Chapter 9

Time series I: time-domain

Aim: The aim of this lecture is to provide a brief introduction to the analysis of time series with emphasis on the time domain approach.

9.1 Introduction

The variation in time of environmental quantities can be studied using the rich branch of statistics known as **time series analysis**. A **discrete** (as opposed to **continuous**) **time series**[1] is a sequence of observed values $\{x_1, x_2, \ldots, x_n\}$ measured at discrete times $\{t_1, t_2, \ldots, t_n\}$. Climatological time series are most often sampled at **regular** intervals $t_k = k\tau$ where τ is the **sampling period**.

The main aims of time series analysis are to **explore** and extract **signals** (patterns) contained in time series, to make **forecasts** (i.e. future predictions in time), and to use this knowledge to optimally **control** processes.

The two main approaches used in time series analysis are **time domain** and **spectral (frequency) domain**. The time domain approach represents time series directly as functions of time, whereas the spectral domain approach represents time series as spectral expansions of either fourier modes or wavelets.

[1] NOTE: time series NOT timeseries !

9.2 Time series components

A lot can be learnt about a time series by plotting x_k versus t_k in a **time series plot**. For example, the time series plot in Figure 9.1 shows the evolution of monthly mean sea-level pressures measured at Darwin in northern Australia.

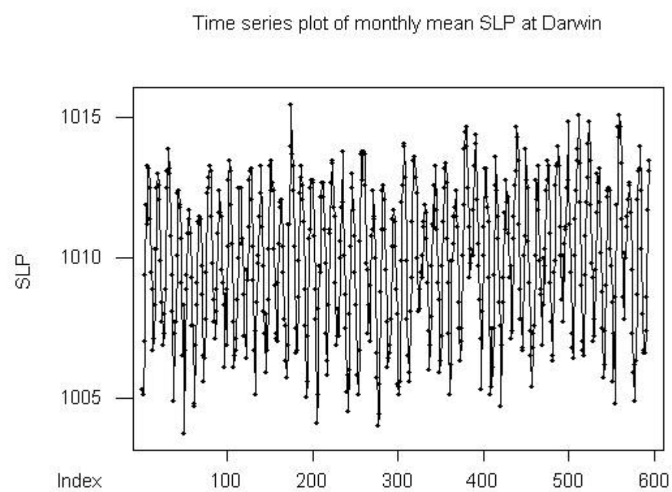

Figure 9.1: Time series of the montly mean sea-level pressure observed at Darwin in northern Australia over the period January 1950 to July 2000.

A rich variety of structures can be seen in the series that include:

- **Trends** - long-term changes in the mean level. In other words, a smooth regular component consisting primarily of fourier modes having periods longer than the length of the time series. Trends can be either **deterministic** (e.g. world population) or **stochastic**. Stochastic trends are not necessarily monotonic and can go up and down (e.g. North Atlantic Oscillation). Extreme care should be exercised in extrapolating trends and it is wise to always refer to them in the past tense.

- **(Quasi-)periodic signals** - having clearly marked cycles such as the **seasonal component** (annual cycle) and interannual phenomena such as El Niño and business cycles. For peridocities approaching the length of the time series, it becomes extremely difficult to discriminate these from stochastic trends.

- **Irregular component** - random or chaotic noisy residuals left over after removing all trends and (quasi-)periodic components. They are (second-order) **stationary** if they have mean level and variance that remain constant in time and can often be modelled as filtered noise using time series models such as ARIMA.

Some time series are best represented as sums of these components (**additive**) while others are best represented as products of these components (**multiplicative**). Multiplicative series can quite often be made additive by normalizing using the logarithm transformation (e.g. commodity prices).

9.3 Filtering and smoothing

It is often useful to either **low-pass filter** (**smooth**) time series in order to reveal low-frequency features and trends, or to **high-pass filter** (**detrend**) time series in order to isolate high frequency transients (e.g. storms).

Some of the most commonly used filters are:

- **Moving average MA(q)**
  This simple class of low-pass filters is obtained by applying a running mean of length q to the original series

$$y_t = \frac{1}{q} \sum_{k=-q/2}^{q/2} x_{t+k} \tag{9.1}$$

  For example, the three month running mean filter MA(3) is useful for crudely filtering out intraseasonal oscillations. Note, however, that the sharp edges in the weights of this filter can causing spurious ringing (oscillation) and leakage into the smoothed output.

