# Wavelets and Time Series Modeling

Guy-Michel Cloarec

▶ **To cite this version:**

Guy-Michel Cloarec. Wavelets and Time Series Modeling. [Research Report] Control Systems Group School of Electronic Engineering. Dublin City University. 1998. hal-03235820

**HAL Id: hal-03235820**
**https://hal.science/hal-03235820v1**

Submitted on 26 May 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Wavelets and Time Series Modeling

Guy-Michel Cloarec

Research Report EE/JVR/98/3

*Control Systems Group School of Electronic Engineering*
*Dublin City University*
*

June 1998

**Abstract**

The aim of this report is to present the most recent ideas related to time-frequency methods and particularily the so called *wavelet transform*. Time-frequency methods provides with analysis and preprocessing tools in the frequency space opening new perspective for time series modeling. These techniques are expected to provide better analysis tool for nonstationary time series or time series with pseudoperiodic or chaotic behaviour. Moreover, when used to perform preprocessing of the input space, these techniques can enable the use of multi-scale or multi-resolution modeling. Finally, we will examine how they can be put to contribution to address the problem of the complexity determination essential for topological determination and by mean of consequence the generalization issue.

Keywords : Wavelets, Time Series, Time-Frequency Analysis, Multi-resolution Modeling, Feature Extraction, Neural Networks, Principal Component Analysis.

*E-mail: cloarecg@eeng.dcu.ie

# Contents

# 1 Introduction

Time Series modeling can be divided into two main parts consisting of the time series analysis and the time series model optimization. In the first part, the analysis aim at characterizing the properties of the considered time series using various mathematical techniques. Again, these techniques can be distinguished into two broad classes of characterization, i.e. the time domain and frequency domain techniques [1].

Both classes involve preprocessing or transformation of the original time series with either expansion or extraction of some predominant features. In the time domain, the classical transformations are the detrending, deseasonalalization and the normalization of the time series. More complex methods lead to linear transformation and expansion such as *Singular Value Decomposition* (SVD). These transformations enable to perform the analysis but also provide with insights on the model choice.

In certain time series exhibiting periodic, pseudoperiodic or even chaotic behaviors, the frequency domain analysis and transformations provide further insights. As a matter of fact, the time domain methods usually fail to deal efficiently with such behavior and require the use of higher regression orders leading to highly correlated input spaces.

Spectral analysis and transformations are amongst the most commonly used frequency domain techniques. It is based on the concept according to which it is possible to decompose the time series into a sum of sines and cosines of different frequencies and amplitudes. Most of the spectral analysis is carried out in the frequency space. The frequency space is obtained by a transformation, usually the Fourier transformation [2].

The Fourier transformation prove to be efficient for stationary signals but miss critical information in non-stationary signals. For such signals, an efficient analysis and the modelling require to process the time and frequency domain information simultaneously. This lead to a unifying approach called *Time-Frequency Analysis* which in time evolves to *multi-resolution* analysis and modelling [3].

To overcome the shortcomings of the Fourier transformation, several methods have been developed. In a first part, we will look at the *modifications* made to the Fourier transform to perform a more reliable time-frequency analysis. In a second part, we will examine another approach based on *wavelets* with an emphasis on the similarities and differences with the Fourier approach.

# 2 The Fourier Approach

## 2.1 Spectral Analysis

The time domain representation of time series is usually the most natural description since almost all physical signals or economic variables are obtained by receivers recording variations in time. However, the frequency domain representation is a very powerful way to express the seasonality, the periodicity or pseudo-periodicity that occurs in physical events.

Spectral analysis is a classical frequency domain method. The variance of the time series is expanded (or broken down) into number of components, the spectral components. Each component is associated with a particular frequency. In other words, the time series is expanded into a series of weighted infinite sine and cosine signals of different frequencies. The relative amplitude of the harmonic represents the contribution of a given frequency (or spectrum) to the time series variance.

The space defined by the harmonics is the frequency space. Practically, the representation of the time series in the frequency domain is obtained by the Fourier Transformation. We start by recalling the definition of the Fourier Transformation.

$$X(\nu) = \int_{-\infty}^{+\infty} x(t) e^{-j2\pi\nu t} dt \tag{1}$$

Where

$x(t)$ is the time series (eventually linearly transformed).

