NOTES OF TIME SERIES ANALYSIS

Olivier Marchal
Department of Geology & Geophysics
Woods Hole Oceanographic Institution
September 21, 2018

These notes collect some of the most basic facts of time series analysis, with emphasis on the analysis in the frequency domain (spectral analysis). Only 'convential' methods of spectral analysis are covered, i.e., procedures such as the multitaper technique and the maximum entropy method are not considered. For a complete treatment and real understanding of the subject, one or several of the textbooks listed below should be consulted. The present notes are largely based on the book of C. Chatfield, which has been found to be particularly useful as an introductory discussion.

TEXTBOOKS

• Box G., Jenkins G., and Reinsel G., *Time series analysis - Forecasting and control*, 3rd ed., Prentice-Hall, 1994

• Chatfield C., *The analysis of time series - An introduction*, Chapman-Hall, 1996.

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

• Percival D. and Walden A., *Spectral analysis for physical applications (multitaper and conventional univariate techniques)*, Cambridge University Press, 1993.

• Priestley M., *Spectral analysis and time series*, 2 volumes, Academic Press, 1981

ADDITIONAL READING

• Ghil M. and Yiou P., Spectral methods: What they can and cannot do for climatic time series, In *Decadal climate variability: Dynamics and Predictability*, Anderson D. and Willebrand J. (eds.), NATO ASI Series, Vol. I44, 1996

• Schulz M. and Stattegger K., SPECTRUM: Spectral analysis of unevenly spaced paleoclimatic time series, *Computers & Geosciences*, vol. 23, 929-945, 1997

• Schulz M. and Mudelsee M., REDFIT: Estimating red-noise spectra directly from unevenly spaced paleoclimate time series, *Computers & Geosciences*, vol. 28, 421-426, 2002

• Yiou P., Baert E., and Loutre M. F., *Spectral analysis of climate data*, Surveys of Geophysics, vol. 17, 619-663, 1996.

1 Introduction

• A **time series** is a collection of observations made sequentially in time.

• A **stochastic** time series is one whose future values cannot be predicted exactly. If such values can be predicted exactly, the time series is **deterministic**.

• The analysis of time series is based on two (complementary) approaches:

i. Analysis in **time domain** (major diagnostic tool = **autocorrelation function**)

ii. Analysis in **frequency domain** (major diagnostic tool = **spectral density function**)

2 Simple Descriptive Techniques

2.1 Types of Variation

- Trend
- Cyclic changes
- Irregular fluctuations

2.2 Stationary Time Series

Most of the theory of time series analysis is concerned with stationary time series. Broadly a time series is stationary if it contains no systematic change in mean level (no trend), no systematic change in variance, and no variations that are strictly periodic. A more precise definition of stationarity will be given below.

2.3 Analysing Series Containing a Trend

- Consider a discrete time series with evenly spaced observations $x_1, x_2, \ldots, x_N$.
- The simplest trend is the one described by the linear regression model:

$$x_t = \alpha + \beta t + \epsilon_t, \tag{1}$$

where (α, β) are constants, ϵ_t is a random error with zero mean, and $t = 1, 2, \ldots, N$.

- Filtering:

$$y_t = \sum_{r=-q}^{s} a_r x_{t+r}. \tag{2}$$

If $\sum a_r = 1$ the filter is commonly called a **moving average**. Moving averages are often symmetric with $s = q$ and $a_r = a_{-r}$. For example,

$$\mathrm{Sm}(x_t) = \frac{1}{2q+1} \sum_{r=-q}^{q} x_{t+r}. \tag{3}$$

The departures of the observations from the trend are often called the residuals. For example,

$$\begin{aligned} \mathrm{Res}(x_t) &= x_t - \mathrm{Sm}(x_t) \\ &= \sum_{r=-q}^{q} b_r x_{t+r}. \end{aligned} \tag{4}$$

Thus, $\mathrm{Res}(x_t)$ is also a filter. $\mathrm{Sm}(x_t)$ is a **low-pass filter**, whereas $\mathrm{Res}(x_t)$ is a **high-pass filter**. A more precise definition of both filters will be given later.

• Differencing:

$$\nabla x_t = x_t - x_{t-1}. \tag{5}$$

Differencing is particulary useful for removing a trend.

2.4 Analysing Series Containing a Cyclic Variation

For example, the seasonal component in a time series can be removed from:

$$\mathrm{Sm}(x_t) = \frac{\frac{1}{2}x_{t-6} + x_{t-5} + x_{t-4} + \ldots + x_{t+4} + x_{t+5} + \frac{1}{2}x_{t+6}}{12}. \tag{6}$$

2.5 Autocovariance and Autocorrelation

• Let $\bar{x}$ be the arithmetic average of a time series $x_1, \ldots, x_N$:

$$\bar{x} = \frac{1}{N}\sum_{t=1}^{N} x_t. \tag{7}$$

• The **autocovariance coefficient** at lag k of the time series is usually computed from

$$c_k = \frac{1}{N}\sum_{t=1}^{N-k} (x_t - \bar{x})(x_{t+k} - \bar{x}). \tag{8}$$

• The **autocorrelation coefficient** at lag k of the time series is computed from

$$r_k = \frac{c_k}{c_0}, \tag{9}$$

where c_0 is the variance of the observations.

• The **correlogram** is a plot of r_k versus k. It is probably pointless to compute r_k for $k > N/4$. The correlogram can help to characterize time series, e.g., a purely random series (to be defined later), short-term correlation, alternating series, non-stationary series, series showing cyclic changes, as well as potential outliers. In particular, the autocorrelation coefficients of a purely random series have a normal (Gaussian) distribution with a mean close to zero and a variance of 1/N, where N is the number of data in the time series, which we write

$$r_k \sim \mathcal{N}\left(0, \frac{1}{N}\right). \tag{10}$$

Thus, 19 out of 20 of the values of r_k from an observed time series are expected to lie between $\pm 2/\sqrt{N}$, if this series is purely random. The correlogram provides a means for testing the **randomness** in a time series.

