WAVELETS AND
TIME-DEPENDENT SPECTRAL ANALYSIS

by

M. B. Priestley

Technical Report No. 311
April 1995

Prepared Under the Auspices
of
National Science Foundation Grant
DMS 93-01366
Ingram Olkin, Project Director

Department of Statistics
Stanford University
Stanford, California
WAVELETS AND TIME-DEPENDENT SPECTRAL ANALYSIS

by

M. B. Priestley

Technical Report No. 311
April 1995

Prepared Under the Auspices
of
National Science Foundation Grant
DMS 93-01366
Ingram Olkin, Project Director

Department of Statistics
Stanford University
Stanford, California
WAVELETS AND TIME-DEPENDENT SPECTRAL ANALYSIS

by

M. B. Priestley

(University of Manchester Institute of Science and Technology)

1 Introduction

The recent explosion of interest in wavelet analysis has stimulated renewed studies of time-dependent spectral analysis. The basic aim of wavelet analysis is to represent a function as a linear superposition of “wavelets” centred on a sequence of time points, and in this sense it forms a natural tool for the investigation of “local” properties of time varying functions. The literature on wavelets is replete with terminology such as “discrete wavelet transforms”, “time-frequency localisation”, “windowed Fourier transforms”, all of which indicate that, at some level, there should be a close connection between wavelet analysis and time-dependent spectra. Indeed, phrases such as “local power spectra” are widely used in the wavelet literature, but so far the motivation for the use of this type of terminology has been highly intuitive and there has been little attempt to investigate the relationship between the two concepts in a formal and mathematically sound fashion. In pursuing this latter approach we encounter two basic difficulties: (a) almost all the literature on wavelet analysis deals with the representation of deterministic square-integrable functions, and (b) any form of time-dependent spectral analysis requires a very carefully constructed background theoretical model if the results of numerically calculated spectra are to have any meaningful physical interpretation. To put it quite simply, what meaning do sample spectra calculated from different time intervals possess? What functions are they estimating? In this paper we will try to address these questions and establish a more analytical link between wavelet analysis and time-dependent spectral analysis. We first review briefly the basic ideas underlying wavelet analysis.
2 Wavelet Analysis

The essence of wavelet analysis is to expand a given function $f(t)$ as a sum of "elementary" functions called wavelets. These wavelets are themselves derived from a single function $\psi(t)$, called the mother wavelet, by translations and dilations as explained below. The mother wavelet $\psi(t)$ is chosen so that it satisfies the following conditions,

\[
\int_{-\infty}^{\infty} \psi(t)dt = 0, \tag{2.1}
\]
\[
\int_{-\infty}^{\infty} |\psi(t)|dt < \infty, \tag{2.2}
\]
\[
\int_{-\infty}^{\infty} \frac{|\Psi(\omega)|^2}{|\omega|}d\omega < \infty, \tag{2.3}
\]

where $\Psi(\omega)$ is the Fourier transform of $\psi(t)$, i.e.

\[
\Psi(\omega) = \int_{-\infty}^{\infty} \psi(t)e^{-i\omega t}dt. \tag{2.4}
\]

Note that a necessary condition for (2.3) to hold is that $\Psi(0) = 0$ which is equivalent to (2.1). However, if we impose a slightly stronger condition than the integrability property (2.2) and assume that for some $\alpha > 0$,

\[
\int_{-\infty}^{\infty} (1 + |t|^\alpha)|\psi(t)|dt < \infty \tag{2.5}
\]

then we have

\[
|\Psi(\omega)| \leq C|\omega|^\beta, \quad \beta = \min(\alpha, 1), \tag{2.6}
\]

$C$ being a constant, and (2.3) is then satisfied. The condition (2.5) effectively ensures that $\psi(t)$ is "well localised", i.e. decays "quickly" to zero at a suitable rate. (Indeed, many mother wavelets have only finite support.)

![Wavelet Diagram](image)

Figure 1: (a) The Haar function. (b) The "Mexican hat function".

There are many different forms of $\psi(x)$ all of which satisfy the above conditions. The
classical mother wavelet is the Haar function (fig 1a) defined by,

\[ \psi^H(t) = \begin{cases} 
1, & 0 \leq t \leq \frac{1}{2}, \\
-1, & -\frac{1}{2} \leq t < 0, \\
0, & \text{otherwise.}
\end{cases} \]  
(2.7)

However, another commonly used wavelet is the “Mexican hat function” (fig 1b),

\[ \psi^M(t) = (1 - t^2)e^{-t^2/2}. \]  
(2.8)

Given a mother wavelet \( \psi(t) \) we now construct a doubly infinite sequence of wavelets by applying varying degrees of translations and dilations to \( \psi(t) \). Specifically, for all real \( a, b, (a \neq 0) \), write

\[ \psi_{a,b}(t) = |a|^{-\frac{1}{2}} \psi \left( \frac{t - b}{a} \right) \]  
(2.9)

so that \( a \) (usually restricted to positive values) represents the scale parameter, and \( b \) the translation parameter. (The normalising factor \( |a|^{-\frac{1}{2}} \) is chosen so that \( \int |\psi_{a,b}(t)|^2 dt = \int |\psi(t)|^2 dt \).)

The crucial property of the set of wavelets \( \{\psi_{a,b}(t)\} \) is that any (deterministic) function \( f(t) \in L_2 \) (the space of square integrable functions) can be expressed as a linear superposition of the \( \{\psi_{a,b}(t)\} \). Thus we can write (see, e.g. Daubechies (1992, p25)),

\[ f(t) = \frac{1}{C_\psi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \langle f, \psi_{a,b} \rangle \psi_{a,b}(t) \frac{dadb}{a^2}, \]  
(2.10)

where \( \langle f, \psi_{a,b} \rangle \) denotes the \( L_2 \) inner product, i.e.

\[ \langle f, \psi_{a,b} \rangle = \int_{-\infty}^{\infty} f(t)\psi_{a,b}(t) dt \]  
(2.11)

and

\[ C_\psi = \int_{-\infty}^{\infty} |\Psi(\omega)|^2 d\omega \]

