

Chapter 11

Stationary and non-stationary time series

G. P. Nason

Time series analysis is about the study of data collected through time. The field of time series is a vast one that pervades many areas of science and engineering particularly statistics and signal processing: this short article can only be an advertisement.

Hence, the first thing to say is that there are several excellent texts on time series analysis. Most statistical books concentrate on stationary time series and some texts have good coverage of “globally non-stationary” series such as those often used in financial time series. For a general, elementary introduction to time series analysis the author highly recommends the book by (Chatfield 2003). The core of Chatfield’s book is a highly readable account of various topics in time series including time series models, forecasting, time series in the frequency domain and spectrum estimation and also linear systems. More recent editions contain useful, well-presented and well-referenced information on important new research areas. Of course, there are many other books: ones the author finds useful are Priestley (1983), Diggle (1990), Brockwell and Davis (1991), and Hamilton (1994). The book by Hannan (1960) is concise (but concentrated) and Pole *et al.* (1994) is a good introduction to a Bayesian way of doing time series analysis. There are undoubtedly many more books.

This article is a brief survey of several kinds of time series model and analysis. Section 11.1 covers stationary time series which, loosely speaking, are those whose statistical properties remain constant over time. Of course, for many real applications the stationarity assumption is not valid. Gener-

ally speaking one should not use methods designed for stationary series on non-stationary series as you run the risk of obtaining completely misleading answers. A tricky question is how can you know whether a time series is stationary or not? There are various tests for stationarity. As well as suffering from all of the usual issues of statistical testing (what significance should I use? what power do I get?) tests of stationarity tend to test against particular alternative models of particular types of non-stationarity. For example, test A might well be powerful at picking up non-stationarities of type A but have no power at detecting those of type B. Section 11.2 briefly considers such tests in the context of describing some non-stationary time series models and also exhibits a technique for the analysis and modelling of both a seismic and a medical time series. Section 11.3 lists some topics that we do not cover in detail but are nevertheless important for the practical use of time series analysis.

11.1 Stationary Time Series

Let us begin this section with the following wonderful quotation:

“Experience with real-world data, however, soon convinces one that both stationarity and Gaussianity are fairy tales invented for the amusement of undergraduates.”

(Thomson 1994)

Bearing this in mind, stationary models form the basis for a huge proportion of time series analysis methods. As is true for a great deal of mathematics we shall begin with very simple building blocks and then build structures of increasing complexity. In time series analysis the basic building block is the purely random process.

11.1.1 Purely random process

A *purely random process* is a stochastic process, $\{\epsilon_t\}_{t=-\infty}^{\infty}$, where each element ϵ_t is (statistically) independent of every other element, ϵ_s for $s \neq t$, and each element has an identical distribution.

Example 1. As a simple example suppose ϵ_t has a Gaussian distribution with mean μ and variance of σ^2 , the same for all integers t (which we will write as $\epsilon_t \sim N(\mu, \sigma^2)$).

There are several important features of this definition and example. First, note that we immediately define a stochastic process here in terms of its behaviour at time t where t is an integer value. Of course, time series exist where t is some other kind of quantity. If t can be mapped to the

integers then there is no problem but important differences arise if t can be a continuous variable (e.g. any real number in the interval $[0, 1]$ for example). The second point, which the example brings out, is that we considered the value of the time series ϵ_t at a given time t to be a continuous random variable (that is, it can potentially take any real value). Many time series, especially those in volcanology, can take other kinds of values: for example, count values such as those recording number of eruptions, or other kinds of discrete events. For the latter kind of discrete valued time series the book by Macdonald and Zucchini (1997) is a useful reference. Note also that we use the braces $\{\}$ to indicate the whole stochastic process but drop them when referring to a generic member, e.g. ϵ_t . In what follows if we do not mention the limits then we assume the process indices range from $t = -\infty$ to $t = \infty$.

By definition it is immediate that the mean and the variance of the above purely random process are

$$\mathbb{E}(\epsilon_t) = \mu \text{ and } \text{var}(\epsilon_t) = \sigma^2. \quad (11.1)$$

These are the basic first and second order statistics of a time series (the order relates to the highest power in the integral defining the statistic: for expectation it is $\int x^1 f(x) dx$, for the variance it is $\int x^2 f(x) dx$.) In time series the key aspect is how observations are related to each other in time. This concept is formalized through the *autocovariance* between elements which measures the degree of second order variation between two elements at two different times. To be mathematically precise the *autocovariance* between X_t and X_s for some process $\{X_t\}$ is defined to be

$$\text{cov}(X_t, X_s) = \mathbb{E}[\{X_t - \mathbb{E}(X_t)\} \{X_s - \mathbb{E}(X_s)\}]. \quad (11.2)$$

Third and higher order quantities can be defined but for a great deal of time series analysis they are largely ignored (it is amazing how only second-order quantities can occupy us).

11.1.2 Stationarity

Loosely speaking a stationary process is one whose statistical properties do not change over time. More formally, a *strictly stationary* stochastic process is one where given t_1, \dots, t_ℓ the joint statistical distribution of $X_{t_1}, \dots, X_{t_\ell}$ is the same as the joint statistical distribution of $X_{t_1+\tau}, \dots, X_{t_\ell+\tau}$ for *all* ℓ and τ . This is an extremely strong definition: it means that all moments of all degrees (expectations, variances, third order and higher) of the process, *anywhere* are the same. It also means that the joint distribution of (X_t, X_s) is the same as (X_{t+r}, X_{s+r}) and hence cannot depend on s or t but *only on the distance* between s and t , i.e. $s - t$.

Since the definition of *strict stationarity* is generally too strict for everyday life a weaker definition of *second order* or *weak stationarity* is usually used. Weak stationarity means that mean and the variance of a stochastic process do not depend on t (that is they are constant) and the autocovariance between X_t and $X_{t+\tau}$ only can depend on the *lag* τ (τ is an integer, the quantities also need to be finite). Hence for stationary processes, $\{X_t\}$, the definition of autocovariance is

$$\gamma(\tau) = \text{cov}(X_t, X_{t+\tau}), \quad (11.3)$$

for integers τ . It is vital to remember that, for the real world, the autocovariance of a stationary process is a model, albeit a useful one. Many actual processes are not stationary as we will see in the next section. Having said this much fun can be had with stationary stochastic processes!

One also routinely comes across the *autocorrelation* of a process which is merely a normalized version of the autocovariance to values between -1 and 1 and commonly uses the Greek letter ρ as its notation:

$$\rho(\tau) = \gamma(\tau)/\gamma(0), \quad (11.4)$$

for integers τ and where $\gamma(0) = \text{cov}(X_t, X_t) = \text{var}(X_t)$.

Example 1, continued For the purely random process $\{\epsilon_t\}$ defined above, formula (11.1) shows that the mean and variance are constant functions in time. Moreover, since ϵ_t is independent of ϵ_s for $t \neq s$ the correlation between them must always be zero too. Hence $\gamma(\tau)$ and $\rho(\tau)$ are always zero for $\tau \neq 0$ and hence “do not depend on τ ”. Thus, the purely random process must be at least second-order stationary. Indeed, it is also strictly stationary because the elements are independently and identically distributed (the joint distribution in the definition of strict stationarity immediately decomposes into a product of ℓ terms with each term being the same). Next we start constructing some more complex and interesting models.

11.1.3 MA, AR and ARMA models

This section considers some basic probability models extensively used for modelling time series.

Moving Average models. Probably the next simplest model is that constructed by simple linear combinations of lagged elements of a purely random process, $\{\epsilon_t\}$ with $\mathbb{E}\epsilon_t = 0$. A *moving average* process, $\{X_t\}$, of order q is defined by

$$X_t = \beta_0\epsilon_t + \beta_1\epsilon_{t-1} + \cdots + \beta_q\epsilon_{t-q} = \sum_{i=0}^q \beta_i\epsilon_{t-i} \quad (11.5)$$

and the shorthand notation is $\text{MA}(q)$. Usually with a newly defined process it is of interest to discover its statistical properties. For an $\text{MA}(q)$ process the mean is simple to find (since the expectation of a sum is the sum of the expectations):

$$\mathbb{E}(X_t) = \mathbb{E} \left(\sum_{i=0}^q \beta_i \epsilon_{t-i} \right) = \sum_{i=0}^q \beta_i \mathbb{E}(\epsilon_{t-i}) = 0, \quad (11.6)$$

because $\mathbb{E}(\epsilon_r) = 0$ for any r . A similar argument can be applied for the variance calculation:

$$\text{var}(X_t) = \text{var} \left(\sum_{i=0}^q \beta_i \epsilon_{t-i} \right) = \sum_{i=0}^q \beta_i^2 \text{var}(\epsilon_{t-i}) = \sigma^2 \sum_{i=0}^q \beta_i^2, \quad (11.7)$$

since $\text{var}(\epsilon_r) = \sigma^2$ for all t .