$X(\nu)$ is the time series representation in the frequency domain.

For Equation 1, the spectrum $X(\nu)$ can be viewed as the coefficient function obtained by expanding the time series $x(t)$ into a family of infinite waves. A characteristic of the waves defined by $e^{-j2\pi\nu t}$ is that they are complex and completely unlocalized in time. For stationary time series, the spectrum is supposedly invariant in any subsets (large enough with respect to the corresponding period range) of the

time series. Under the assumption of stationarity, the measured variation in the spectrum can be used to evaluate the error of the spectrum estimate.

For non stationary time series, the Fourier transform indicate the frequencies contained in the time series, their amplitude (their relative contribution to the variance)and phase. However, no indication on their localization in time is given. This will be possible only with *frequency localization* techniques carried out in the time-frequency space [4].

## 2.2 Frequency Localization

For certain time series, it is necessary to characterize the signal simultaneously in the time and frequency domain. Such analysis is called frequency localization or, using its counterparts based on a statistical approach, frequency dispersion.

In order to express qualitatively the issues involved in frequency dispersion, we first have to define some notations. We will consider a probabilistic approach where the localization is related to the probability of a given spectrum to be present in the time series at a given moment.

We first consider the squared modules of the representations of the time series in the time domain and frequency domain respectively. $\|x(t)\|^2$ and $\|X(\nu)\|^2$ are two probability distributions. When considered in the two-dimensional time-frequency space, their product defines the *localization density of probability function*.

We define $E_x$ to be the energy of the signal (assumed finite).

$$E_x = \int_{-\infty}^{+\infty} \|x(t)\|^2 dt \tag{2}$$

From Equation 2, it is possible to define the averaged time and averaged frequency of the time series, $t_m$ and $\nu_m$ respectively.

$$t_m = \frac{1}{E_x} \int_{-\infty}^{+\infty} t\|x(t)\|^2 dt \tag{3}$$

$$\nu_m = \frac{1}{E_x} \int_{-\infty}^{+\infty} \nu\|X(\nu)\|^2 d\nu \tag{4}$$

In the $(t, \nu)$ time-frequency space (plane), $t_m$ and $\nu_m$ are the averaged values, i.e. the locus of main energy localization, or, in the statistical approach, the locus of maximum variation in the time series. From these averaged values, we can define the deviation of the spreading both in time and frequency.

$$T^2 = \frac{4\pi}{E_x} \int_{-\infty}^{+\infty} (t - t_m)^2 \|x(t)\|^2 dt \tag{5}$$

$$B^2 = \frac{4\pi}{E_x} \int_{-\infty}^{+\infty} (\nu - \nu_m)^2 \|X(\nu)\|^2 d\nu \tag{6}$$

In Equations 5 6, $T^2$ and $B^2$ are the time spreading and frequency spreading respectively. In the signal processing field, B is defined to be the bandwidth of the signal.

The product $T \times B$, called the *Time-Bandwidth* product, is an important qualitative characteristic of the time series.

A version of the *Heisenberg Principle of Uncertainty* can be applied here to bound the Time-Bandwidth product. It can been shown [2] that a time series (or a signal in general) cannot have simultaneously small support in time and frequency.

It is possible to qualitatively express that condition using the *Heisenberg-Gabor inequality* :

$$T \times B \geq 1 \tag{7}$$

The lower bound of the of the inequality is reached for a Gaussian function. Therefore, if the Time-bandwidth product is used as a complexity criterion, a Gaussian function will the reference function to which the complexity of the signal is compared.

## 2.3 Instantaneous Frequency

### 2.3.1 Introduction

We have seen in the previous section that the Fourier analysis provided with the full spectrum present in the entire signal or data set. In order to localize in time the frequency components in nonstationary time series, we introduce the concept of *instantaneous frequency*. This concept is based on the assumption that there exists only one spectra or harmonic for a given time index. This unique spectrum is called the *principal component* and the instantaneous frequency techniques are referred as *mono component* analysis techniques [3].