3 Probability Models for Time Series

3.1 Stochastic Processes

• In the theory of time series, each observation of a given time series is regarded as a random variable. The set of random variables that characterizes the whole series is called a random or **stochastic process**.

• Probability models for time series are thus models describing the temporal evolution of random variables, i.e., stochastic processes.

• If the values of the time series occur only at discrete times, the random variables are written as X_t, where $t = 0, \pm 1, \pm 2, \ldots$, whereas if they occur continuously they are referred to as $X(t)$, where $-\infty < t < \infty$. Thus, a discrete process is written as the ensemble $\{X_t\}$, whereas a continuous process is written as the ensemble $\{X(t)\}$.

• A given observed time series is regarded as a particular **realization** of a stochastic process. An important part of time series analysis is concerned with inferring the structure of the process that generated the observed time series.

• Because the values originating from a time series are viewed as random variables, each of these variables possesses a probability distribution and these variables jointly are characterized by a joint probability distribution.

• Consider a continuous stochastic process $\{X(t)\}$. The process is **strictly stationary** if its joint probability density function is invariant under a shift of the time origin, i.e., if

$$p\left[X(t_1) = x(t_1), X(t_2) = x(t_2), \ldots, X(t_n) = x(t_n)\right] \tag{11}$$

is equal to

$$p\left[X(t_1+\tau) = x(t_1+\tau), X(t_2+\tau) = x(t_2+\tau), \ldots, X(t_n+\tau) = x(t_n+\tau)\right]. \tag{12}$$

• A process is **second-order stationary** it it has a constant mean, a constant variance, and an autocovariance function that depends only on the lag k. For a discrete process, these three conditions are, respectively,

$$\mathrm{E}[X_t] = \mu, \tag{13a}$$
$$\mathrm{Var}[X_t] = \sigma^2, \tag{13b}$$
$$\mathrm{Cov}[X_t, X_{t+k}] = \gamma(k). \tag{13c}$$

3.2 Autocorrelation Function

• The **sample autocorrelation function** of a discrete time series, r_k, has already been described in section 2.5

- The **theoretical autocorrelation function** of a discrete stochastic process is

$$\rho(k) = \frac{\gamma(k)}{\gamma(0)}, \tag{14}$$

where $\gamma(0) = \mathrm{Var}[X_t]$.

- The function $\rho(k)$ has three important properties:

  i. $\rho(k) = \rho(-k)$ ($\rho(k)$ is an even function),

  ii. $-1 \leq \rho(k) \leq 1$,

  iii. Lack of uniqueness: the same function $\rho(k)$ can characterize different stochastic processes.

3.3 Examples of Stochastic Processes

3.3.1 Purely Random Process (white noise)

- The random variables composing a purely random process are mutually independent and identically distributed.

- For a discrete process:

$$\begin{aligned} \mathrm{E}[Z_t] &= \mu, & (15a)\\ \mathrm{Var}[Z_t] &= \sigma^2, & (15b)\\ \mathrm{Cov}[Z_t, Z_{t+k}] &= 0 \ \forall \ k \neq 0. & (15c) \end{aligned}$$

3.3.2 Random Walk

- Let $\{Z_t\}$ be a discrete purely random process with mean μ and variance σ_Z^2. The process $\{X_t\}$ is a random walk if

$$X_t = X_{t-1} + Z_t. \tag{16}$$

- The mean and variance of X_t are (for $X_0 = 0$)

$$\begin{aligned} \mathrm{E}[X_t] &= \mu\, t, & (17a)\\ \mathrm{Var}[X_t] &= \sigma_Z^2 t. & (17b) \end{aligned}$$

Thus, the random walk is not a stationary process.

3.3.3 Moving Average Process

• Let $\{Z_t\}$ be a discrete purely random process with mean zero and variance σ_Z^2. The process $\{X_t\}$ is a moving average process of order q, noted MA(q), if

$$X_t = \beta_0 Z_t + \beta_1 Z_{t-1} + \ldots + \beta_q Z_{t-q}, \tag{18}$$

where the coefficients β_i are constants.

• The mean and variance of X_t are

$$\begin{aligned} \mathrm{E}[X_t] &= 0, && (19a) \\ \mathrm{Var}[X_t] &= \sigma_Z^2 \sum_{i=0}^{q} \beta_i^2. && (19b) \end{aligned}$$

The function $\rho(k)$ has vanishing values for $k > q$.

3.3.4 Autoregressive Process

• Let $\{Z_t\}$ be a discrete purely random process with mean zero and variance σ_Z^2. The process $\{X_t\}$ is an autoregressive process of order p, noted AR(p), if

$$X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \ldots + \alpha_p X_{t-p} + Z_t, \tag{20}$$

where the coefficients α_i are constants.

• Example: Autoregressive process of order one or Markov process, noted AR(1):

$$X_t = \alpha X_{t-1} + Z_t. \tag{21}$$

The mean, variance, and autocorrelation of a AR(1) process are (for $|\alpha| < 1$):

$$\begin{aligned} \mathrm{E}[X_t] &= 0, && (22a) \\ \mathrm{Var}[X_t] &= \frac{\sigma_Z^2}{1-\alpha^2}, && (22b) \\ \rho(k) &= \alpha^k \quad \text{for } k = 0, 1, 2, \ldots. && (22c) \end{aligned}$$

• AR processes can be applied in situations where one expects that the present value of the time series depends on the past values together with some random effect.

4 Estimation in Time Domain

• The problem here is to fit a suitable probability model to an observed time series. The major tool for doing this is the **sample autocorrelation function** (noted r_k for the discrete case). Inference based on this function is often called an **analysis in the time domain**.

• This problem is not exposed here. An introduction is provided in Chatfield (1996). Box et al. (1994) discuss this problem at length.

5 Stationary Processes in Frequency Domain

• The problem here is to determine the frequency properties of an observed time series. The major diagnostic tool for doing this is the **sample spectral density function**. Inference based on this function is often called an **analysis in the frequency domain**.