The autocovariance is slightly more tricky to work out.

$$\gamma(\tau) = \text{cov}(X_t, X_{t-\tau}) \quad (11.8)$$

$$= \text{cov} \left(\sum_{i=0}^q \beta_i \epsilon_{t-i}, \sum_{j=0}^q \beta_j \epsilon_{t-\tau-j} \right) \quad (11.9)$$

$$= \sum_{i=0}^q \sum_{j=0}^q \beta_i \beta_j \text{cov}(\epsilon_{t-i}, \epsilon_{t-\tau-j}) \quad (11.10)$$

$$= \sigma^2 \sum_{i=0}^q \sum_{j=0}^q \beta_i \beta_j \delta_{j, i+\tau} \quad (11.11)$$

where $\delta_{u,v}$ is the Kronecker delta which is 1 for $u = v$ and zero otherwise (this arises because of the independence of the ϵ values. Thus since $\delta_{j, i+\tau}$ is involved only terms in the j sum where $j = i + \tau$ survive). Hence continuing the summation gives

$$\gamma(\tau) = \sum_{i=0}^{q-\tau} \beta_i \beta_{i+\tau}. \quad (11.12)$$

In other words, the β_j becomes $\beta_{i+\tau}$ and *also* the index of summation ranges only up to $q - \tau$ since the largest $\beta_{i+\tau}$ occurs for $i = q - \tau$.

The formula for the autocovariance of an $\text{MA}(q)$ process is fascinating: it is effectively the convolution of $\{\beta_i\}$ with itself (an “autoconvolution”). One of the most important features of an $\text{MA}(q)$ autocovariance is that it is zero for $\tau > q$. The reason for its importance is that when one is confronted with an actual time series x_1, \dots, x_n one can compute the *sample autocovariance* given by:

$$c(\tau) = \sum_{i=1}^{n-\tau} (x_i - \bar{x})(x_{i+\tau} - \bar{x}). \quad (11.13)$$

for $\tau = 0, \dots, n - 1$. The *sample autocorrelation* can be computed as $r(\tau) = c(\tau)/c(0)$. If, when one computes the sample autocovariance, it cuts off at a certain lag q , i.e. it is effectively zero for lags of $q + 1$ or higher, then one can postulate the MA(q) model in (11.5) as the underlying probability model. There are other checks and tests that one can make but comparison of the sample autocovariance with reference values, such as the model autocovariance given in (11.12), is a major first step in *model identification*.

Also, at this point one should question what one means by “effectively zero”. The sample autocovariance is an empirical statistic calculated from the random sample at hand. If more data in the time series were collected, or another sample stretch used then the sample autocovariance would be different (although for long samples and stationary series the probability of a large difference should be very small). Hence, sample autocovariances (and autocorrelations) are necessarily random quantities and hence “is effectively zero” translates into a statistical hypothesis test on whether the true autocorrelation is zero or not.

Finally, whilst we are on the topic of sample autocovariances notice that at the extremes of the range of τ :

$$c(0) = \sum_{i=0}^n (x_i - \bar{x})^2, \quad (11.14)$$

which also happens to be the *sample variance* and

$$c(n - 1) = x_1 x_n. \quad (11.15)$$

The lesson here is that $c(0)$, which is an estimate of $\gamma(0)$, is based on n pieces of information, whereas, $c(n - 1)$, an estimate of $\gamma(n - 1)$, is only based on 1 piece of information. Hence, it is easy to see that estimates of sample autocovariances for higher lags are more unreliable for those of smaller lags.

Figure 11.1 (top) shows a simulation from the MA(2) process $X_t = \epsilon_t - 0.8\epsilon_{t-1} + 0.9\epsilon_{t-2}$. Figure 11.1 (bottom) shows the empirical sample autocorrelation function (acf) of the simulated series. Notice that the sample acf does seem to “cut-off” at lag 2. The “true” theoretical values of the autocorrelation function $\rho(\tau)$ given by formulae (11.4) and (11.12) are $\rho(0) = 1$ (of course), $\rho(1) \approx -0.62$ and $\rho(2) \approx 0.367$. The horizontal dashed lines in the bottom plot of Figure 11.1 are approximate 95% significance bands. Any coefficient which is outside of these bands (like the ones in the figure at lags 1 and 2) are deemed significant, or at least worth further contemplation. It is also interesting to note that although $\rho(\tau) = 0$ for $\tau > 2$ the sample autocorrelations for lags 10 to 14 inclusive all look quite large:

this is merely an illustration of the comment that the sample acf function is a random quantity and some care and experience is required to prevent reading too much into them.

Autoregressive models. The other basic development from a purely random process is the autoregressive model which, as its name suggests, models a process where future values somehow depend on the recent past. More formally, $\{X_t\}$ follows an AR(p) model which is characterised by

$$X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \cdots + \alpha_p X_{t-p} + \epsilon_t, \quad (11.16)$$

where $\{\epsilon_t\}$ is a purely random process. For the purposes of discussion we shall follow time-honoured history and first consider the simplest AR model: AR(1). Here we simplify notation and define $\{X_t\}$ by

$$X_t = \alpha X_{t-1} + \epsilon_t. \quad (11.17)$$

How can we answer the simple sounding question: what is the expectation of X_t ? The obvious answer to this is to dumbly apply the expectation operator \mathbb{E} to formula (11.17). However, knowing that $\mathbb{E}X_t = \mathbb{E}X_{t-1}$ does not get us very far, especially if it is assumed (or discovered) that $\{X_t\}$ is stationary in which case we've got $\mu = \mu!$ Another, more successful approach, is to recurse formula (11.17):

$$X_t = \alpha X_{t-1} + \epsilon_t \quad (11.18)$$

$$= \alpha (\alpha X_{t-2} + \epsilon_{t-1}) + \epsilon_t \quad (11.19)$$

$$= \alpha^2 X_{t-2} + \alpha \epsilon_{t-1} + \epsilon_t \quad (11.20)$$

$$= \alpha^2 (\alpha X_{t-3} + \epsilon_{t-2}) + \alpha \epsilon_{t-1} + \epsilon_t \quad (11.21)$$

$$= \alpha^3 X_{t-3} + \alpha^2 \epsilon_{t-2} + \alpha \epsilon_{t-1} + \epsilon_t. \quad (11.22)$$

A clear pattern emerges and one can write (backwards):

$$X_t = \epsilon_t + \alpha \epsilon_{t-1} + \alpha^2 \epsilon_{t-2} + \cdots + \alpha^r \epsilon_{t-r} + \cdots = \sum_{i=0}^{\infty} \alpha^i \epsilon_{t-i}. \quad (11.23)$$

In effect, we have turned an AR(1) process into a moving average process of *infinite* order (an MA(∞) process). This new representation means that it is easy to compute the expectation:

$$\mathbb{E}X_t = \mathbb{E} \sum_{i=0}^{\infty} \alpha^i \epsilon_{t-i} = \sum_{i=0}^{\infty} \alpha^i \mathbb{E}\epsilon_{t-i} = 0, \quad (11.24)$$

since we assumed that $\mathbb{E}\epsilon_t = 0$. As for the variance we can do this easily

since we the ϵ_t are independent:

$$\text{var } X_t = \text{var} \sum_{i=0}^{\infty} \alpha^i \epsilon_{t-i} \quad (11.25)$$

$$= \sigma^2 \sum_{i=0}^{\infty} \alpha^{2i}. \quad (11.26)$$

This latter sum is only finite if $|\alpha| < 1$ in which case using basic knowledge about geometric sums one obtains

$$\text{var } X_t = \sigma^2 / (1 - \alpha^2). \quad (11.27)$$

If $|\alpha| \geq 1$ then the sum in (11.26) does not converge and the process X_t is not stationary (one can see this as the variance would increase with t). The case $\alpha = 1$ is an interesting one. Here the model is

$$X_t = X_{t-1} + \epsilon_t \quad (11.28)$$

and is known as a *random walk* often mooted as a model for “the stock market”.