### 2.3.2 Analytic Signal

It is well known that the Fourier transform provides symmetrical spectra or double-sided spectra. The spectra is symmetrical since the time series or the signal are usually real. To define the instantaneous amplitude and frequency functions, we first have to introduce the *analytical signal* or time series of $x(t)$.

$$x_a(t) = x(t) + j\mathcal{H}_t(x(t)) \tag{8}$$

Where

$\mathcal{H}_t$ is the Hilbert transform.

$x_a$ is the analytic signal or time series associated with x.

The main property of $x_a$ is that in the frequency domain $X_a$ is single-sided. Basically, the negative frequencies are removed while the amplitude of the positive frequencies are doubled to ensure that the DC component is unchanged. This can be summarized by :

$$\begin{aligned} X_a(\nu) &= X(0) \quad if\ \nu = 0 \\ X_a(\nu) &= 2X(\nu) \quad if\ \nu > 0 \end{aligned}$$

### 2.3.3 Mono-Component Analysis

Since $x(t)$ is a real signal, $X_a$ contains all the information contained in X. Using the analytic signal, it is possible to define real values instantaneous amplitude and frequency distribution.

$$a(t) = \|x_a(t)\| \tag{9}$$

$$f(t) = \frac{1}{2\pi}\frac{\partial arg(x_a(t))}{\partial t} \tag{10}$$

Equation 9 10 represents the instantaneous amplitude and frequency function used to characterize locally the frequency behaviour as a function of the time. To illustrate the concept of instantaneous frequency, we can examine the application to a chaotic time series [5] that exhibit particular pseudoperiod features.

Figure 1 represents the 2000 samples of the chaotic time series. An interesting feature is the succession of oscillatory modes with increasing amplitude that occur locally. In order to analyze in more detail these feature, we use the instantaneous frequency function on the analytic time series.

The analytic time series associated with the chaotic time series is represented on Figure 2. It appears clearly that the frequency in varying in time but the instantaneous frequency function does not provide an accurate and qualitative reading of the modulation. This can be a typical effect of simultaneous harmonic as well as noise (large frequency range).

It is clear that for noise free stationary signals, the instantaneous frequency function give good results but for real life application, better methods must be used. The first step to overcome the effects of noise (white noise) and the single component limiting assumption would be to implement multi component analysis techniques.

## 2.4 Multi-Component Transformations

### 2.4.1 Atomic Decomposition

As we already explained, the Fourier transform is not well adapted to non-stationary signals principally because it is based on the expansion in infinite waves (sinusoids) which are delocalized in time. In the