5.1 Spectral Distribution Function

• Let us consider a stochastic process that is discrete, real-valued, and stationary, with an autocovariance fonction $\gamma(k)$. A theoretical result (Wiener-Khintchine theorem) shows that there then exists a function $F(\cdot)$ that increases monotically with (angular) frequency ω such that

$$\gamma(k) = \int_0^\pi \cos(\omega k)\ \mathrm{d}F(\omega). \tag{23}$$

• The function $F(\omega)$ is the **spectral distribution function**. It has an important physical interpretation: $F(\omega)$ is the contribution to the variance of the process which is accounted for by frequencies in the range $(0, \omega)$.

• There is no variation at negative frequencies, so that

$$F(\omega) = 0 \text{ for } \omega < 0. \tag{24}$$

• For a discrete process measured at unit intervals, the highest possible frequency is π, and so all the variation is accounted for by frequencies less than π, i.e.,

$$F(\pi) = \mathrm{Var}(X_t) = \sigma_X^2. \tag{25}$$

• The **normalized spectral distribution function** is

$$F_*(\omega) = \frac{F(\omega)}{\sigma_X^2}. \tag{26}$$

$F_*(\omega)$ has similar properties to a cumulative distribution function.

5.2 Spectral Density Function

• The **spectral density function** (also called the 'power spectrum' or more simply the 'spectrum') is the first derivative of the spectral distribution function:

$$f(\omega) = \frac{\mathrm{d}F(\omega)}{\mathrm{d}\omega}. \tag{27}$$

• The spectrum is thus related to the autocovariance function of the process by

$$\gamma(k) = \int_0^{\pi} \cos(\omega k)\; f(\omega)\mathrm{d}\omega. \tag{28}$$

• The spectrum has also an important physical meaning. To show this, consider the autocovariance function at zero lag, i.e., the variance of the process,

$$\gamma(0) = \sigma_X^2 = \int_0^{\pi} f(\omega)\mathrm{d}\omega. \tag{29}$$

Thus, $f(\omega)\mathrm{d}\omega$ represents the contribution to variance of components with frequencies in the range $(\omega, \omega + d\omega)$. A peak in the spectrum $f(\omega)$ indicates an important contribution at frequencies in the appropriate interval.

• The spectrum is the **Fourier transform** of the autovariance function:

$$f(\omega) = \frac{1}{\pi} \sum_{k=-\infty}^{\infty} \gamma(k) e^{-i\omega k}. \tag{30}$$

Thus, $\gamma(k)$ and $f(\omega)$ are a Fourier transform pair. Since $\gamma(k)$ is an even function,

$$f(\omega) = \frac{1}{\pi} \left[\gamma(0) + 2 \sum_{k=1}^{\infty} \gamma(k) \cos \omega k \right]. \tag{31}$$

• The **normalized spectral density function** is

$$f_*(\omega) = \frac{f(\omega)}{\sigma_X^2}. \tag{32}$$

Thus, $f_*(\omega)\mathrm{d}\omega$ is the proportion of variance in the interval $(\omega, \omega + d\omega)$. $f_*(\omega)$ has similar properties to a probability density function. It is the Fourier transform of the autocorrelation function. The autocorrelation function $\rho(k)$ and the normalized spectral density function $f_*(\omega)$ are another Fourier transform pair.

5.3 Spectra for Selected Stochastic Processes

• For a purely random process Z_t,

$$f_*(\omega) = \frac{1}{\pi}. \tag{33}$$

The spectrum is 'flat', hence the name 'white noise' (the spectrum has 'no color').

• For a first-order moving average process ($X_t = Z_t + \beta Z_{t-1}$),

$$f_*(\omega) = \frac{1}{\pi} \left(1 + \frac{2\beta \cos \omega}{1 + \beta^2} \right). \tag{34}$$

• For a first-order autoregressive process ($X_t = \alpha X_{t-1} + Z_t$),

$$f_*(\omega) = \frac{1}{\pi} \frac{1 - \alpha^2}{1 - 2\alpha \cos \omega + \alpha^2}. \tag{35}$$

6 Spectral Analysis

• Spectral analysis is concerned with estimating the spectral density function of a given time series.

• It is essentially a modification of Fourier analysis so as to make it suitable for stochastic rather than deterministic functions of time.

6.1 Periodogram Analysis

• Consider the possibility of fitting a Fourier series to a time series of N observations $x_1, x_2, \dots, x_N$ with unit sampling interval ($\Delta t = 1$). It is assumed that N is even. The component with lowest frequency completes only one cycle over the whole length of the series, which is approximated as $N\Delta t = N$; its frequency ω is thus $2\pi/N$. The component in the series with highest frequency completes one cycle in two sampling intervals, i.e., over a time corresponding to $2\Delta t = 2$; its frequency ω is thus π. This latter component is the **Nyquist frequency**. The finite Fourier series representation of the observed time series is therefore

$$x_t = a_0 + \sum_{p=1}^{N/2-1} [a_p \cos \omega_p t + b_p \sin \omega_p t] + a_{N/2} \cos \pi t, \tag{36}$$

where $\omega_p = 2\pi p/N$.

• The Fourier coefficients are given by

$$\begin{aligned} a_0 &= \frac{1}{N} \sum x_t, && \text{(37a)} \\ a_p &= \frac{2}{N} \sum x_t \cos \omega_p t \qquad \text{for } p = 1, 2, \dots, N/2 - 1, && \text{(37b)} \\ b_p &= \frac{2}{N} \sum x_t \sin \omega_p t \qquad \text{for } p = 1, 2, \dots, N/2 - 1, && \text{(37c)} \\ a_{N/2} &= \frac{1}{N} \sum (-1)^t x_t, && \text{(37d)} \end{aligned}$$

where all summations are from $t = 1$ to $t = N$.

• The Fourier series has N parameters to describe N observations and so can be made to fit the data exactly.

• It expresses the partition of the variability of the series into components at frequencies $2\pi/N, 4\pi/N, \dots, \pi$.

• The pth Fourier component ('harmonic') can be written as

$$a_p \cos \omega_p t + b_p \sin \omega_p t = R_p \cos(\omega_p t + \phi_p), \tag{38}$$

where $R_p = \sqrt{a_p^2 + b_p^2}$ is the amplitude of the pth harmonic and $\phi_p = \tan^{-1}(-b_p/a_p)$ is the phase of the pth harmonic.