Another way of deriving the MA(∞) model for an AR process is by introducing the shift operator, B , defined as

$$BX_t = X_{t-1} \quad (11.29)$$

valid for any process $\{X_t\}$. Then we can rewrite the AR model in (11.17) as

$$X_t - \alpha X_{t-1} = \epsilon_t \quad (11.30)$$

$$(1 - \alpha B)X_t = \epsilon_t \quad (11.31)$$

$$X_t = (1 - \alpha B)^{-1} \epsilon_t \quad (11.32)$$

$$X_t = (1 + \alpha B + \alpha^2 B^2 + \dots) \epsilon_t \quad (11.33)$$

$$X_t = \epsilon_t + \alpha \epsilon_{t-1} + \alpha^2 \epsilon_{t-2} + \dots, \quad (11.34)$$

where the quantity $(1 - \alpha B)^{-1}$ in (11.32) gets transformed into $1 + \alpha B + \alpha^2 B^2 + \dots$ in (11.33) by the binomial expansion.

The final quantity that we need to concern ourselves with is the autocovariance. A simple way of seeing what the autocovariance is to assume that the process is stationary and multiply both sides of (11.17) by $X_{t-\tau}$ and take expectations:

$$\mathbb{E}(X_t X_{t-\tau}) = \alpha \mathbb{E}(X_{t-1} X_{t-\tau}) + \mathbb{E}(\epsilon_t X_{t-\tau}). \quad (11.35)$$

Since the $\mathbb{E}X_t = 0$ the first terms on both sides of this equation are the autocovariances of X_t : in other words $\mathbb{E}(X_t X_{t-\tau}) = \text{cov}(X_t, X_{t-\tau}) = \gamma(\tau)$ and similarly $\mathbb{E}(X_{t-1} X_{t-\tau}) = \gamma(\tau - 1)$. Finally, note from the model formula (11.17) that $X_{t-\tau}$ will only contain ϵ terms but only for ϵ_i with $i < t - \tau$, i.e. $X_{t-\tau}$ only includes *past* ϵ relative to ϵ_t . This means that ϵ_t is independent of $X_{t-\tau}$ and hence $\mathbb{E}(\epsilon_t X_{t-\tau}) = \mathbb{E}\epsilon_t \mathbb{E}X_{t-\tau} = 0$ since the purely random process here has zero mean. Thus (11.35) turns into

$$\gamma(\tau) = \alpha\gamma(\tau - 1), \quad (11.36)$$

this is all for $\tau > 0$, but formula (11.27) gives $\gamma(0)$ and hence using (11.36) one can obtain all values of $\gamma(\tau)$. Formula (11.36) is a simple example of a *Yule-Walker* equation, more complex versions can be used to obtain formulae for the autocovariances of more general $\text{AR}(p)$ processes.

For the $\text{AR}(1)$ model we can divide through both sides of (11.36) by $\gamma(0)$ to obtain an equivalent expression for the autocorrelation function:

$$\rho(\tau) = \alpha\rho(\tau - 1). \quad (11.37)$$

Since $\rho(0) = 1$ always it is clear that $\rho(1) = \alpha$, $\rho(2) = \alpha^2$ and hence continuing in this manner that $\rho(k) = \alpha^k$. Thus for an $\text{AR}(1)$ process the parameter, α , can be discerned from the autocorrelation plot.

As an example we have simulated two $\text{AR}(1)$ processes each of 200 observations. The first `ar1pos` is a realization where $\alpha = 0.9$ and the second, `ar1neg` is a realization where $\alpha = -0.9$. Plots of each realization and their autocorrelation coefficients appear in Figure 11.2.

The exponential decay nature of the autocorrelation of `ar1pos` is clear and one can identify $\rho(1) = \alpha = 0.9$ from the plot. For `ar1neg` similar considerations apply except that $\rho(\tau)$ has an extra $(-1)^\tau$ factor which causes the oscillation.

ARMA models. Both AR and MA models express different kinds of stochastic dependence. AR processes encapsulate a Markov-like quality where the future depends on the past, whereas MA processes combine elements of randomness from the past using a moving window. An obvious step is to combine both types of behaviour into an $\text{ARMA}(p, q)$ model which is obtained by a simple concatenation. The process $\{X_t\}$ is $\text{ARMA}(p, q)$ if

$$X_t = \alpha_1 X_{t-1} + \cdots + \alpha_p X_{t-p} + \epsilon_t + \beta_1 \epsilon_{t-1} + \cdots + \beta_q \epsilon_{t-q}. \quad (11.38)$$

Fitting ARMA models So far, we have only described the probability models underlying ARMA processes and some mathematical quantities related to them. Of course, once one has real data one would like to know the answers to questions like “Are these models appropriate for my data?”, “Can I fit these models to my data?” “How do I fit these models?”, “Does

the model I fitted fit the data well?”. Indeed, these sorts of questions lead to the Box-Jenkins procedure, see Box *et al.* (1994). Part of this procedure involves studying the sample autocorrelation functions (and a related function called the *partial autocorrelation function*) to decide on the order of any MA (or AR) terms. However, there are many other aspects such as removing deterministic trends, removing outliers, and/or checking residuals. Indeed, a full treatment is way beyond the length constraints of this article so we refer the interested reader to Chatfield (2003) in the first instance.

11.1.4 The spectrum

So far we have been exploring time series with one hand tied behind our back! Statisticians, in particular, love the probability models part of time series analysis. This is probably because of the familiarity of the concepts like mean and covariance and the whole stochastic process paraphernalia. However, there is another, parallel, way of looking at time series. This “other way” determines how much energy is contained within a time series as a function of frequency (the *spectrum*). The “frequency” or spectral approach is, apparently, much more natural and familiar to those approaching time series analysis from a signal processing or engineering approach. However, it is important to realize that the two approaches are merely two sides to the same coin. Indeed, most of the important quantities in the spectral side are Fourier Transforms of equivalent quantities on the “statistics” side.

The basic principles of the spectral approach are very simple. For further information the author recommends any of the previously mentioned general texts but Priestley (1983) in particular. We start from the *harmonic process* model in Priestley (1983)

$$X_t = \sum_{i=1}^K A_i \cos(\omega_i t + \phi_i), \quad (11.39)$$

where K , $\{A_i\}$ and $\{\omega_i\}$ are constants, and the $\{\phi_i\}$ are independent random variables each having the uniform distribution on $[-\pi, \pi]$. We emphasize that (11.39) is but *one* model which has been found to be useful. There are many other possibilities and randomness could, in principle, be added to other quantities (for example, K , the number of frequencies could be made random for different parts of the series). The key points for us with this model is that X_t is comprised of sinusoidal waves where the wave with frequency ω_i has amplitude A_i . Here you have to imagine a time series being comprised of several sine waves of different frequencies (indexed by ω_i) each with different amplitudes. In general scientists are interested in analysing time series with model (11.39) in mind and figuring out what A_i and ω_i are.

The concept of building time series using oscillatory building blocks (sines) and varying amplitudes has been born in this paper. Essentially, the rest of this short paper deals with extensions and generalizations of this theme.