- **Binomial filters** $(1/2, 1/2)^m$

  These smoother low-pass filters are obtained by repeatedly applying the MA(2) filter that has weights $(1/2, 1/2)$. For example, with $m = 4$ applications the binomial filter weights are given by $(1/2, 1/2)^4 = (1, 4, 6, 4, 1)/16$ which tail off smoothly towards zero near the edges. For large large numbers of applications, the weights become Gaussian and the filtering approximates Gaussian kernel smoothing.

- **Holt exponential smoother**

  This simple and widely used recursive filter is obtained by iterating

$$y_t = \alpha x_t + (1 - \alpha) y_{t-1} \tag{9.2}$$

  where α is a tunable smoothing parameter. This low-pass filter gives most weight to most recent historical values and so provides the basis for a sensible forecasting procedure when applied to trend, seasonal, and irregular components (Holt-Winters forecasting).

- **Detrending (high-pass) filters**

  High-pass filtering can most easily be performed by subtracting a suitably low-pass filtered series from the original series. The detrended residuals $x_t - y_t$ contain the high-pass component of x. For example, the backward difference detrending filter $\Delta x = x_t - x_{t-1}$ is simply twice the residual obtained by removing a MA(2) low-pass filtered trend from a time series. It is very efficient at removing stochastic trends and is often used to detrend non-stationary time series (e.g. random walks in commodity prices).

9.4 Serial correlation

Successive values in time series are often correlated with one another. This **persistence** is known as **serial correlation** and leads to increased spectral power at lower frequencies (**redness**). It needs to be taken into account when testing significance, for example, of the correlation between two time series. Among other things, serial correlation (and trends) can severely reduce the

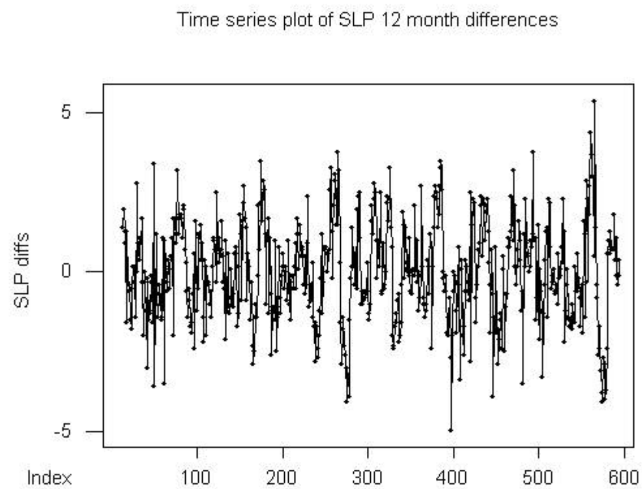

Figure 9.2: Time series plot of one-year backward differences in monthly mean sea-level pressure at Darwin from the period January 1951 to July 2000. The differencing has efficiently removed both the seasonal component and the long-term trend thereby revealing short-term interannual variations.

effective number of degrees of freedom in a time series. Serial correlation can be explored by estimating the sample **autocorrelation coefficients**

$$r_k = \frac{\frac{1}{n}\sum_{i=k+1}^{n}(x_i - \overline{x})(x_{i-k} - \overline{x})}{\frac{1}{n}\sum_{i=k+1}^{n}(x_i - \overline{x})^2} \quad (9.3)$$

where $k = 0, 1, 2, \ldots$ is the **time lag**. The zero lag coefficient r_0 is always equal to one by definition, and higher lag coefficients generally damp towards small values with increasing lag. Only autocorrelation coefficients with lags less than $n/4$ are sufficiently well-sampled to be worth investigation.

The autocorrelation coefficients can be plotted versus lag in a plot known as a **correlogram**. The correlogram for the Darwin series is shown in Fig. 9.3. Note the fast drop off in the **autocorrelation function (a.c.f.)** for time lags greater than 12 months. The lag-1 coefficient is often (but not always) adequate for giving a rough indication of the amount of serial correlation in a series. A rough estimate of the **decorrelation time** is given by