• Parseval's theorem:

$$\frac{1}{N}\sum_{t=1}^{N}(x_t - \bar{x}) = \sum_{p=1}^{N/2-1}\frac{R_p^2}{2} + a_{N/2}^2. \tag{39}$$

Thus, $R_p^2/2$ is the contribution of the pth harmonic to the variance, and the Fourier series shows how the variance is partitioned between different frequencies.

• A plot of $R_p^2/2$ versus $\omega_p = 2\pi p/N$ is a **line spectrum**.

• A plot of $I(\omega_p) = NR_p^2/(4\pi)$ versus $\omega_p = 2\pi p/N$ is a **periodogram**.

• The periodogram ordinate at the Nyquist frequency is computed by regarding $a_{N/2}^2$ as the contribution to variance in the range $[\pi(N-1)/N, \pi]$, i.e.,

$$I(\pi) = \frac{Na_{N/2}^2}{\pi}. \tag{40}$$

• The periodogram $I(\omega_p)$ is the finite Fourier transform of the sample autocovariance function c_k:

$$I(\omega_p) = \frac{1}{\pi}\sum_{k=-(N-1)}^{N-1} c_k e^{-i\omega_p k} \tag{41a}$$

$$= \frac{1}{\pi}\left(c_o + 2\sum_{k=1}^{N-1} c_k \cos\omega_p k\right) \tag{41b}$$

• Sampling properties of $I(\omega_p)$:

i. The estimator $I(\omega_p)$ is asymptotically unbiased:

$$E[I(\omega_p)] \to f(\omega) \qquad \text{as } N \to \infty \tag{42}$$

ii. The estimator $I(\omega_p)$ is not consistent:

$$\mathrm{Var}[I(\omega_p)] \text{ does not vanish as } N \to \infty \tag{43}$$

6.2 Some Consistent Estimation Procedures

6.2.1 Truncating & Weighting the Autocovariance Function

• The estimator of the true spectrum is

$$\hat{f}(\omega) = \frac{1}{\pi}\left[\lambda_0 c_0 + 2\sum_{k=1}^{M}\lambda_k c_k \cos\omega k\right], \tag{44}$$

where $\{\lambda_k\}$ is a set of weights called the **lag window**, and $M < N$ is the **truncation point**.

• Example 1: The Tukey, Tukey-Hanning, or Blackman-Tukey window is

$$\lambda_k = \frac{1}{2}\left(1 + \cos\frac{\pi k}{M}\right) \qquad \text{for } k = 0, 1, \ldots, M. \tag{45}$$

• Example 2: The Parzen window is

$$\begin{aligned} \lambda_k &= 1 - 6\left(\frac{k}{M}\right)^2 + 6\left(\frac{k}{M}\right)^3 \qquad & 0 \le k \le M/2, \qquad &(46a) \\ \lambda_k &= 2\,(1 - k/M)^3 & M/2 \le k \le M. \qquad &(46b) \end{aligned}$$

• The bias of the spectral estimates increases with decreasing M and the variance of the spectral estimates decreases with decreasing M.

6.2.2 Hanning

• The autocovariance function is first truncated:

$$\hat{f}_1(\omega) = \frac{1}{\pi}\left[c_0 + 2\sum_{k=1}^{M} c_k \cos\omega k\right]. \tag{47}$$

Estimates of the spectrum are then obtained by smoothing with the weights $(\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$:

$$\hat{f}(\omega) = \frac{1}{4}\hat{f}_1(\omega - \frac{\pi}{M}) + \frac{1}{2}\hat{f}_1(\omega) + \frac{1}{4}\hat{f}_1(\omega + \frac{\pi}{M}). \tag{48}$$

• At zero frequency and at the Nyquist frequency,

$$\begin{aligned} \hat{f}(0) &= \frac{1}{2}\left[\hat{f}_1(0) + \hat{f}_1(\frac{\pi}{M})\right], \qquad &(49a) \\ \hat{f}(\pi) &= \frac{1}{2}\left[\hat{f}_1(\pi) + \hat{f}_1(\frac{\pi(M-1)}{M})\right]. \qquad &(49b) \end{aligned}$$

• Hanning is equivalent to using the Tukey window.

6.2.3 Hamming

• Hamming is similar to Hanning except that the weights for $\hat{f}(\omega)$ are changed from $(\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$ to $(0.23, 0.54, 0.23)$. At frequencies $\omega = 0$ and $\omega = \pi$, the weights are 0.54 and 0.46, respectively.

6.2.4 Smoothing the Periodogram

• The spectral estimates are given by

$$\hat{f}(\omega) = \frac{1}{m}\sum_{j} I(\omega_j), \tag{50}$$

where $\omega_j = 2\pi j/N$ and j varies over consecutive integers so that the ω_j are symmetric about ω.

• The spectral estimates at zero frequency and at the Nyquist frequency assume that the periodogram can be treated as being symmetric about 0 and π. Thus, taking m to be odd with $m_* = (m-1)/2$,

$$\hat{f}(0) = I(0) + \frac{2}{m}\sum_{j=1}^{m_*} I(2\pi j/N). \tag{51}$$

Likewise,

$$\hat{f}(\pi) = \frac{1}{m}\left[I(\pi) + 2\sum_{j=1}^{m_*} I(\pi - 2\pi j/N)\right]. \tag{52}$$

• The bias of the spectral estimates increases with increasing m, and the variance of the spectral estimates decreases with increasing m.

6.3 Confidence Interval for the Spectrum

• The construction of the confidence intervals for the spectral estimates assumes that the observations are normally distributed with constant mean and constant variance. For a discrete stochastic process,

$$X_t \sim \mathcal{N}\left(\mu_X, \sigma_X^2\right). \tag{53}$$

The assumption of normality may or may not apply to the observations being investigated.