In general, there is no reason why only a finite number, K , of amplitudes need be involved. In fact it turns out that any discrete-time *stationary* process has a representation of the form

$$X_t = \int_{-\pi}^{\pi} A(\omega) \exp(i\omega t) d\xi(\omega), \quad (11.40)$$

where A is a possibly complex-valued function. Once again (11.40) is a probability *model* but this time involving an infinity of frequencies. For real *finite length* data, spectral information on a finite number of frequencies equispaced over the range usually gets estimated. This is in contrast to the estimation involved for (11.39) which is performed on a set of pre-specified frequencies. The reader should immediately compare this formula to the harmonic process model in (11.39) as the formulae are not all that different. The cosine in (11.39) gets replaced by $e^{i\omega t}$ but of course $e^{i\omega t} = \cos(\omega t) + i \sin(\omega t)$ so the sinusoidal nature is retained. The discrete A_i in (11.39) gets replaced by $A(\omega)$ in (11.40) because the sum gets replaced by an integral. The time series X_t in (11.40) is a stochastic quantity, so where does the randomness come in? The randomness in (11.40) gets supplied by the quantity $d\xi(\omega)$ which is technically known as an *orthonormal increments* process but *very loosely speaking* can be thought of multiplying the “sinusoid” $A(\omega^*)e^{it\omega^*}$ by an random infinitesimal quantity $d\xi(\omega^*)$ (a full discussion of which can be found in Priestley (1983)).

The above model is all very well but what is an analyst actually interested in? Given a time series X_t the problem is similar to that for the harmonic model except that only the quantity $A(\omega)$ is of interest (because all frequencies from $-\pi$ to π are automatically considered). If a particular range of frequencies $\omega \in \Omega$ does not contribute to the time series then we can always set $A(\omega) = 0$ for $\omega \in \Omega$). In practice one can actually estimate the quantity $f(\omega) = |A(\omega)|^2$ which is called the *spectrum* or *spectral density function*. It has a useful interpretation in that $\int_{\omega_1}^{\omega_2} f(\omega) d(\omega)$ quantifies the contribution to variance in the whole process from frequencies in the underlying process between ω_1 and ω_2 . Indeed the quantity $F(\omega) = \int_{-\pi}^{\omega} f(u) du$ is called the *spectral distribution function*. Both f, F are similar in character to the probability density and distribution functions.

Relationship between spectrum and autocovariance. Formula (11.40) can be viewed in a different light: the time series X_t can be seen as some sort of “Fourier transform” of $A(\omega)$. This is not quite the standard Fourier transform since the series X_t and the innovations process $\xi(\omega)$ are random.

However, this is enough to explain that the autocovariance and the spectrum are linked by the standard Fourier transform, i.e.

$$f(\omega) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \gamma(k)e^{-ik\omega} \text{ and } \gamma(k) = \int_{-\pi}^{\pi} f(\omega)e^{ik\omega} d\omega. \quad (11.41)$$

The Fourier link should come as no surprise as the covariance describes how much weight should be given to internal association in a time series at certain lags and the spectrum says how much weight is put on oscillations of certain frequencies. A series containing a dominant frequency at ω has clear implications for the autocovariance at lag $1/\omega$.

For example, Figure 11.2 showed realizations and autocorrelation estimates of two AR(1) time series. Figure 11.3 shows estimates of the associated spectra. Figure 11.3 (a) shows that most of the power is concentrated in the low frequencies and this corresponds to the plot Figure 11.2 (a) where the oscillations are largely confined to the lowest frequencies. In contrast, the spectrum Figure 11.3 (b) shows that most of the power is concentrated in the high frequencies and this corresponds to the rapidly oscillating series in Figure 11.2 (b). However, reexamination of the autocovariance plot in Figure 11.2 (c) is instructive. It shows that observations at lag one are highly negatively autocorrelated with the preceding observations, those separated by lag 2 are highly positively correlated, those at lag 3 negatively correlated and so on. This plot indicates fast oscillation which is precisely what the spectral plot shows. Similar comments can be made to compare the `ar1pos` autocorrelation and spectrum.

Estimation Although we have not yet said a lot about it, estimation is a critical component of time series analysis. In statistics it is no good just having interesting theoretical quantities if one cannot estimate them! As is typical in statistics there are usually several ways to estimate the same quantity and choice of estimate can sometimes be justified by mathematical statistical theory. However, choice of estimate sometimes depends on the situation one is faced with. For example, we have already covered one of the most popular estimates of the sample autocovariance in (11.13), but there are several others. There are also a variety of methods used for estimating the spectrum. The spectral estimate plots in Figure (11.3) were produced by assuming a parametric form for the spectrum and estimating parameters (since we knew that the underlying “true” spectrum was AR this seemed like a good thing to do). Other spectral methods involve taking the direct Fourier transform of the observations and studying the squared modulus of the resulting coefficients. This estimate is known as the periodogram. Unfortunately, it is a fact of life that with the periodogram if one obtains more data points (further observations in time) then one does not better estimate the spectrum at the existing frequencies. However, as more data become

available one can actually estimate more frequencies. In other words, the ratio of data available to estimate the spectral content at each frequency remains constant as more data become available (contrast this to estimation of, e.g. the sample mean of a sample which only gets better as more data are collected). Technically, the periodogram lacks a property called consistency. There are various techniques to create a consistent spectral estimate from the periodogram which usually involve smoothing the periodogram by pooling information from nearby frequencies. However, most of these techniques introduce bias into what is an asymptotically uncorrelated quantity. So, as with most smoothing methods there is a tradeoff: too much smoothing and the estimate of the spectrum becomes biased, too little and the estimate becomes too variable. In particular, the spectral estimate plots in Figure 11.3 are very smooth since we have used the parametric AR method for estimation. Typical periodogram based estimates are far noisier.

There are other important practical issues concerned with spectral estimation. The rate at which samples are taken for a time series puts an upper bound on the frequencies that can be estimated. If time series samples are taken at smaller and smaller intervals then higher frequencies can be observed. The upper bound is called the Nyquist rate and it is twice sampling rate (“twice” depends on the constant you assume in your definition of the Fourier transform). For example, sharp-eared humans can hear sounds up to a frequency of about 22kHz so the sampling frequency for a CD is about twice this: 44kHz. Suppose now you fix your sampling rate. Then if the time series you observe contains frequencies higher than the Nyquist rate then these higher frequencies cannot get sampled ‘often enough’ and so they actually appear in the sample as lower frequencies. In other words, the spectral information in the sampled series gets distorted. This phenomenon is called *aliasing* and basically means that you need to make sure your sampling rate is high enough to capture what you think will be the highest frequencies in the signal you are recording (it also is one reason why CD music played down the telephone sounds awful). Also, the converse of the aliasing problem is that to capture really long, very low frequency, cycles in a stationary time series you need enough time series observations to “see” the whole cycle.

Another effect is *spectral leakage*. Leakage occurs when there is a lack of periodicity in the time series over the whole interval of observation (this periodicity is “assumed” by the Fourier transform which is typically used to compute spectral estimates). Many real series are not periodic in that their first observation is almost never equal to their last! In other words, the behaviour at the start of the series is not the same as at the end which almost always occurs with real time series. However, this mismatch causes the Fourier series to “see” a large discontinuity which is an extremely high

frequency feature. This high frequency feature then gets aliased and spreads to low frequencies as noted in the previous paragraph. So, power seems to leak to a wider range of frequencies and a ‘leaked’ spectral estimate can seem more spread out and vague when compared to the ‘true’ spectrum. The effects of leakage can be minimized by using a technique called *tapering* which gently tries to taper or reduce the values at the start and end of a series to be the same value.

11.2 Non-stationary Time Series

The previous sections have concentrated on the very special class of *stationary* processes. There are many ways in which a time series fail to be stationary. We have already mentioned one. If the α parameter of an AR(1) process is equal to one then one obtains the random walk model in (11.28) described earlier. The process earlier was not stationary in that the variance was not finite (one can see that the sum in (11.26) will blow up to infinity if $|\alpha| \geq 1$).

11.2.1 Is it stationary?

As mentioned in the Introduction one should first test a time series to see if it is stationary. Formal hypothesis tests tend to concentrate on testing one kind of alternative but are often insensitive to other kinds (but, of course, they are often very powerful for the phenomena that they are designed to detect). For more information on these kinds of tests see Priestley (1983) or Van Bellegem (2003). As in many areas of statistics one can achieve quite a lot just through fairly simple plots. For example, one might look at a time series plot to see whether the mean or the variance of the time series changes over time. Another useful indication is to compute the autocovariance or spectrum (or both) on two different parts of the time series (that themselves “seem” stationary). If the two quantities from the different regions look very different then this provides some evidence of non-stationarity. Additional graphical procedures might be to look at some kind of time-frequency or time-scale plot (as later) and see if this exhibits constancy over time or not. If a test or a plot indicates non-stationarity in a particular way then that non-stationarity can be modelled in a number of ways as described next.