The expression (2.11) is called the continuous wavelet transform of \( f(t) \). Remarkably, we can still obtain a representation of \( f(t) \) similar to (2.10) even when \( a \) and \( b \) are restricted to discrete sets of values. Thus, if we choose fixed values \( a_0, b_0 \), and set \( a = a_0^m, b = nb_0a_0^m \), \( n, m = 0, \pm 1, \pm 2, \ldots \) and set

\[ \psi_{m,n} = a_0^{-m/2}\psi \left( \frac{t - nb_0a_0^m}{a_0^m} \right) \]

then with suitably chosen \( \psi(t), a_0, b_0 \), we can still write any function \( f(t) \in L_2 \) in the form

\[ f(t) = \sum_m \sum_n \tilde{\psi}_{m,n}(t)\langle f, \psi_{m,n} \rangle \]  
(2.12)
where \( \{\tilde{\psi}_{m,n}(t)\} \) form the so-called "dual frame" wavelets (see Daubechies (1992, p54)), and
\[
\langle f, \psi_{m,n} \rangle = \int_{-\infty}^{\infty} f(t)\psi_{m,n}(t)dt
\]
is called the discrete wavelet transform.

The most commonly used discrete set of wavelets, associated with "multiresolution analysis," is generated by setting \( a_0 = \frac{1}{2}, b_0 = 1 \), yielding for \( m, n = 0, \pm 1, \pm 2, \ldots \)
\[
\psi_{m,n}(t) = 2^{m/2}\psi(2^m t - n) \tag{2.14}
\]
It may be shown that the class of wavelets given by (2.14) forms a complete orthonormal basis for \( L_2 \) (Mallat (1989), Daubechies (1992, p129)), and hence any \( f(t) \in L_2 \) admits the representation
\[
f(t) = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} a_{m,n}\psi_{m,n}(t) \tag{2.15}
\]
where
\[
a_{m,n} = \int_{-\infty}^{\infty} f(t)\psi_{m,n}(t)dt
\]
Note that in (2.14) the mother wavelet \( \psi \) is dilated by the factor \( 2^{-m} \) and translated to the position \( 2^{-m}n \).

In the context of multiresolution analysis the approximation of the function \( f(t) \) at resolution \( 2^{-\ell} \) is given by
\[
\hat{f}_\ell(t) = \sum_{m=-\infty}^{\ell} \sum_{n=-\infty}^{\infty} a_{m,n}\psi_{m,n}(t), \tag{2.17}
\]
and the inner sum in (2.17) is called the "detail signal at level \( 2^{-m} \). The approximation \( \hat{f}_\ell(t) \) can be written in the alternative form,
\[
\hat{f}_\ell(t) = \sum_{n=-\infty}^{\infty} b_{\ell,n}\phi_{\ell,n}(t) \tag{2.18}
\]
where \( \{\phi_{\ell,n}(t)\} \) constitute an orthonormal basis and are derived from the corresponding scale function \( \phi(t) \) by setting \( \phi_{\ell,n}(t) = 2^{\ell/2}\phi(2^\ell t - n), \ n = 0, \pm 1, \pm 2 \ldots \) and
\[
b_{\ell,n} = \int_{-\infty}^{\infty} f(t)\phi_{\ell,n}(t)dt
\]
(See, e.g., Mallat (1989)).
3 Applications of Wavelet Analysis

Most of the literature on wavelets is concerned with their applications to the analysis of deterministic functions, and one of the main objectives is simply to provide a characterisation of a given function $f(t)$ in terms of its wavelet transform. In particular, if we choose the discrete wavelet system (2.14) then $f(t)$ may be characterised by its set of wavelet coefficients $\{a_{m,n}\}$ (as given by (2.16)).

We now recall that the mother wavelet is required to be “well-localised” in the time domain, i.e. it decays “quickly” to zero. The wavelet $\psi_{m,n}(t)$ will then be “localised” around the time point $t = \frac{n}{2^m}$, and hence the wavelet coefficient $a_{m,n}$ will depend only on the “local” properties of $f(t)$ in the neighbourhood of $t = \frac{n}{2^m}$. Indeed, as $m$ increases the “width” of $\psi_{m,n}(t)$ decreases and the dependence of $a_{m,n}$ on the “local” properties of $f(t)$ becomes accentuated.

Consider, for example, the Haar wavelet $\psi^{H}(t)$ given by (2.7). Then

$$
\psi^{H}_{m,n}(t) = \begin{cases} 
2^{m/2}, & \frac{n}{2^m} \leq t \leq \frac{n+\frac{1}{2}}{2^m}, \\
-2^{m/2}, & \frac{n-\frac{1}{2}}{2^m} \leq t < \frac{n}{2^m}, \\
0, & \text{otherwise,}
\end{cases}
$$

(3.1)

so that $a_{m,n}$ depends only on the integral of $f(t)$ over an interval of width $\frac{1}{2^m}$ centred on $t = \frac{n}{2^m}$. Now if $f(t) \in L_2$ we would expect that, in general, $f(t) \to 0$ as $t \to \pm\infty$, and hence we would expect that the $a_{m,n}$ will become vanishingly small as $n \to \infty$. If we further truncate the range of $m$ at some finite point then hopefully we can obtain a sufficiently accurate characterisation of $f(t)$ in terms of a finite number of wavelet coefficients. As an extension of this idea we can use a wavelet representation also in connection with the problem of smoothing noisy data derived from observations on a function $f(t)$. This may be effected by “smoothing” the wavelet coefficients and then constructing the inverse discrete wavelet transform – see Donoho and Johnstone (1994), Nason and Silverman (1994).