• Consider the estimator

$$\hat{f}(\omega) = \frac{1}{\pi}\sum_{k=-M}^{M} \lambda_k c_k \cos \omega k. \tag{54}$$

The number of degrees of freedom for this estimator is

$$\nu = \frac{2N}{\sum\limits_{k=-M}^{M} \lambda_k^2}. \tag{55}$$

• A typical $100(1-\alpha)\%$ confidence interval for $f(\omega)$ is given by:

$$\frac{\nu \hat{f}(\omega)}{\chi^2_{\nu,\alpha/2}} \text{ to } \frac{\nu \hat{f}(\omega)}{\chi^2_{\nu,1-\alpha/2}}, \tag{56}$$

where χ^2_ν is the variate of the χ^2 distribution with ν degrees of freedom and α is the specified level of confidence.

• The degrees of freedom for the Tukey and Parzen windows are $2.67N/M$ and $3.71N/M$, respectively. The degree of freedom for the smoothed periodogram is $2m$.

6.4 Comparison between Different Estimation Procedures

• The comparison is commonly based on the **spectral window** or **kernel** $K(\omega)$, which is the Fourier transform of the lag window.

• If the lag window λ_k is zero for $k > M$ and $\lambda_k = \lambda_{-k}$, it can be shown that

$$\hat{f}(\omega_0) = \int_{-\pi}^{\pi} K(\omega) I(\omega_0 - \omega) \mathrm{d}\omega. \tag{57}$$

Thus, all estimation procedures are essentially smoothing the periodogram using the weight function $K(\omega)$.

• The expected value of the spectral estimate $\hat{f}(\omega_0)$ is, for $N \to \infty$,

$$\mathrm{E}\left[\hat{f}(\omega_0)\right] = \int_{-\pi}^{\pi} K(\omega) f(\omega_0 - \omega) \mathrm{d}\omega. \tag{58}$$

The bias depends on the spectral window.

• The **bandwidth** is, roughly speaking, the width of the spectral window. The bandwidths of the Tukey and Parzen windows are, respectively, $8\pi/(3M)$ and $2\pi(1.86/M)$. The bandwidth of the smoothed periodogram is $2\pi m/N$.

• The choice of bandwidth appears more critical than the choice of the window.

• The choice of bandwidth is based on a trade-off between frequency resolution and variance. For the Tukey and Parzen windows, the lower the value of the truncation point the lower the variance but the larger the bias. For the smoothed periodogram, the lower the value of m the lower the bias but the higher the variance. The choice of bandwidth is rather like the choice of class interval when constructing a histogram.

6.5 Aliasing, Tapering, and Prewithening

6.5.1 Aliasing

Consider a continuous time series with spectrum $f_c(\omega)$ for $0 < \omega < \pi$, which is sampled at equal intervals of length Δt. The resulting discrete time series has a spectrum $f_d(\omega)$

defined over $0 < \omega < \pi/\Delta t$. It can be shown that

$$f_d(\omega) = \sum_{s=0}^{\infty} f_c\left(\omega + 2\pi s/\Delta t\right) + \sum_{s=1}^{\infty} f_c\left(-\omega + 2\pi s/\Delta t\right). \tag{59}$$

Thus, the effect of sampling will be that variation at frequencies above the Nyquist frequency will be 'folded back' and produce an effect at a frequency lower than the Nyquist frequency in $f_d(\omega)$. Such an effect is called **aliasing**. Thus, the time interval Δt should be chosen such that $f_c(\omega) \approx 0$ for $\omega > \pi/\Delta t$.

6.5.2 Tapering and Prewithening

• Consider a stochastic process with a continuous spectrum defined over the frequency interval $[\omega_1, \omega_2]$. It can be shown that the expected value of the periodogram at frequency ω is proportional to

$$\int_{\omega_1}^{\omega_2} \mathcal{F}(\omega - \omega') f(\omega) \mathrm{d}\omega'. \tag{60}$$

The function $\mathcal{F}(\cdot)$ is the **Fejér's kernel**. Thus, the expected value of the periodogram is a convolution of the Fejér's kernel with the true spectrum. The Fejér's kernel is composed of a central lobe symmetric around frequency ω and of side lobes that are symmetric on each side of the central lobe. Because of these side lobes, the convolution also involves distant frequencies. Thus a transfer of power from one region of the spectrum to another via the kernel occurs, which is often called *leakage*. As a result, the periodogram can be biased for a finite time series. The bias is particularly severe for processes that have a high dynamic range, which is defined as

$$10 \log_{10} \left[\frac{\max_\omega f(\omega)}{\min_\omega f(\omega)} \right]. \tag{61}$$

For a white noise, the dynamic range is zero, so that the periodogram is an unbiased estimator of a purely random process. However, the dynamic range can be quite large for other processes, in which case the bias of the periodogram can be notable at some frequencies.

• There are two common techniques for lessening the bias in the periodogram: tapering and prewithening. **Tapering** modifies the kernel in the convolution, whereas **prewithening** preprocesses the data so as to reduce the dynamic range of the spectrum to be estimated. Both techniques are discussed in Percival and Walden (1993).

6.6 What to Look for in the Spectrum?

i. Are there any peaks?

ii. Is the spectrum large at low frequency (possible non-stationarity in the mean)?

iii. Is the spectrum large at high frequency (possible aliasing)?

iv. What is the general shape of the spectrum (possible suggestion of a particular stochastic model)?

• How to plot the spectrum? The axis for the frequency should be linear. The axis for the spectral estimates could be linear or logarithmic.

• How many values (N) are required to get a reasonable estimate of the spectrum? Many investigators consider that between 100 and 200 observations is the minimum.

• Which values to use for M, m? Values of M close to $2\sqrt{N}$ are sometimes recommended for the Tukey and Parzen windows. Values of m close to $N/40$ are sometimes recommended for the smoothed periodogram.