11.2.2 “Global non-stationarity”

A process is only stationary or not. However, the random walk example is, to the author’s mind at least, an example of a “global non-stationarity”.

One can imagine that the process and parameters of such a global non-stationarity process are fixed some time ago (usually infinity) and then the process itself evolves but the rule for evolution does not change. Another important example of such processes are the ARIMA processes which are a generalization of ARMA processes. As an example, suppose we difference the random walk process and give it another name:

$$W_t = X_t - X_{t-1} = \epsilon_t. \quad (11.42)$$

After differencing the process that is left is only ϵ_t . So the process W_t is merely a purely random process (which is stationary). The differencing operation is so important it has a notation $\Delta X_t = X_t - X_{t-1}$. There are also variants: Δ^2 means apply the differencing operation twice in succession $\Delta^2 X_t = \Delta(X_t - X_{t-1}) = X_t - 2X_{t-1} + X_{t-2}$ and Δ_k means difference with lag k , i.e. $\Delta_k X_t = X_t - X_{t-k}$.

An general ARIMA process is constructed in a similar way (the “I” stands for integrated since the process is a “summed” version of an ARMA process). First, the process X_t is differenced a number, say d , times: $W_t = \Delta^d X_t$, then a standard ARMA process is fitted to the result. The key fact here is that although a *stationary* model is fitted to the differenced data it is the case that the model for the original time series X_t is not stationary (like the above example where $W_t = \epsilon_t$ was stationary but $X_t = X_{t-1} + \epsilon_t$ is not). In estimating an ARIMA model one has to successively difference the data until one hits on a difference, d , that makes the time series “look” stationary. If an ARMA(p, q) process got fitted with d differences then the process is known as ARIMA(p, d, q) (the random walk is ARIMA(0, 1, 0)).

Sometimes it is not the time series itself that is the only quantity of interest. Recently, models have evolved that update other properties of the time series. For example, in econometrics and financial time series both stochastic volatility and ARCH (autoregressive conditional heteroscedasticity) models and their generalizations are very important (see, e.g. Tsay (2002)). For example, the ARCH model for an observed time series is:

$$X_t = \sigma_t \epsilon_t, \quad (11.43)$$

where ϵ_t is the purely random process. However, here the variance of X_t , σ_t^2 , is itself random! The variance is a time series in its own right and its value depends on the recent past history of the X_t process:

$$\sigma_t^2 = \alpha_0 + \sum_{j=1}^q \alpha_j X_{t-j}^2. \quad (11.44)$$

This kind of model can explain all kinds of phenomena common in financial time series (e.g. the autocorrelation of the series is insignificant, but the

autocorrelations of the absolute values of the series are not) and there are many more developments on this theme of providing a model for the variance too. Clearly, since the variance of this process, σ_t , changes through time the process is not stationary. However, from (11.44) the parameters α_i stay fixed for all time. This assumption is surely only approximately true for a great many real time series.

11.2.3 Local non-stationarity.

Clearly there is a need for time series models with parameters that themselves evolve over time. Although a fairly well-established idea (Page 1952; Silverman 1957; Priestley 1965) the concept has recently gained impetus due to the introduction of a new innovative asymptotic statistical modelling framework due to Dahlhaus (1997). See also Nason and von Sachs (1999) for further introduction and discussion.

The main issue for evolutionary processes is not so much dreaming up the process models themselves but how to estimate the parameters. As an example of the potential difficulties suppose one has a time series model $X_t = \sigma_t Z_t$ but this time suppose that σ_t changes in an unstructured way so that it is different for every time point. What hope is there in forming an estimate of σ_t from the single data point X_t ? The answer is “not much”. Estimating a quantity using a *single* sample value is not recommended. However, suppose that we postulate a model in which σ_t varies *slowly*. Then we can use values of X_t near to t^* to estimate σ_{t^*} .

More general processes than the one in the previous paragraph can be constructed. For example, the reason why the process defined in (11.40) is stationary is that the amplitude function, $A(\omega)$ does not depend on t : that is the statistical properties of the time series remain constant over all time because $A(\omega)$ is a constant function of time. A simple way to introduce time-dependence, causing the statistical properties of X_t to evolve over time is to replace $A(\omega)$ by $A_t(\omega)$. Without going into too many technical details the speed of evolution of the stochastic process X_t can be controlled by the mathematical rate of change of $A_t(\omega)$ as a function of t . The *locally stationary Fourier* processes of Dahlhaus (1997) is one possible method. However, there is no necessity to use Fourier building blocks. For example, the *locally stationary wavelet* processes of Nason *et al.* (2000) use an elaboration of the model:

$$X_t = \sum_j \sum_k w_{jk} \psi_{jk}(t) \xi_{jk}, \quad (11.45)$$

where $\psi_{jk}(t)$ are wavelets at scale j and location k , ξ_{jk} are simply identically distributed independent Gaussian random variables and w_{jk} specifies the amplitude at scale j and location k . Here the spectral quantity is called

the evolutionary wavelet spectrum (EWS), $S_j(k)$, which measures the power of oscillation in the time series operating at scale 2^j and location k . Approximately $S_j(k) \approx w_{jk}^2$. Many spectral methods just propose computing coefficients (whether they be Fourier or wavelet or other) and examining simple functions of these (usually the square) as spectral estimates. The advantage with models (11.40) and (11.45) is that we can compute statistical properties of our estimators (functions of coefficients), e.g. obtain expectations, variances and compute confidence intervals. A properly formulated model makes it clear whether certain kinds of phenomena can be actually estimated. For example, if your series evolves so fast in that it cannot be estimated with a slow sampling rate then it is important to be able to know this. Model based estimation provides the framework for this.

11.2.4 Example: Seismic Time Series

Figure 11.4 shows a time series of Realtime Seismic Amplitude Measurement (RSAM) counts. The counts are 10 minute averages, sampled every 5 minutes, of ground shaking caused by earthquakes and volcanic tremor. An estimate of the EWS is shown in Figure 11.5. For this time series the Nyquist frequency is one oscillation per 10 minutes. Hence the frequencies that can be observed in this series without aliasing range from 0 to a wavelength of 10 minutes. Figure 11.5 shows the contribution to overall variance in the time series at different scale (frequency) bands. The bands are labelled on the vertical axis and are arranged in a dyadic fashion. The finest scale (lowest band) contains the highest half of the frequencies, ranging from wavelengths from 20 minutes to 10 minutes (thus the mid period, labelled, is 15 minutes). Then the next finest scale illustrates the variance contribution for waves ranging from 40 minutes to 20 minutes wavelength (and hence mid period of 30 minutes), and so on. Of course, for a time series of 60 days one might ask how we can present information concerning waves at wavelengths of 85 and even 42 days? The answer is that we can not really and the information in these plots at the coarsest scales should not be taken too literally, especially away from the centre. This idea can be formalized and information in the plot above the dashed lines (either of them) should be disregarded. This idea of a “cone of influence” (below the lines) stems from Torrence and Compo (1998) who describe a different approach to wavelet spectral estimation.

A few interesting features can be picked out of the EWS estimate. A large peak exists in the time series at about 43 days (September 25th, 2001) and this gets picked out clearly in the EWS as a high frequency feature (indeed, it appears to be the dominant feature in the highest frequency bands at the bottom). Within the 8 hours to 2 day 16 hours bands there appears

to be oscillation at about 52 days that might reflect “banded tremor” and there are also early power peaks in the 2d16 and 5d8 bands at about 8 days that might indicate other geophysical processes. Of course, this is a preliminary analysis and much further work on interpretation and linking to real events on the ground would need to be done.

11.2.5 Example: Medical Time Series

Figure 11.6 shows an ECG (electro-cardiogram – heart-beats) recording of an infant child. Both ECG and the seismic series of the previous section measure the intensity of a number of events per unit time. The main difference with this medical time series compared with the seismic series is that we have at our disposal an extra time series: the infant sleep state. The infant sleep state is determined by an expert after the event using video information, EEG (electro-encephalogram – “brain-waves”) and EOG (electro-oculogram – eye movements) and indicates whether the infant is in quiet sleep, active sleep or awake.