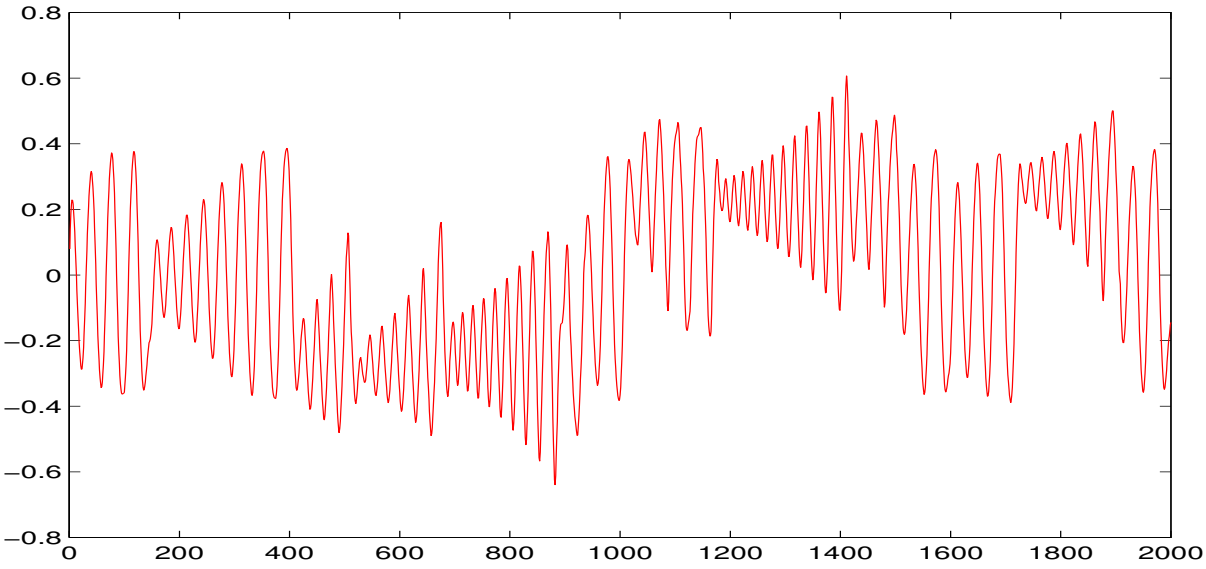


Figure 1: Chaotic Time Series.

section 2.3.3, we have seen that the primary attempt to localize the spectrum using the instantaneous frequency principal lead to lose valuable information. To combine the multi-spectra or multi-resolution capacities of Fourier analysis and the need for localization of the frequency, we introduce another concept called *Atomic Decomposition* also known as *Linear Time-frequency Representation* [3, 6].

The main difference with the instantaneous amplitude and frequency functions is the fact that the new function issued from the atomic decomposition will be bi-dimensional. The bidimensionality implies that for a given time index, the full spectrum (or at least a portion of it) with be available. Atomic relates to the fact that the signal or the time series will be somehow *truncated* into elementary time intervals on which the spectrum will be evaluated. Of course, behind atomic decomposition lies more mathematically sound principals and definitions but we will limit the spectrum of this report on practical applications. Therefore, we will introduce the transforms based on atomic decomposition, the *Short-term Fourier Transform*, also called the *Windowed Fourier Transform*.

### 2.4.2 Short-Time Fourier Transform

The main idea behind Short-Time Fourier Transform (STFT) (also called Gabor transform) is to introduce a time dependency in the Fourier transform. The term dependency relates to a more general approach of the identification and modeling issues based on statistical theory.

The simpler solution consists to pre-windowing the signal around a particular point in time and calculating the Fourier transform over that window. The STFT is obtained by translating the window over the full signal or time series length. The resulting bidimensional function is called STFT or *Short-Time Spectrum*.

$$F_x(t,\nu,h) = \int_{-\infty}^{+\infty} x(u)h^*(u-t)e^{-j2\pi\nu u}du \tag{11}$$

where

$x(u)$ is the time series.

$h(t)$ is the short-time analysis window.

$F_x(t,\nu,h)$ is the short-time Fourier transform of x.

Equation 11 represents the definition of the STFT function. In other words, it is a *local* spectrum of x around time t. This transform can be inverted by :

$$x(t) = \frac{1}{E_h}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} F_x(u,\xi,h)h(t-u)e^{j2\pi t\xi}dud\xi \tag{12}$$