7 Bivariate Processes

7.1 Cross-Covariance and Cross-Correlation

- The **cross-covariance function** is (for a discrete process)

$$\gamma_{xy}(k) = \text{Cov}\left[X_t, Y_{t+k}\right]. \tag{62}$$

- $\gamma_{xy}(k)$ is **not** an even function ($\gamma_{xy}(k) \neq \gamma_{xy}(-k)$) but

$$\gamma_{xy}(k) = \gamma_{yx}(-k) \tag{63}$$

- The **cross-correlation function** is

$$\rho_{xy}(k) = \frac{\gamma_{xy}(k)}{\sqrt{\gamma_{xx}(0)\gamma_{yy}(0)}}, \tag{64}$$

where

$$\begin{aligned} \gamma_{xx}(k) &= \text{Cov}\left[X_t, X_{t+k}\right], && \text{(65a)} \\ \gamma_{yy}(k) &= \text{Cov}\left[Y_t, Y_{t+k}\right]. && \text{(65b)} \end{aligned}$$

- The cross-correlation function has the following properties:

$$\begin{aligned} \rho_{xy}(k) &= \rho_{yx}(-k), && \text{(66a)} \\ \mid \rho_{xy}(k) \mid &\leq 1. && \text{(66b)} \end{aligned}$$

- The cross-covariance function $\gamma_{xy}(k)$ is estimated by the sample cross-covariance function:

$$\begin{aligned} c_{xy}(k) &= \frac{1}{N}\sum_{t=1}^{N-k}\left(x_t - \bar{x}\right)\left(y_{t+k} - \bar{y}\right) \qquad \text{for } k = 0, 1, \ldots, N-1, && \text{(67a)} \\ c_{xy}(k) &= \frac{1}{N}\sum_{t=1-k}^{N}\left(x_t - \bar{x}\right)\left(y_{t+k} - \bar{y}\right) \qquad \text{for } k = -1, -2, \ldots, -(N-1). && \text{(67b)} \end{aligned}$$

- The cross-correlation function $\rho_{xy}(k)$ is estimated by the sample cross-correlation function:

$$r_{xy}(k) = \frac{c_{xy}(k)}{\sqrt{c_{xx}(0)c_{yy}(0)}}. \tag{68}$$

• To test for zero correlation between two time series, these should first be **filtered** to convert them to white noise before computing the cross-correlation function. For two uncorrelated series of white noise, it can be shown that

$$\mathrm{E}\left[r_{xy}(k)\right] \approx 0, \tag{69a}$$

$$\mathrm{Var}\left[r_{xy}(k)\right] \approx \frac{1}{N}. \tag{69b}$$

These expressions provide a test for zero correlation.

7.2 Cross-Spectrum

• The **cross-spectrum** is the Fourier transform of the cross-covariance function:

$$f_{xy}(\omega) = \frac{1}{\pi} \sum_{k=-\infty}^{\infty} \gamma_{xy}(k) e^{-i\omega k}, \tag{70}$$

over the range $0 < \omega < \pi$.

• Several functions derived from the cross-spectrum are helpful for interpreting the cross-spectrum. Note that $f_{xy}(\omega)$ is a complex function (unlike the autospectrum function), because $\gamma_{xy}(k)$ is not an even function. Thus, one can write

$$f_{xy}(\omega) = c(\omega) - \mathrm{i}q(\omega). \tag{71}$$

The real part of $\gamma_{xy}(k)$, $c(\omega)$, is the **co-spectrum**:

$$\begin{aligned} c(\omega) &= \frac{1}{\pi} \sum_{k=-\infty}^{\infty} \gamma_{xy}(k) \cos \omega k, && (72a) \\ &= \frac{1}{\pi} \left\{ \gamma_{xy}(0) + \sum_{k=1}^{\infty} \left[\gamma_{xy}(k) + \gamma_{yx}(k)\right] \cos \omega k \right\}. && (72b) \end{aligned}$$

The imaginary part of $\gamma_{xy}(k)$, $q(\omega)$, is the **quadrature spectrum**:

$$\begin{aligned} q(\omega) &= \frac{1}{\pi} \sum_{k=-\infty}^{\infty} \gamma_{xy}(k) \sin \omega k, && (73a) \\ &= \frac{1}{\pi} \left\{ \gamma_{xy}(0) + \sum_{k=1}^{\infty} \left[\gamma_{xy}(k) - \gamma_{yx}(k)\right] \sin \omega k \right\}. && (73b) \end{aligned}$$

The cross-spectrum function can also be written in polar form:

$$f_{xy}(\omega) = \alpha_{xy}(\omega) e^{i\phi_{xy}(\omega)}, \tag{74}$$

where

$$\alpha_{xy}(\omega) = \sqrt{c^2(\omega) + q^2(\omega)} \tag{75}$$

is the **cross-amplitude spectrum** and

$$\phi_{xy}(\omega) = \tan^{-1}\left[\frac{-q(\omega)}{c(\omega)}\right] \tag{76}$$

is the **phase spectrum**. A particularly useful function derived from the cross-spectrum is the (squared) **coherency**,

$$C(\omega) = \frac{c^2(\omega) + q^2(\omega)}{f_x(\omega) f_y(\omega)} \tag{77a}$$

$$= \frac{\alpha_{xy}^2(\omega)}{f_x(\omega) f_y(\omega)}, \tag{77b}$$

where $f_x(\omega)$ and $f_y(\omega)$ are the spectra of the individual processes $\{X_t\}$ and $\{Y_t\}$. The coherency measures the square of the linear correlation between the two components of the bivariate process at frequency ω. It is therefore analogous to the square of the usual correlation coefficient. Finally, consider the **gain spectrum**

$$G_{xy}(\omega) = \sqrt{\frac{f_y(\omega) C(\omega)}{f_x(\omega)}}, \tag{78a}$$

$$= \frac{\alpha_{xy}(\omega)}{f_x(\omega)}, \tag{78b}$$

which is essentially the regression coefficient of the process $\{Y_t\}$ on the process $\{X_t\}$ at frequency ω. An equivalent definition exists for $G_{xy}(\omega)$:

$$G_{yx}(\omega) = \frac{\alpha_{xy}(\omega)}{f_y(\omega)}. \tag{79}$$

• In general, three functions have to be plotted against frequency for a complete description of the relationship between two time series in the frequency domain. A useful trio, whose each component has a relatively straightforward physical interpretation, is coherency, phase, and gain (see below, the analysis of linear systems).