4 Relationship with Fourier Analysis

There is an obvious analogy between wavelet analysis and Fourier analysis in the sense that both techniques aim to represent a function as a linear superposition of “basis” functions. In the case of wavelet analysis the basis functions are the wavelets $\{\psi_{a,b}(t)\}$ or $\{\psi_{m,n}(t)\}$
whereas in Fourier analysis they are the complex exponentials \( \{ e^{i\omega t} \} \). Thus, given a function \( f(t) \in L_1 \) we may represent it in the form

\[
f(t) = \int_{-\infty}^{\infty} F(\omega) e^{i\omega t} d\omega \tag{4.1}
\]

where \( F(\omega) \), the Fourier transform of \( f(t) \), is given by

\[
F(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} f(t) e^{-i\omega t} dt \tag{4.2}
\]

The analogue of the discrete wavelet transform arises when \( f(t) \) is periodic with period \( 2\pi \) (say), in which case we can represent \( f(t) \) as a Fourier series of the form,

\[
f(t) = \sum_{n=-\infty}^{\infty} c_n e^{int} \tag{4.3}
\]

where

\[
c_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(t) e^{-int} dt \tag{4.4}
\]

However, although this analogy is frequently referred to in the literature (see, e.g. Strang (1993)), it should be treated cautiously since despite the common feature mentioned above there are important differences between these two approaches. The most obvious difference is that the wavelet basis functions are indexed by two parameters \( (a, b \text{ or } m, n) \) whereas the Fourier basis functions are indexed by the single parameter \( \omega \). In physical terms this means that wavelet transforms (or coefficients) are characteristics of the local behaviour of function whereas Fourier transforms (or coefficients) are characteristics of the global behaviour of the function. In the case of Fourier analysis the parameter \( \omega \) has the physical interpretation of frequency; in wavelet analysis the second parameter \( (b \text{ or } n) \) represents a time location and the first parameter \( (a \text{ or } m) \) determines the "width" of the wavelet. There is thus no immediate physical connection between the wavelet parameters and the Fourier frequency parameter \( \omega \), but we may establish a tenuous relationship between \( \omega \) and the wavelet dilation parameter \( a \) (or \( m \)) by noting that if the mother wavelet has "oscillatory" characteristics — as, e.g., in the case of Mexican hat function (2.8) — then as \( a \) decreases the "oscillations" become compressed in the time domain, i.e. exhibit "high frequency" behaviour, whereas as \( a \) increases the "oscillations" become drawn out, i.e. exhibit "low frequency" behaviour. In a lose sense, therefore, we may tentatively identify the wavelet parameters \( a, b \) (or \( m, n \)) with "frequency" and "time", respectively, but it should be firmly noted that at this stage the physical interpretation of the wavelet parameters is purely heuristic. In particular it should be emphasised that the physical concept of "frequency" is related purely to the family of coupled exponential functions and has no precise meaning when
applied to other families of functions — unless, of course, we can define a more general concept of “frequency” which agrees with our physical understanding. (We return to this discussion in section 6.).

One notable feature of wavelet representations is that, as noted in section 3, the wavelet transforms (or coefficients) are “localised”, i.e. are time-varying and depend only on the local properties of \( f(t) \) in the neighbourhood of each time point. Thus, if \( f(t) \) has singularities (such as discontinuities or “spikes”) these will affect only the wavelet transforms at time points near the singularities; wavelet transforms at time points well removed from the singularities will be (essentially) unaffected. By contrast, the standard Fourier transforms (or coefficients) depend on the global properties of \( f(t) \), and any singularity in \( f(t) \) will affect all such transforms. As is well-known, if a periodic function \( f(t) \) has a discontinuity then we would require a very large number of terms in its Fourier series in order to obtain an adequate approximation of \( f(t) \) in the region of the discontinuity. However, if \( f(t) \) is “smooth” except at the discontinuity (say, piece-wise continuous) then we may obtain quite a good approximation by using a relatively small number of wavelet coefficients (see, e.g. Nason and Silverman (1994)).

One way in which we can strengthen the analogy between wavelet and Fourier analysis is by introducing the notion of time-dependent or windowed Fourier transforms, constructed as follows. Choose a function \( g(t) \) which is concentrated around \( t = 0 \) and decays to zero as \( t \rightarrow \pm \infty \) — rather like a mother wavelet but without the condition that it integrates to zero. We now define the windowed Fourier transform of \( f(t) \) by

\[
F_t(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} f(s)g(s-t)e^{-i\omega s} \, ds \quad (4.5)
\]

(What we have done, in effect, is to “isolate” a portion of the function \( f \) in the neighbourhood of the time point \( t \) by giving high weight to the values of \( f \) near \( t \) and low weight to values far removed from \( t \). The windowed Fourier transform \( F_t(\omega) \) now shares with wavelet transforms the property that it is time-dependent, i.e. is a function of two variables, \( t \) and \( \omega \), but it is, of course, dependent on the arbitrary choice of the “window” \( g(t) \) and does not therefore characterise purely the properties of the function \( f(t) \). The physical interpretation of \( F_t(\omega) \) is open to question; it is tempting (and indeed customary) to think of \( F_t(\omega) \) as a “short time” Fourier transform — particularly if \( g(t) \) is chosen to have a rectangular form and vanishes outside a finite interval. However, the question arises as to exactly what \( F_t(\omega) \) is measuring, and what type of representation for \( f(t) \) leads to a meaningful interpretation of \( F_t(\omega) \).
The use of windowed Fourier transforms raises a further and quite fundamental difficulty. If we wish to obtain "high resolution" in the time domain, i.e. we wish \( F_t(\omega) \) to reflect the properties of \( f \) strictly near the time point \( t \) (and not be contaminated by values of \( f \) far removed from the point \( t \)) then the window \( g(t) \) must be chosen so that it decays to zero very quickly, i.e. it must have a very small time domain "width". In this case the integral in (4.5) operates only over a very small time interval and \( F_t(\omega) \) then loses resolution in the frequency domain. Conversely, if we wish \( F_t(\omega) \) to have high frequency domain resolution then \( g(t) \) will have to be chosen so that it has a large "width" and we then lose resolution in the time domain. This feature is a consequence of a fundamental result known as the "Uncertainty Principle" (Priestley (1988, p151)), and applies equally well to wavelet transforms. The often stated objective that (mother) wavelets should be chosen so as to achieve "good localisation" in both the time and frequency domains is, in fact, impossible to achieve in an ideal sense.

