for the other months.