7.2.1 Estimation

• Two basic approaches exist to estimate the cross-spectrum. In a first approach the cross-spectrum is estimated by truncating and weighting the (sample) cross-covariance function. Thus, the co-spectrum and quadrature spectrum are estimated as:

$$\hat{c}(\omega) = \frac{1}{\pi} \sum_{k=-M}^{M} \lambda_k c_{xy}(k) \cos \omega k, \tag{80a}$$

$$\hat{q}(\omega) = \frac{1}{\pi} \sum_{k=-M}^{M} \lambda_k c_{xy}(k) \sin \omega k. \tag{80b}$$

The truncation point M and the lag window $\{\lambda_k\}$ are chosen in a similar way to those used for the individual spectral estimates, with the Tukey and Parzen windows being the most popular. From $\hat{c}(\omega)$ and $\hat{q}(\omega)$, the other cross-spectral functions can be estimated, i.e.,

$$\hat{\alpha}_{xy}(\omega) = \sqrt{\hat{c}^2(\omega) + \hat{q}^2(\omega)}, \tag{81a}$$

$$\tan \hat{\phi}_{xy}(\omega) = -\frac{\hat{q}(\omega)}{\hat{c}(\omega)}, \tag{81b}$$

$$\hat{C}(\omega) = \frac{\hat{\alpha}^2_{xy}(\omega)}{\hat{f}_x(\omega)\hat{f}_y(\omega)}. \tag{81c}$$

• The second approach for estimating the cross-spectrum is to smooth a function called the cross-periodogram $I_{xy}(\omega)$. The real and imaginary parts of the cross-periodogram at the discrete frequency $\omega_p = 2\pi p/N$ are

$$\hat{c}(\omega_p) = \frac{N}{4\pi m} \sum_{q=p-m_*}^{p+m_*} \left(a_{qx}a_{qy} + b_{qx}b_{qy}\right), \tag{82a}$$

$$\hat{q}(\omega_p) = \frac{N}{4\pi m} \sum_{q=p-m_*}^{p+m_*} \left(a_{qx}b_{qy} - a_{qy}b_{qx}\right), \tag{82b}$$

where (a_{px}, b_{px}) and (a_{py}, b_{py}) are, respectively, the Fourier coefficients of x_t and y_t at ω_p, and $m = 2m_* + 1$. These estimates may then be used to estimate the cross-amplitude, phase, etc., as before.

• Note that estimates of phase and cross-amplitude are imprecise when the coherency is relatively small.

• Confidence intervals for the estimators of phase spectrum and squared coherency have been developed. For both estimators the normality assumption is made. A discussion is available, for example, in Jenkins and Watts (1968).

7.2.2 Interpretation

• For two time-series that can be viewed as being on a similar footing, the coherency spectrum is probably the most useful function.

• The other functions, such as the phase spectrum and the gain spectrum, are probably better understood via the analysis of linear systems.

8 Linear Systems

8.1 Introduction

• One process, noted $\{x_t\}$ (discrete) or $\{x(t)\}$ (continuous), is regarded as the '**input**', whereas another process, noted $\{y_t\}$ (discrete) or $\{y(t)\}$ (continuous), is regarded as the '**output**'.

• The systems examined here are **linear** and **time-invariant**. A system is considered as linear if a linear combination of the inputs, say $\lambda_1 x_1(t) + \lambda_2 x_2(t)$, produces the linear combination of the outputs $\lambda_1 y_1(t) + \lambda_2 y_2(t)$, where λ_1, λ_2 are any constants. A system is considered as time-invariant if, in the case where $\{x(t)\}$ produces $\{y_t\}$, $\{x(t+\tau)\}$ produces $\{y(t+\tau)\}$, i.e., the relation between the input and output does not change with time.

• The study of linear systems is useful for (1) a better understanding of the relationship between two time series (e.g., bivariate spectral analysis) and (2) a better understanding of the properties of linear filters such as used for detrending a time series.

8.2 Linear Systems in Time Domain

• A linear time-invariant filter is defined in continuous time as

$$y(t) = \int_{-\infty}^{\infty} h(u)x(t-u)\mathrm{d}u, \tag{83}$$

and in discrete time as

$$y_t = \sum_{k=-\infty}^{\infty} h_k x_{t-k}. \tag{84}$$

The weight function $h(u)$ in continuous time or h_k in discrete time is called the **impulse response function**.

• The system is said to be **physically realizable** if

$$h(u) = 0 \qquad \text{for } u < 0, \tag{85a}$$

$$h_k = 0 \qquad \text{for } k < 0. \tag{85b}$$

8.3 Linear Systems in Frequency Domain

8.3.1 Frequency Response Function

• The **frequency response function** or **transfer function** is the Fourier transform of the impulse response function. In continuous time,

$$H(\omega) = \int_{-\infty}^{\infty} h(u) e^{-i\omega u} \mathrm{d}u \quad (0 < \omega < \infty). \tag{86}$$

In discrete time,

$$H(\omega) = \sum_{k=-\infty}^{\infty} h_k e^{-i\omega k} \qquad (0 < \omega < \pi). \tag{87}$$

• Theorem:

A sinuoidal input to a linear system gives rise, in the steady state, to a sinusoidal output at the **same** *frequency. The amplitude of the sinusoid may change and there may also be a phase shift*

• Example 1. Consider the continuous input

$$x(t) = \cos \omega t. \tag{88}$$

It can be shown that the corresponding output is

$$y(t) = G(\omega) \cos \left[\omega t + \phi(\omega)\right], \tag{89}$$

where $G(\omega)$ is the **gain** of the system and $\phi(\omega)$ is the **phase shift**.

• Example 2. Consider the continuous input

$$x(t) = e^{i\omega t}. \tag{90}$$

Then it can be shown that the corresponding output is

$$\begin{aligned} y(t) &= G(\omega) e^{i\phi(\omega)} x(t), \\ &= H(\omega) x(t), \end{aligned} \tag{91}$$

where $H(\omega)$ is the frequency response function of the system. The gain is the real part of $H(\omega)$, i.e., $G(\omega) = |H(\omega)|$. It describes how an input at frequency ω is damped or amplified by the filter. It can correspond, for example, to a **low-pass filter** or to a **high-pass filter**.