Despite the fact that wavelet and Fourier transforms are both time and frequency dependent there is an important difference between the two techniques. In the case of windowed Fourier transforms the "width" of the window \( g(t) \) remains constant as the frequency variable \( \omega \) changes. Thus, both high and low frequency components are evaluated over the same (effective) time interval. However, in the case of wavelet transforms a decrease in the value of the parameter \( a \) (or an increase in the value of \( m \)) increases the frequency of the wavelet oscillations and simultaneously shrinks the time domain width of the wavelet. Similarly, an increase in the value of \( a \) decreases the frequency of the wavelet oscillations and stretches the time domain width — as shown in fig 2 for the Mexican hat wavelet. This

![Figure 2: The Mexican hat wavelet.](image)

is a crucial feature of wavelet analysis and means that in this approach high frequency components are evaluated over "short" time intervals whereas low frequency components are evaluated over "large" time intervals. This enables wavelet transforms to highlight "short time" high frequency phenomena such as short duration transients or singularities.
We could, of course, incorporate the same feature in windowed Fourier transforms by changing the width of the window \(g(t)\) as the frequency variable \(\omega\) changes - but the relationship between \(\omega\) and the width of \(g(t)\) would be quite arbitrary and would not be effected in the "automatic" way in which wavelets achieve this effect.

5 Spectral Analysis of Stationary Processes

Fourier transforms are a natural tool for the frequency analysis of (suitable) deterministic functions. However, if instead of deterministic functions we now consider a stochastic process, \(X(t)\), then in general Fourier transforms no longer exist and the appropriate tool for frequency analysis becomes Spectral Analysis. Thus a general stochastic process would not admit a standard Fourier integral representation of the form (4.1), but if we restrict attention to zero mean (stochastically continuous) stationary processes then there exists a generalised Fourier-Stieltjes integral representation of the form,

\[
X(t) = \int_{-\infty}^{\infty} e^{i\omega t} dZ(\omega)
\]

where \(Z(\omega)\) is a complex-valued process with orthogonal increments over \((-\infty, \infty)\) — see, e.g. Priestley (1981, p246). The integrated spectrum, \(H(\omega)\), is then defined by

\[
dH(\omega) = E[|dZ(\omega)|^2]
\]

When \(H(\omega)\) is differentiable, \(h(\omega) = H'(\omega)\) is called the (power) spectral density function. Since \(|dZ(\omega)|\) is the (random) amplitude of the component in \(X(t)\) with frequency \(\omega\), \(h(\omega)d\omega = E[|dZ(\omega)|^2]\) may be interpreted as the average contribution to the total power of \(X(t)\) with frequencies between \(\omega, \omega + d\omega\), — hence the description of \(h(\omega)\) as a power density function.

A basic result, known as Wiener-Khintchine theorem, allows us to evaluate \(h(\omega)\) (when it exists) as the standard Fourier transform of the autocovariance function of \(X(t)\), i.e. we have,

\[
h(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} R(\tau) e^{-i\omega \tau} d\tau
\]

where \(R(\tau) = E[X(t)X(t + \tau)]\) denotes the autocovariance function. Given a continuously observed finite record of \(X(t)\), say for \(0 \leq t \leq T\), we may construct an estimate of \(h(\omega)\) by passing the record through a narrow band filter centred on frequency \(\omega_0\), say, which (ideally) will remove all components except those with frequencies in the neighbourhood of \(\omega_0\). The power in \(X(t)\) near frequency \(\omega_0\) is then estimated by estimating the total
power (i.e. variance) of the filter output. More specifically, we choose a filter with impulse response function \( g(u) \) such that the corresponding transfer function

\[
\Gamma(\theta) = \int_{-\infty}^{\infty} g(u)e^{-iu\theta} du,
\]

(5.4)
is highly concentrated in the region \( \theta = 0 \). We then multiply \( X(t) \) by \( e^{j\omega_0 t} \) and pass it through this filter, giving the output

\[
Y(t) = \int_{t-T}^{t} g(u)X(t - u)e^{-j\omega_0(t-u)}du
\]

(5.5)
(Assuming that the impulse response function \( g(u) \) decays to zero sufficiently fast, and that \( t \gg 0 \), the limits in the integral in (5.5) may be effectively replaced by \(-\infty, \infty\)). We now estimate \( h(\omega_0) \) by evaluating the sample variance of \( Y(t) \), i.e. we form

\[
\hat{h}(\omega_0) = \frac{1}{T} \int_{0}^{T} |Y(t)|^2 dt
\]

(5.6)
The above estimate is, in fact, equivalent to computing the (continuous time) periodogram of \( X(t) \) and then smoothing this periodogram over neighbouring frequencies via the "spectral window" \( |\Gamma(\theta)|^2 \), i.e. if we evaluate the periodogram of \( X(t) \),

\[
I(\theta) = \frac{1}{2\pi T} \left| \int_{0}^{T} X(t)e^{-j\theta t} dt \right|^2
\]

(5.7)
then \( \hat{h}(\omega) \) may be written in the equivalent form

\[
\hat{h}(\omega_0) = \int_{-\infty}^{\infty} I(\theta)|\Gamma(\omega_0 - \theta)|^2 d\theta
\]

(5.8)
— see Priestley (1981, p496).

6 Time Dependent Spectral Analysis

The theory of spectral analysis described in the preceding section is valid only for stationary processes, i.e. processes whose second order statistical properties do not change over time. A non-stationary process does not possess a spectrum in the conventional sense, but since its statistical properties are now changing over time it is natural to try to describe its power/frequency properties by introducing the notion of "time dependent spectra", \( h_t(\omega) \), constructed so that, for each \( t \), \( h_t(\omega) \) describes a local power/frequency distribution which is characteristic of the behaviour of the process in the neighbourhood of the time point \( t \). From an empirical point of view the construction of time dependent spectral estimates is quite straightforward; given a sample record of \( X(t) \) observed over the interval \( 0 \leq t \leq T \) we
simply divide the interval \((0, T)\) into a number of sub-intervals and compute expressions of the form (5.6) separately for each sub-interval — leading to a sequence of functions \(\hat{h}_{t}(\omega)\), say, with \(t\) corresponding (say) to the mid-part of the sub-interval. (Note, however, that in view of the "Uncertainty Principle" — referred to briefly in section 4 — the smaller we make each sub-interval the more we lose resolution in the frequency domain). This approach is echoed in the wavelet analysis literature where the squared modulus of the wavelet transform (2.11), \(\langle \langle f, \psi_{a,b} \rangle \rangle^2\) is often referred to as a time-dependent "spectrogram" — see, e.g. Daubechies (1992, p86), Scargle (1994). However, at this stage the interpretation of any of these qualities as time dependent power spectra is merely wishful thinking. The crucial question is; what are (e.g.) the \(\hat{h}_{t}(\omega)\) estimating? To answer this we need first to define theoretical time dependent spectra in order to provide a framework against which we can interpret the sample estimates described above.