A key goal of this data analysis was to determine whether there was any relationship between the time series of ECG and sleep state. Sleep state is a useful indicator of potential problems such as apnea but measuring it can be (i) expensive (ii) potentially distressing to the subject as EEG and EOG require attachment of sensors to the scalp and face (iii) is difficult to do in a home environment. On the other hand ECG is relatively unobtrusive, since the leads are attached to the infant’s chest and parents can be readily taught to do this, even in the home environment.

The full EWS estimate can be found in Nason *et al.* (2000). However, the estimate of power at the finest scale, $\hat{S}_{-1}(k)$ is shown in figure 11.7. Also shown in this figure is the expert determination of sleep state according to a coarse four-point scale. Clearly, periods of active and quiet sleep correspond to periods where the evolutionary spectrum is large. Clearly the wavelet spectrum of the ECG does have predictive power for the sleep state, see Nason *et al.* (2001) for further details.

11.3 Other topics

It is inevitable, in a short article such as this, that many important and interesting topics will not have been covered. By way of amelioration we mention a few of the ones that we consider to be more important. **Forecasting:** Given a time series it is often the case that the key objective is to *forecast* or *predict* future values of the series. For stationary series the principles for how to do this are well established, see Chatfield (2003)

for example. In general, for good forecasts one needs really good parameter estimates. For locally stationary processes this is doubly true, see Fryzlewicz *et al.* (2003). **Multivariate systems:** We have only studied situations concerned with a single time series. In practice, this is unrealistic as most experiments collect a multitude of data sets. Methods exist for the modelling and interpretation of vector-valued multivariate time series. For example, a basic quantity is the cross-correlation which measures the association between different components of a vector time series at different lags. **Discrete-valued time series** From brief conversations with various Earth Science colleagues it appears that many volcanological time series are discrete-valued (for example, counts of things, state of things). A lot of current time series work is for continuous-valued time series. However, this is changing, for example, see MacDonald and Zucchini (1997). (We also have not mentioned time series that are measured continuously in time. At least some of the standard texts, e.g. Priestley (1983) consider some models and examples).

11.4 Concluding Remarks

“All models are wrong. Some models are useful” (attributed to George Box). This is a useful thought to have in mind whilst modelling time series. This short article has attempted to introduce statistical time series modelling as a tool that can be used to model real-life time series, including those in volcanology. We studied various kinds of stationary time series model and said a word on how to estimate key quantities such as the autocorrelation and the spectrum. We have described a few practical details such as aliasing and leakage. In less detail we have described the basic principles of some non-stationary time series models and also shown some examples.

Free software exists to do many of the analyses described here (and in the text books). As a statistician I recommend the freeware R package which can be downloaded from <http://www.r-project.org/>. However, like the text books there are many other different packages that provide time series analysis to varying degrees.

Further reading

As mentioned in the introduction my favourite time series book is Chatfield (2003). Diggle (1990) is also a nice introduction. The book by Priestley (1983) approaches the field through oscillation and spectral analysis but also has a comprehensive coverage of the time-domain concepts and models. Priestley (1983) has extremely well written sections on the intuitive fundamentals of time series but also contains a lot of well-explained mathematical theory. The book by

Hamilton (1994) is pretty well comprehensive and up-to-date but might be too detailed for those seeking a quick introduction.

Acknowledgements

The author is grateful to Heidy Mader and Stuart Coles, co-organisers and helpers for organising the very interesting Workshop in Statistics and Volcanology. It must have been very hard work: thank-you. The author is grateful to Sofia Olhede of Imperial College for bringing the Thomson quote to his attention. The author would like to thank P. Fleming, A. Sawczenko and J. Young of the Institute of Child Health, Royal Hospital for Sick Children, Bristol for supplying the ECG data. The author would like to thank the Montserrat Volcano Observatory for the RSAM time series and he would like to thank Willy Aspinall for supplying this data and also for many helpful explanations concerning it. The author would like to thank the referees and editors for suggesting many important and helpful things. This work was supported by EPSRC Advanced Research Fellowship GR/A01664.

Bibliography

- [1] Box, G.E.P., Jenkins, G.M., & Reinsel, G.C. 1994. *Time-Series Analysis, Forecasting and Control*. Holden-Day, San Francisco.
- [2] Brockwell, P.J. & Davis, R.A. 1991. *Time Series: Theory and Methods*. Springer.
- [3] Chatfield, C. 2003. *The Analysis of Time Series: An Introduction*. Chapman & Hall/CRC.
- [4] Dahlhaus, R. 1997. Fitting time series models to nonstationary processes. *Annals of Statistics*, **25**, 1-37.
- [5] Diggle, P.J. 1990. *Time Series: a Biostatistical Introduction*. Clarendon Press.
- [6] Fryzlewicz, P., Van Bellegem, S. & von Sachs, R. 2003. Forecasting non-stationary time series by wavelet process modelling. *Annals of the Institute of Statistical Mathematics*, **55**, 737-764.
- [7] Hamilton, J.D. 1994. *Time series analysis*. Princeton University Press, Princeton, New Jersey.
- [8] Hannan, E.J. 1960. *Time series analysis*. Chapman and Hall.
- [9] MacDonald, I.L. and W. Zucchini 1997. *Hidden Markov and Other Models for Discrete-valued Time Series*. Chapman and Hall/CRC.
- [10] Nason, G.P. & von Sachs, R. 1999. Wavelets in time series analysis. *Philosophical Transactions: Mathematical, Physical and Engineering Sciences*, **357(1760)**, 2511-2526, doi:10.1098/rsta.1999.0445.
- [11] Nason, G.P., von Sachs, R. & Kroisandt, G. 2000. Wavelet processes and adaptive estimation of the evolutionary wavelet spectrum. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **62(2)**, 271-292, doi:10.1111/1467-9868.00231.

- [12] Nason, G.P., Sapatinas, T. & Sawczenko, A. 2001. Wavelet packet modelling of infant sleep state using heart rate data. *Sankhya*, **63**, 199-217.
- [13] Page, C.H. 1952. Instantaneous power spectra. *Journal of Applied Physics*, **23**, 103-106.
- [14] Pole, A., West, M. & Harrison, J 1994. **Applied Bayesian Forecasting and Time Series Analysis**. Chapman and Hall/CRC.
- [15] Priestley, M.B, 1965. Evolutionary spectra and non-stationary processes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **27**, 204-237.
- [16] Priestley, M.B. 1983. *Spectral Analysis and Time Series*. Academic Press.
- [17] Silverman, R.A. 1957. Locally stationary random processes. *IEEE Transactions on Information Theory*, **3(3)**, 182-187.
- [18] Thomson, D.J. 1994. Jackknifing multiple-window spectra. *In: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, **VI**, 73-76.
- [19] Torrence, C. & Compo, G.P. 1998. A practical guide to wavelet analysis. *Bulletin of the American Meteorological Society*, **79**, 61-78.
- [20] Tsay, R.S. 2002. *Analysis of Financial Time Series*. Wiley, New York.
- [21] Van Bellegem, S 2003. Adaptive methods for modelling, estimating and forecasting locally stationary processes. PhD thesis, Institut de Statistique, Université Catholique de Louvain, Voie du Roman Pays, 20, B-1348, Louvain-la-Neuve, Belgium.