• More generally, an input consisting of a sum of sinusoidal components,

$$x(t) = \sum_j A_j(\omega_j) e^{i\omega_j t}, \tag{92}$$

gives rises to an output

$$y(t) = \sum_j A_j(\omega_j) H(\omega_j) e^{i\omega_j t}. \tag{93}$$

• Theorem:

Consider a stable linear system with gain function $G(\omega)$. Suppose that the input $x(t)$ is a stationary process with continuous spectrum $f_x(\omega)$. Then the output $y(t)$ is also a stationary process, whose spectrum is given by

$$f_y(\omega) = G^2(\omega) f_x(\omega). \tag{94}$$

8.4 Identification of Linear Systems

• In many situations, the structure of the system is not known, whereas the input and output are. The **identification** of the system is concerned with inferring the properties of the system from the relationship between the input and the output.

• Consider the physically realizable continuous output

$$Y(t) = \int_0^\infty h(u)X(t-u)\mathrm{d}u + N(t) \tag{95}$$

where $N(t)$ is a noise that has zero mean and is uncorrelated with the input. For convenience it is assumed that $\mathrm{E}[X(t)] = 0$, so that $\mathrm{E}[Y(t)] = 0$. It is found that

$$f_{xy}(\omega) = H(\omega) f_x(\omega). \tag{96}$$

Thus, knowledge about the frequency response function of the system is available from knowledge about the spectrum of the input, $f_x(\omega)$, and knowledge about the cross-spectrum of the input and output, $f_{xy}(\omega)$. Thus, we have

$$\hat{H}(\omega) = \frac{\hat{f}_{xy}(\omega)}{\hat{f}_x(\omega)}. \tag{97}$$

Furthermore, the real and imaginary parts of $H(\omega) = G(\omega)e^{i\phi(\omega)}$ can be estimated separately. The gain can be estimated as follows

$$\begin{aligned} \hat{G}(\omega) &= |\hat{H}(\omega)| \\ &= \left|\frac{\hat{f}_{xy}(\omega)}{\hat{f}_x(\omega)}\right| \\ &= \frac{|\hat{f}_{xy}(\omega)|}{\hat{f}_x(\omega)} \\ &= \frac{\hat{\alpha}_{xy}(\omega)}{\hat{f}_x(\omega)}. \end{aligned} \tag{98}$$

Thus, knowledge about the gain is provided by the estimate of the cross-amplitude spectrum, $\hat{\alpha}_{xy}(\omega)$, and the estimate of the spectrum of the input, $\hat{f}_x(\omega)$. The phase shift can be estimated from

$$\tan \hat{\phi}_{xy}(\omega) = -\frac{\hat{q}(\omega)}{\hat{c}(\omega)}. \tag{99}$$

Knowledge about the phase shift is available from estimates of the co-spectrum, $\hat{c}(\omega)$, and of the quadrature spectrum, $\hat{q}(\omega)$.

• The effect of the noise can be elucidated. Consider the discrete system

$$Y_t = \sum_{k=0}^{\infty} h_k X_{t-k} + N_t. \tag{100}$$

Again it is assumed that the noise has zero mean and is uncorrelated with the input. It is also assumed that $\mathrm{E}[X_t] = 0$, so that $\mathrm{E}[Y_t] = 0$. It can be shown that

$$f_y(\omega) = G(\omega)^2 f_x(\omega) + f_n(\omega), \tag{101}$$

where $f_n(\omega)$ is the noise spectrum. This result shows how the presence of the noise corrupts the relationship between the input and the output in the frequency domain. The noise contribution can be expressed as

$$f_n(\omega) = f_y(\omega)\left[1 - C(\omega)\right]. \tag{102}$$

Thus the noise spectrum can be estimated from an estimate of the (auto)spectrum of the output and an estimate of the coherency,

$$\hat{f}_n(\omega) = \hat{f}_y(\omega)\left[1 - \hat{C}(\omega)\right]. \tag{103}$$

• The above results show how to identify a linear system by cross-spectral analysis. They also provide guidance for the interpretation of the various functions derived from the cross-spectrum, in particular, the gain, phase, and coherency.

9 Some Practical Recommendations

• Do not underestimate the value of the time plot (use different aspect ratios, etc.).

• Consider whether your time series is long enough to conduct meaningful analyses in the time domain and (or) frequency domain.

• Consider that most of the theory is generally based on several assumptions, e.g.,

i. The data are evenly spaced in time,

ii. The process that generated the data is stationary,

iii. The data are normally distributed.

• Produce the correlogram. Compare with correlogram for standard probability models (test for randomness, etc.)

• Produce lag scatter plots, in particular for lags for which $|r_k| > 2/\sqrt{N}$. Check for the (non)linear nature of the relationship at these lags.

• Produce the periodogram, keeping in mind its poor bias and variance properties.

• Consider different bandwidths and perhaps also different procedures of spectral estimation. Report the actual procedure and bandwith used in your work.

• Consider different interpolation schemes and time steps for interpolation. Possible interpolation schemes are

i. Linear interpolation,

ii. Smoothing splines,

iii. Data averaging.

• Use spectral procedures that you understand (otherwise, risk of mis-interpretation).

• Use both a linear scale and a logarithmic scale for the spectral estimates.

• Be aware of the assumptions involved in the construction of confidence intervals. Check for features such as non-normality, the presence of outliers, variable variance, etc.

• Filter the data before testing the cross-correlation between two time series.

• Align the data before generating the estimate of the cross-spectrum so at to reduce its bias.

• Do not hesitate to consult knowledgeable people for opinion & advice!

10 Final Remarks

• The field has no consistent terminology: different authors use different symbols and (or) different terms for the same concept.

• Several techniques are not considered in these notes, such as

i. Maximum entropy method,

ii. Multitaper technique,

iii. Wavelet analysis,

iv. Singular spectrum.