Various definitions of theoretical time dependent spectra have been proposed in the literature (see Priestley (1988, p855) for further discussion) but most of these fail to capture the physical interpretation of "local" power/frequency distributions. The crucial difficulty concerns the interpretation of "frequency": in the case of stationary processes we have the spectral representation (5.1), which tells us that a stationary process \(X(t)\) can be expressed as a "sum" of sine and cosine functions with varying frequencies and (random) amplitudes. We can then identify that component in \(X(t)\) which has frequency \(\omega\), and meaningfully discuss the contribution of this component to the total power of the process. In the absence of such a representation we cannot immediately talk about "power distributions over frequency". Now a non-stationary process cannot be represented as a "sum" of sine and cosine functions (with orthogonal coefficients) — instead we have to represent it as a "sum" of other kinds of functions. Since, according to its conventional definition, the term "frequency" is associated specifically with sine and cosine functions we cannot talk about the "frequency components" of a non-stationary process unless we first define a more general concept of "frequency" which agrees with our physical understanding.

Consider, for example, a deterministic function which has the form of a damped sine wave, say,

\[
f(t) = Ae^{-t^2/\alpha^2} \cos(\omega_0 t + \phi)
\]

If we perform a conventional Fourier analysis of \(f(t)\) we see that it contains Fourier components at all frequencies — the Fourier transform of \(f(t)\) consists of two Gaussian functions, one centred on \(\omega_0\) and the other on \((-\omega_0)\), the width of these functions being inversely proportional to the parameter \(\alpha\). Thus, if we represent \(f(t)\) as a "sum" of
sine and cosine functions with constant amplitudes we need to include components at all frequencies. However, we can equally well describe \( f(t) \) by saying that it consists of just two "frequency" components \( (\omega = \pm \omega_0) \), each component having a time varying amplitude, \( Ae^{-t^2/\alpha^2} \). In fact, if we were to examine the local behaviour of \( f(t) \) in the neighbourhood of a particular time point this is precisely what we would observe, i.e. if the interval of observations was small compared with \( \alpha \) then \( f(t) \) would appear simply as a cosine function with frequency \( \omega_0 \) and amplitude \( Ae^{-t^2/\alpha^2} \). Nevertheless, it would not be physically meaningful to attempt to assign a "frequency" to a function \( f(t) \) of arbitrary form — it would make little sense to talk about the "frequency" of the function \( \log \omega t \). For the term "frequency" to be physically meaningful the function \( f(t) \) must possess what we can describe loosely as an "oscillatory form", and we can characterise this property by saying that the Fourier transform of such a function will be concentrated around a particular point \( \omega_0 \) (or around \( \pm \omega_0 \) in the real case). Thus if we have a non-periodic function \( f(t) \) whose Fourier transform has an absolute maximum at the point \( \omega_0 \) we may define \( \omega_0 \) as the "frequency" of this function, the argument being that locally \( f(t) \) behaves like a sine wave with (conventional) frequency \( \omega_0 \), modulated by a "smoothly varying" amplitude function. It is this type of reasoning which forms the basis of the "evolutionary spectra" approach to the analysis of non-stationary processes, and this approach rests on a spectral representation which is a direct generalisation of (5.1).

7 Evolutionary Spectral Analysis

We consider a fairly general class of stochastic processes \( \{X(t)\} \) for which \( \exists \) a family of functions \( \{\phi_t(\omega)\} \) such that \( X(t) \) admits a representation of the form

\[
X(t) = \int_{-\infty}^{\infty} \phi_t(\omega) dZ(\omega)
\]

(7.1)

where \( Z(\omega) \) is again a process with orthogonal increments. Suppose that, for each fixed \( \omega \), \( \phi_t(\omega) \) (considered as a function of \( t \)) possesses a Fourier transform whose modulus has an absolute maximum at frequency \( \theta(\omega) \), say. Then \( \phi_t(\omega) \) may be regarded as an amplitude modulated sine wave with frequency \( \theta(\omega) \), i.e. we may write

\[
\phi_t(\omega) = A_t(\omega)e^{i\theta(\omega) t}
\]

where now the Fourier transform of \( A_t(\omega) \) has an absolute maximum at zero frequency. The function \( \phi_t(\omega) \) is then called an oscillatory function, and if the family \( \{\phi_t(\omega)\} \) is such that
\( \theta(\omega) \) is a single valued function of \( \omega \) then we may transform the variable in the integral in (7.1) from \( \omega \) to \( \theta(\omega) \), and by suitably redefining \( A_t(\omega) \) we may write

\[
X(t) = \int_{-\infty}^{\infty} A_t(\omega)e^{i\omega t}dZ(\omega) \tag{7.2}
\]

When \( X(t) \) admits a representation of the form (7.2) (with \( A_t(\omega) \)) satisfying the required condition we call it an oscillatory process. By virtue of the orthogonality of the \( \{dZ(\omega)\} \) it follows immediately from (7.2) that

\[
\text{var} \{X(t)\} = \int_{-\infty}^{\infty} |A_t(\omega)|^2d\mu(\omega) \tag{7.3}
\]

where \( d\mu(\omega) = E[|dZ(\omega)|^2] \). Since \( \text{var}\{X(t)\} \) may be interpreted as the "total power" of the process at time \( t \), (7.3) gives a decomposition of total power in which the contribution from frequency \( \omega \) is \( |A_t(\omega)|^2d\mu(\omega) \). This result is consistent with the interpretation of (7.2) as an expression for \( X(t) \) as the "sum" of sine and cosine terms with varying frequencies and time dependent (random) amplitudes \( \{A_t(\omega)dZ(\omega)\} \). We now define the evolutionary spectrum at time \( t \) by (cf Priestley (1988, p148)),

\[
dH_t(\omega) = |A_t(\omega)|^2d\mu(\omega) \tag{7.4}
\]

In particular, when \( H_t(\omega) \) is differentiable with respect to \( \omega \) (for each \( t \)), \( h_t(\omega) = dH_t(\omega)/d\omega \) is called the evolutionary spectral density function at time \( t \). Note that the evolutionary spectrum has the same physical interpretation as the conventional spectrum of a stationary process, namely that it describes a distribution of power over frequency, but whereas the latter is determined by the behaviour of the process over all time the former represents specifically the spectral content of the process in the neighbourhood of the time point \( t \).