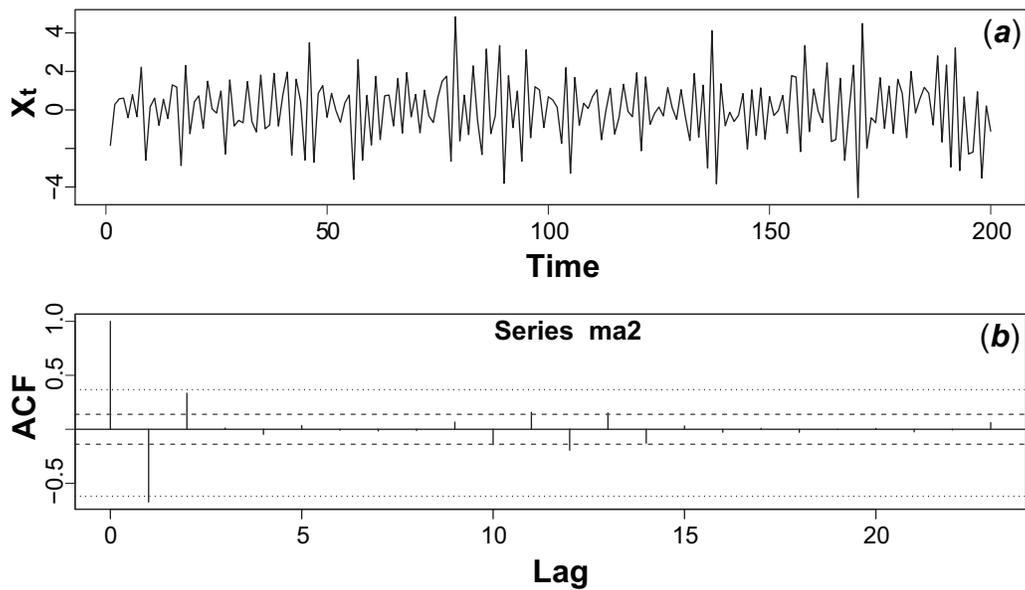


Figure 11.1: Plot (a) is a simulation realization of an 200 observations of an MA(2) process with parameters $\beta_0 = 1$, $\beta_1 = 0.9$, $\beta_2 = -0.8$. Here the horizontal axis is time, t , the vertical axis is the time series, X_t . Plot (b) is the sample autocorrelation of realization. Here the horizontal axis is lag, τ , the vertical axis is autocorrelation, $\rho(\tau)$.

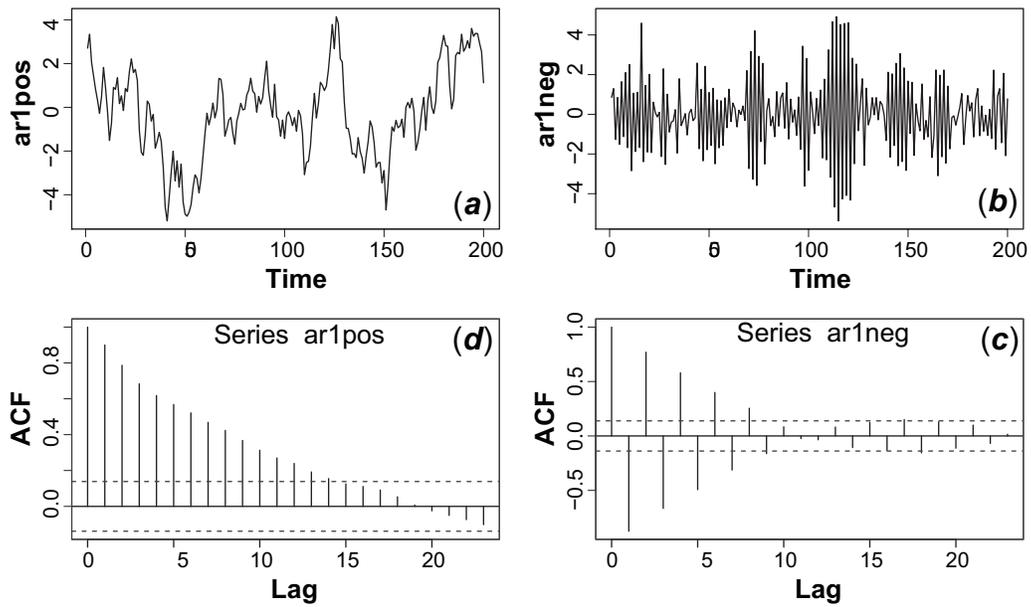


Figure 11.2: The plots are (a): realization of `ar1pos`, (b): realization of `ar1neg`, (c): autocorrelation of `ar1neg`, and (d): autocorrelation of `ar1pos`. Recall Time is t and lag is τ and ACF stands for autocorrelation function $\rho(\tau)$ in both cases.

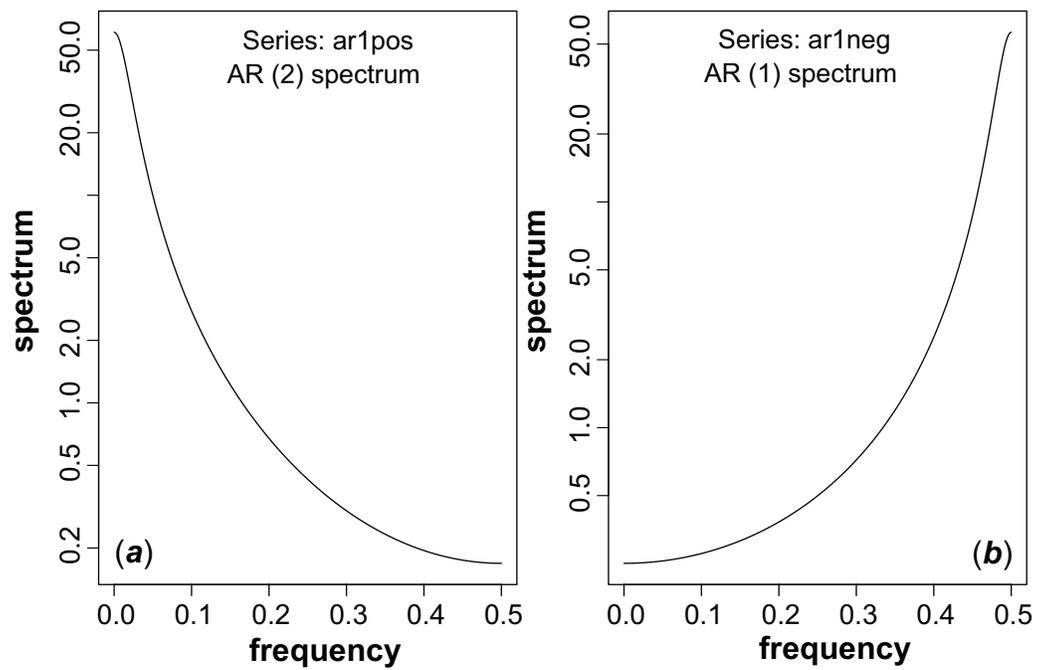


Figure 11.3: AR spectrum (estimates) of the time series presented in Figure 11.2. Spectrum estimate of (a) `ar1pos`, and (b) `ar1neg`. In both plots the horizontal axis, frequency, is ω and the vertical axis, spectrum, is $f(\omega)$.

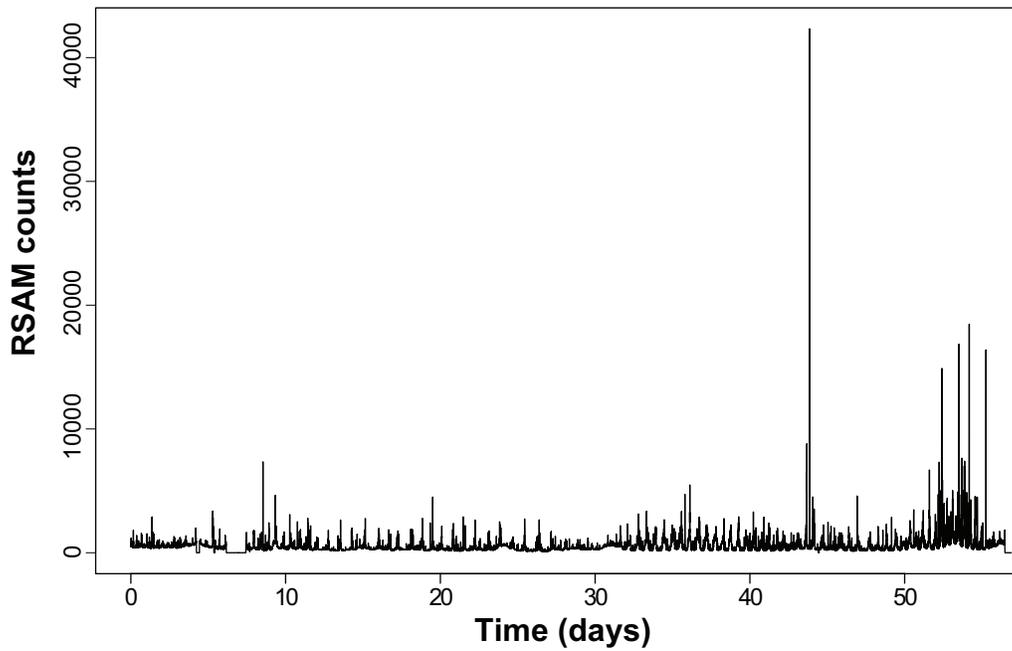


Figure 11.4: RSAM count series (from Montserrat from the short-period station MBLG, Long Ground, 16:43.50N, 62:09.74W, 287m altitude). Series starts 31th July 2001 00:00, ends 8th October 2001 23:54. Sampling time: every 5 minutes.