Evolutionary spectral density functions may be estimated from data by an extension of the method described in section 5 for the estimation of spectral density functions of stationary processes. Thus, given observations over the interval \((0, T)\) we pass the data through a linear filter centred on frequency \( \omega_0 \), say, yielding an output \( U(t) \), say. We then compute a weighted average of \(|U(t)|^2\) in the neighbourhood of the time point \( t \) to provide an estimate of the local power density function at frequency \( \omega_0 \). Specifically, we set

\[
U(t) = \int_{t-T}^{t} g(u)X(t-u)e^{-i\omega_0(t-u)}du, \tag{7.5}
\]

\[
h_t(\omega_0) = \int_{t-T}^{t} w(v)|U(t-v)|^2dv \tag{7.6}
\]

where \( g(u) \) is a filter whose transfer function \( \Gamma(\theta) \) (defined as in (5.4)) is peaked in the neighbourhood of \( \theta = 0 \) and is normalised so that \( \int_{-\infty}^{\infty} |\Gamma(\theta)|^2d\theta = 1 \). The filter width,
\[ B_g = \int_{-\infty}^{\infty} |u| |g(u)| du, \] is chosen so that \( B_g \) is much smaller than "characteristic width" \( B_X \) of \( X(t) \), and the weight function \( w(v) \) is chosen so that \( \int_{-\infty}^{\infty} w(v) dv = 1 \). (The characteristic width, \( B_X \) is defined so that, roughly speaking, \( 2\pi B_X \) may be interpreted as the maximum time interval over which the process may be treated as "approximately stationary". For a precise definition of \( B_X \), together with a more detailed discussion of the estimation procedure see Priestley (1988, Ch 6)).

Assuming that both \( g(u) \) and \( w(v) \) decay to zero sufficiently fast so that the limits in the integrals in (7.5), (7.6) may be replaced effectively by \((−\infty, \infty)\), it may be shown that

\[
E[|U(t)|^2] \sim \int_{-\infty}^{\infty} h_t(\omega + \omega_0)|\Gamma(\omega)|^2 d\omega
\]  
(7.7)

so that

\[
E[\hat{h}_t(\omega_0)] \sim \int_{-\infty}^{\infty} \tilde{h}_t(\omega + \omega_0)|\Gamma(\omega)|^2 d\omega
\]  
(7.8)

where

\[
\tilde{h}_t(\omega + \omega_0) = \int_{-\infty}^{\infty} w(v) h_{t-v}(\omega + \omega_0) dv
\]  
(7.9)

Making the usual assumption that \( h_t(\omega) \) is "flat" over the bandwidth of \( |\Gamma(\omega)|^2 \) we may write

\[
E \left[ \hat{h}_t(\omega_0) \right] \sim \bar{\hat{h}}_t(\omega_0),
\]

ie \( \hat{h}_t(\omega_0) \) is an approximately unbiased estimate of the weighted average of \( h_t(\omega_0) \) in the neighbourhood of the time point \( t \).

For discrete parameter processes (i.e. processes defined for say integer values of \( t \)), the expressions corresponding to (7.5), (7.6) becomes

\[
U_t = \sum_u g_u X_{t-u} e^{-i\omega_0(t-u)},
\]  
(7.10)

\[
\hat{h}_t(\omega_0) = \sum_v w_v |U_{t-v}|^2,
\]  
(7.11)

where now \( \{g_u\}, \{w_v\} \) are suitably chosen sequences.

8 Relationship with Wavelet Analysis

Most of the literature on wavelets is concerned with the analysis of deterministic functions and there is very little discussion of their applications to stochastic processes — although it should be noted that Cambanis and Masry (1994) discuss the wavelet representation of both stationary and non-stationary processes but in the case of stationary processes restrict their analysis to stochastically continuous processes defined only on a finite time interval,
(0, T). (This condition is imposed so that, with probability 1, \(\int_0^T X^2(t)dt < \infty\) and thus \(X(t) \in L_2\), with probability 1.)

Let us now return to the discrete wavelet representation (2.15), viz,

\[
f(t) = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} a_{m,n} \psi_{m,n}(t),
\]

where

\[
\psi_{m,n}(t) = 2^{m/2} \psi(2^m t - n),
\]

\(\psi(t)\) being a given mother wavelet. As noted in section 4, it is conventional to identify the wavelet parameter \(m\) with "frequency", although in general this interpretation is highly dubious. However, if we focus attention on the particular mother wavelet,

\[
\psi(t) = \begin{cases} 
\sqrt{2} \sin 2\pi t, & 0 \leq t \leq 1, \\
0, & \text{otherwise},
\end{cases}
\]

then the "frequency" interpretation of \(m\) becomes much stronger. For with this class of \(\psi(t)\) we see that, for fixed \(m\), \(\sum_{n=-\infty}^{\infty} a_{m,n} \psi_{m,n}(t)\) has the form of an amplitude modulated sine wave with the modulating function taking the form of a step-function as shown in fig 3. (I am indebted to my colleague Dr P J Laycock for suggesting the form (8.3).) Thus, we may

![Figure 3: The form of \(\sum_{n=-\infty}^{\infty} a_{m,n} \psi_{m,n}(t)\).](image)

write

\[
\sum_{n=-\infty}^{\infty} a_{m,n} \psi_{m,n}(t) = A_t(\omega) \sin \omega t, \quad \omega = 2\pi 2^m,
\]

with

\[
A_t(2\pi 2^m) = 2^{m+1} a_{m,n}, \quad \frac{n}{2^m} \leq t < \frac{n+1}{2^m}
\]
Hence, we may now re-write the representation (8.1) in the form

\[ f(t) = \int_{-\infty}^{\infty} A_t(\omega) \sin \omega t dZ(\omega) \]  

(8.6)

with \( A_t(\omega) \) given by (8.5) and

\[ dZ(\omega) = \begin{cases} 
1, & \omega = 2\pi 2^m, \ m = 0, \pm 1, \pm 2, \ldots \\
0, & \text{otherwise}.
\]  

(8.7)

The similarity between (8.6) and the evolutionary spectral representation (7.2) is now quite compelling: in effect, (8.6) is a discrete approximation to (7.2), the discretisation occurring both over time (with \( A_t(\omega) \) taking the form of a step-function over the time intervals \((n/2^m, (n+1)/2^m)\)) and over frequency (with \( \omega \) restricted to values \(2\pi 2^m\)). Comparing (8.6) with (7.2) and using the definition (7.4) we now have the heuristic relationship,

\[ h_t(\omega) d\omega = |A_t(\omega)|^2 E[|dZ(\omega)|^2] \sim 2^{m+1}|a_{m,n}|^2, \ n = \frac{t}{2^m}, \ \omega = 2\pi 2^m, \]  

(8.8)

thus justifying the time-dependent spectral interpretation of the squared modulus of the wavelet coefficients. It is important, however, to recall that in order for the function \( A_t(\omega) \sin \omega t \) to be said to have “frequency” \( \omega \), the function \( A_t(\omega) \) must not oscillate “too fast”, i.e. its Fourier transform must satisfy the condition referred to in section 7. (Note that, with \( \psi(t) \) chosen as in (8.3), \( A_t(\omega) \) is constant over precisely one cycle of the function \( \sin \omega t \).)

Whether or not the Fourier transform of \( A_t(\omega) \) satisfies the required condition depends on the form of the wavelet coefficients, \( a_{m,n} \) which in turn depends on the function (or process) being analysed.

We may note further that in the representation (8.6) the frequency variable \( \omega \) is restricted to discrete values at the end points of octave bands. Thus, as \( m \) takes values through the range \(-\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots\), \( \omega \) takes values \(\ldots, \frac{\pi}{4}, \frac{\pi}{2}, \pi, 2\pi, 4\pi, 8\pi, 16\pi, \ldots\). Also, as is characteristic of this form of wavelet analysis, the time variable \( t \) is discretised at intervals which become progressively narrower as the frequency variable increases; for frequency \( \omega = 2\pi 2^m \) the time points are sampled at intervals of width \( \frac{1}{2^m} = \frac{2\pi}{\omega} \). Thus, high frequency components are sampled at small time intervals while low frequency components are sampled at large time intervals.

We now re-examine the heuristic relationship between \( h_t(\omega) \) and \( |a_{m,n}|^2 \) (equation 8.8) in a more detailed form. If we apply the above wavelet analysis to a zero mean stochastic process \( X(t) \) then replacing \( f(t) \) by \( X(t) \) in (2.16) and using (2.14) we may write

\[ a_{m,n} = \int_{-\infty}^{\infty} X(u) \psi_m \left( u - \frac{n}{2^m} \right) du \]  

(8.9)
where we have written \( \psi_m(t) = 2^{m/2} \psi(2^mt) \), \( \psi \) being the mother wavelet. The RHS of (8.9) has an obvious similarity with the process \( U(t) \) defined by (7.5). If, for the moment, we set \( \omega_0 = 0 \) in (7.5) and re-write the expression for \( U(t) \) in the form

\[
U(t) = \int_{-\infty}^{\infty} X(u)g(u-t)du
\]

then (8.9) becomes equivalent to (8.10) on identifying \( g(u) \) with \( \psi_m(u) \) and \( t \) with \( \frac{\omega}{2^m} \). We now have from (7.7),

\[
E[|a_{m,n}|^2] \sim \int_{-\infty}^{\infty} h_t(\omega)|\Psi_m(\omega)|^2 d\omega
\]

where \( \Psi_m(\omega) \) is the Fourier transform of \( \psi_m(t) \).

Now consider the complex form of the mother wavelet given by (8.3), namely \( \psi(t) = e^{2\pi it}, 0 \leq t \leq 1 \). Then

\[
\psi_m(t) = 2^{m/2} \exp\{2\pi i(2^mt)\}, \quad 0 \leq t \leq \frac{1}{2^m}
\]

and its Fourier transform is easily shown to be

\[
\Psi_m(\omega) = 2^{m/2} D_{\frac{1}{2^{m+1}}} (\omega - \omega_m) e^{-i\omega/2^m}
\]

where \( D_M(\omega) \), the Dirichlet kernel, is given by \( D_M(\omega) = \sin \frac{M\omega}{\omega} \), and \( \omega_m = 2\pi 2^m \) (see, e.g., Priestley (1981, p437)). Hence in this case \( |\Psi_m(\omega)|^2 \) is proportional to the square of a Dirichlet kernel centred on frequency \( \omega_m \), and thus \( |\Psi_m(\omega)|^2 \) is peaked in the neighbourhood of \( \omega = \omega_m \). In fact, \( |\Psi_m(\omega)|^2 \) has effectively the form of a Fejer kernel, \( F_M(\omega) \), centred on \( \omega = \omega_m \) and with \( M = \frac{1}{2^m} \). We may therefore conclude tentatively from (8.11) that if \( |\Psi_m(\omega)|^2 \) is suitably concentrated around \( \omega = \omega_m \),

\[
E[|a_{m,n}|^2] \sim 2\pi h_t(\omega_m), \quad t = \frac{n}{2^m}
\]

(Note that \( \int_{-\infty}^{\infty} |\Psi_m(\omega)|^2 d\omega = 2\pi \).) However, we have to examine the behaviour of \( |\Psi_m(\omega)|^2 \) for varying \( m \) very carefully. It is well-known that as \( M \to \infty \) the Fejer kernel becomes highly concentrated and acts as a pseudo \( \delta \)-function, while as \( M \) decreases towards zero the Fejer kernel becomes "flat". These features are evident once it is noted that the (half-power) bandwidth of the Fejer kernel is \( O(1/M) \). The band width of \( |\Psi_m(\omega)|^2 \) is correspondingly \( O(2^m) \), and consequently \( |\Psi_m(\omega)|^2 \) will be concentrated around \( \omega = \omega_m \) for small or negative \( m \) but for large positive \( m \) \( |\Psi_m(\omega)|^2 \) will be a "flat" function.