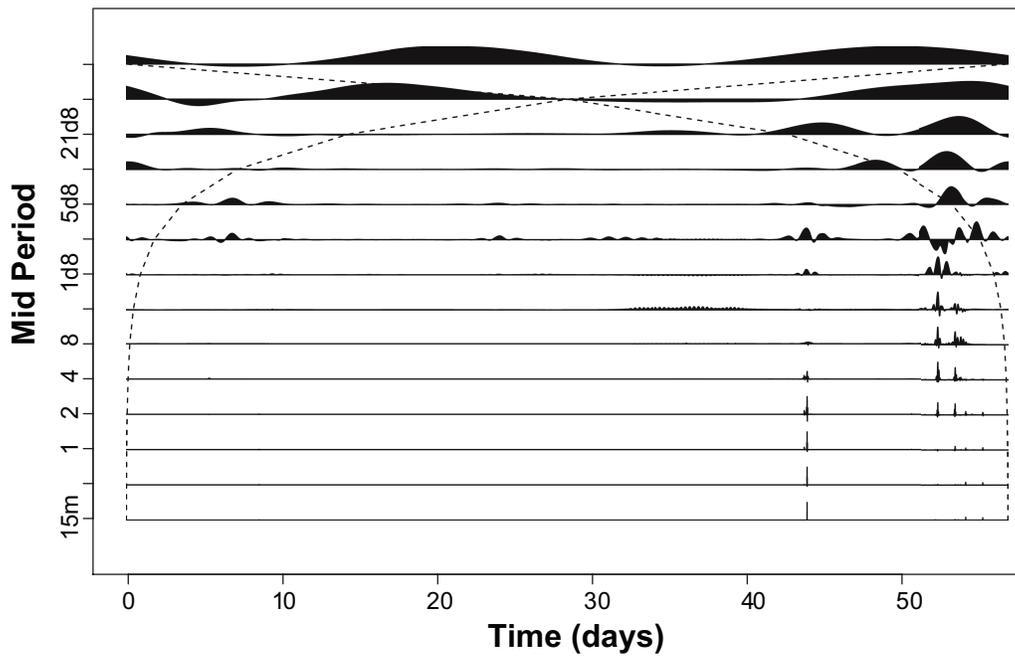


Figure 11.5: EWS estimate, $\hat{S}_j(k)$, for time series shown in Figure 11.4. Horizontal axis is time in days. Curved dashed lines indicate “cone of influence”. “Mid-period” is explained in the text.

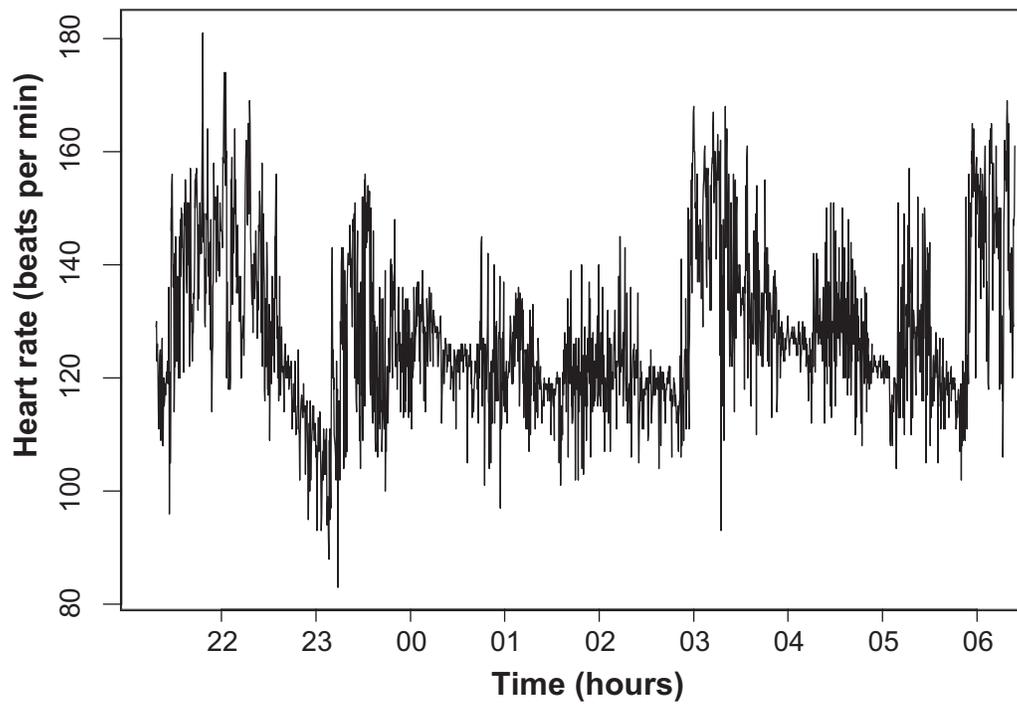


Figure 11.6: Heart rate recording of a 66 day old infant: the series is sampled at 1/16 Hz and is recorded from 21:17:59 to 06:27:18; there are $T = 2048$ observations. Reproduced from Nason *et al.* (2000) with permission.

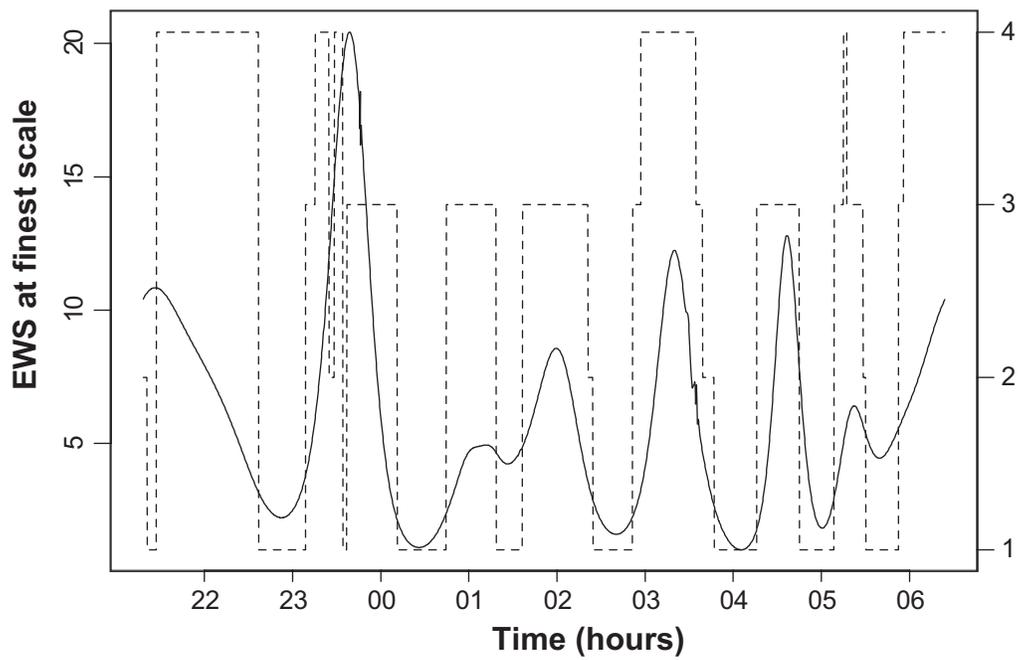


Figure 11.7: Estimate of the finest scale EWS for the ECG data in figure 11.6 (continuous curve, left axis) and sleep state (dashed line, right axis): 1, quiet sleep; 2, state between 1 and 3; 3, active sleep; 4 awake. Reproduced from Nason *et al.* (2000) with permission.