We may thus conclude that, as spectral estimates, the discrete wavelet transforms will have good frequency resolution at low frequencies but very poor frequency resolution at high frequencies. (This effect cannot be remedied by "smoothing" the \( |a_{n,m}|^2 \) over neighbouring values of \( n \).)
— as is done in the case of evolutionary spectral estimates — cf (7.6), (7.11). Such smoothing would simply reduce the sampling fluctuations but could not improve the resolution in the frequency domain.)

The behaviour of \(|\Psi_m(\omega)|^2\) with increasing values of \(m\) is a built-in feature of wavelet transforms and holds quite generally. Indeed, it is an inevitable consequence of the fact that as \(m\) increases (or equivalently as \(a\) decreases) the wavelet \(\Psi_{m,n}(t)\) "shrinks" in the time domain (cf the discussion in section 4) and correspondingly its Fourier transform becomes "flat". It is also perfectly consistent with the often quoted property of wavelets that they have the ability to "zoom in" on local "fine detail": high resolution in the time domain inevitably implies poor resolution in the frequency domain.

Whilst the discussion on the behaviour of \(|\Psi_m(\omega)|^2\) for large \(m\) is quite general, the behaviour of \(|\Psi_m(\omega)|^2\) for small \(m\) depends very much on the choice of the mother wavelet. The mother wavelet (8.3) has an "oscillatory" character, and as we have seen the Fourier transform of \(\psi_m(t)\), \(\Psi_m(\omega)\), is then centred on frequency \(\omega_m\) and thus provides some degree of frequency domain resolution. However, if we consider instead the Haar wavelet (2.7), or for simplicity the slightly modified version,

\[
\psi(t) = \begin{cases} 
1, & 0 \leq t \leq 1, \\
0, & \text{otherwise}
\end{cases} \tag{8.13}
\]

then the situation becomes quite different. (Strictly speaking, (8.13) is not a "wavelet" since it does not integrate to zero, but it is in fact the so-called "scale function" associated with the Haar wavelet.). It is now easily seen that

\[
|\Psi_m(\omega)|^2 = 2^{m+2}|D_{\frac{1}{2^{m+1}}} (\omega)|^2, \tag{8.14}
\]

so that, for all \(m\), \(|\Psi_m(\omega)|^2\) has a peak solely at the origin. We now have in place of (8.12),

\[
E[|a_{n,m}|^2] \sim 4 \pi h_t(0), \quad t = \frac{n}{2^m}, \tag{8.15}
\]

provided \(m\) is sufficiently small or negative for \(|\Psi_m(\omega)|^2\) to be sufficiently concentrated around \(\omega = 0\). For \(m\) large,

\[
E[|a_{n,m}|^2] \propto \int_{-\infty}^{\infty} h_t(\omega) d\omega = \text{total power at time} \quad t = \frac{n}{2^m} \tag{8.16}
\]

(Note that those results assume that \(X(t)\) is a zero mean process.). Thus in the case of the Haar wavelet the discrete wavelet transforms can at best estimate only the spectral ordinates at zero frequency and are totally ineffective at other frequencies.
9 Use of Complex Demodulation

The behaviour of wavelet transforms derived in the preceding section shows that although they may be useful for estimating time dependent spectra at low or zero frequencies they are ineffective at other frequencies - even in the case of an "oscillatory" wavelet of the form (8.3). To overcome this difficulty we may adopt the technique of "complex demodulation" (Priestley (1981, p848)). Thus, if we are interested in estimating time dependent spectra near frequency \( \omega_0 \), say, then we first transform \( \omega_0 \) down to zero frequency by setting

\[
X'(t) = e^{-i\omega_0 t}X(t)
\]

It follows immediately from the spectral representation (7.2) that

\[
X'(t) = \int_{-\infty}^{\infty} A_t(\omega + \omega_0)e^{i\omega}dZ(\omega + \omega_0),
\]

so that the evolutionary spectral density functions of \( X'(t) \) are given by

\[
h_t'(\omega) = |A_t(\omega + \omega_0)|^2 E \left[ \left| dZ(\omega + \omega_0) \right|^2 \right] = h_t(\omega + \omega_0)
\]

We can now use discrete wavelet transforms applied to \( X'(t) \) to estimate \( h_t(\omega) = h'(\omega - \omega_0) \) for frequencies \( \omega \) in the neighbourhood of \( \omega_0 \). In effect we now compute the wavelet transforms

\[
a_{m,n}' = \int_{-\infty}^{\infty} X(u)e^{-i\omega_0 u} \psi_m \left( u - \frac{n}{2m} \right) du
\]

and base the analysis on values of \( |a_{m,n}'|^2 \) evaluated from large negative values of \( m \). Thus to estimate \( h_t(\omega) \) we would form weighted averages of \( |a_{m,n}'|^2 \) over neighbouring values of \( n \), giving

\[
h_t(\omega) = \sum \omega_v |a_{m,n-v}'|^2, \quad t = \frac{n}{2m}
\]

— cf the evolutionary spectral estimate (7.11). Choice of \( m \)

Note that choosing \( m \) small or negative increases the spacings, \( \frac{1}{2m} \), between the successive time points at which the spectra are estimated. (Once again, increased resolution in the frequency domain leads to reduced resolution in the time domain.) The value of \( m \) must therefore be chosen so as to effect a compromise between high frequency domain and high time domain resolution. This approach may be particularly useful in applications to e.g. image analysis where we can "track" a particular frequency component over time and characterise its behaviour over time by a finite number of wavelet coefficients, \( \{a_{m,n}\} \). Moreover, we may exploit the existence of certain specially constructed wavelets designed to have efficient time and frequency localisation.
References


Further references in wavelets