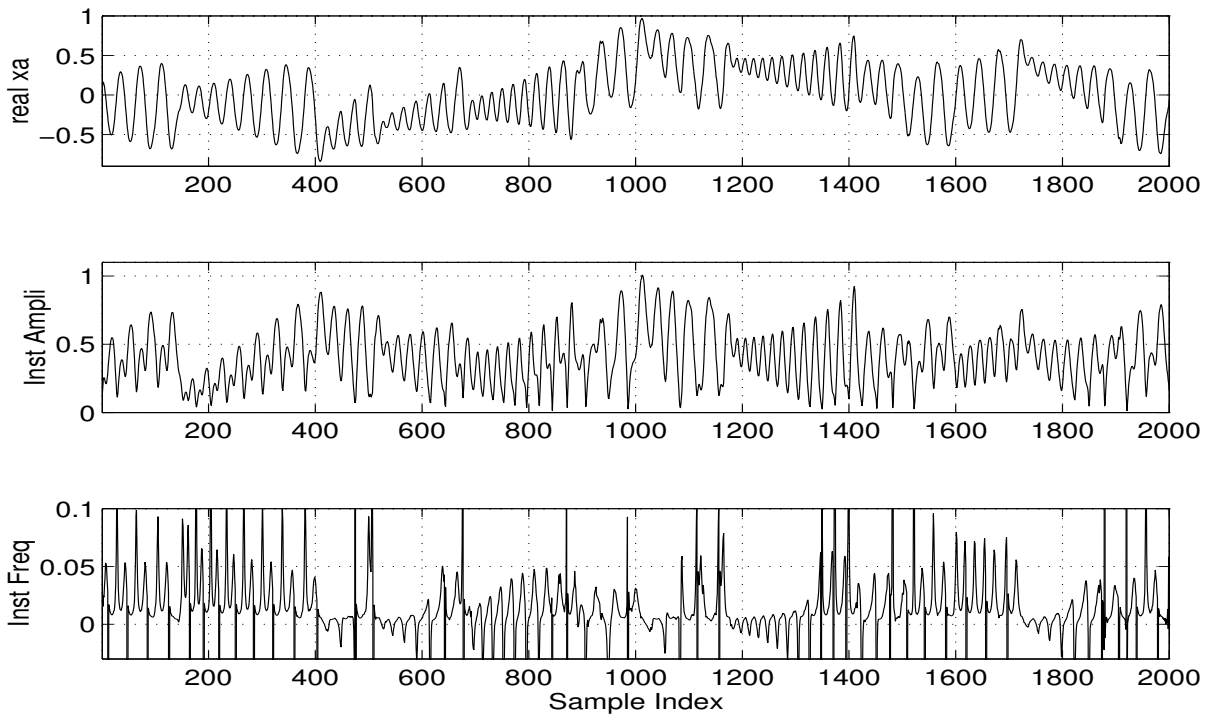


Figure 2: Analytic Time Series and Instantaneous Amplitude and Frequency.

Where $E_h$ is the term related to the short-time analysis window h and used to normalize the relation.

$$E_h = \int_{-\infty}^{+\infty} \|h(t)\|^2 dt \tag{13}$$

Equation 12 can be seen as the expansion equation : the signal is expanded into a sum of weighted elementary waveforms. The elementary waveforms are obtained form a primary or elementary function, here the window function h. This interpretation, although it seems to complexify what is a simple Fourier transform variation, is important to understand how another transform, the *wavelet transform*, is defined.

In order to detail the elementary waveform approach, we introduce the concept of *atoms* : An atom is an elementary waveform obtained by translation in time and frequency (modulation) of a primary (mother) function. In the SFTF context, the mother function is the window function h. Therefore, the elementary building block or atoms as defined by :

$$h_{t,\nu}(u) = h(u-t)e^{j2\pi\nu u} \tag{14}$$

The corresponding transformation group of $(t, \nu)$ translation is called the *Weyl-Heisenberg* group. Being aware of the fact that the results of such transformation are not obvious, we can examine an example. A simple window function is the Gaussian window centered in the origin and with a variance of 1. From that Gaussian we will defines two atoms or *atomic functions*, obtained from restriction of the $h_{t,\nu}$ function. Practically, we chose to fix the time index, i.e. location to 5 and we set the *frequencies* (or what will be called later *scales*) to respectively 1 and 3. The window function and the two atoms respectively called *Atom 1* and *Atom 2* are represented on Figure 3.

Figure 3 represents two atomic functions obtained from a specific Gaussian window. The two atoms correspond to the point in time $t = 5$ and for frequencies $\nu = 1$ and $\nu = 3$ respectively. Of course, the atomic functions are complex so that so simplicity, the real part (line) and the imaginary part (dash line) have been represented. Broadly speaking, the atoms can be seen as attenuated sinusoids of frequency 1Hz and 3 Hz respectively with symmetrical envelopes centered on $t = 5$ and variances equal to the variance of the initial window, the mother function.

The time series or the signal is then expanded as the summation of the weighted atoms. These weights are correlated with the presence of the given spectrum within the time series. Of course, when we use the term frequency, it will have some minor differences with the *Fourier frequency* as we will see later.

Before examining an example, it is interesting to examine yet another understanding of the atomic decomposition. If we define the STFT function in the Fourier space, we have :

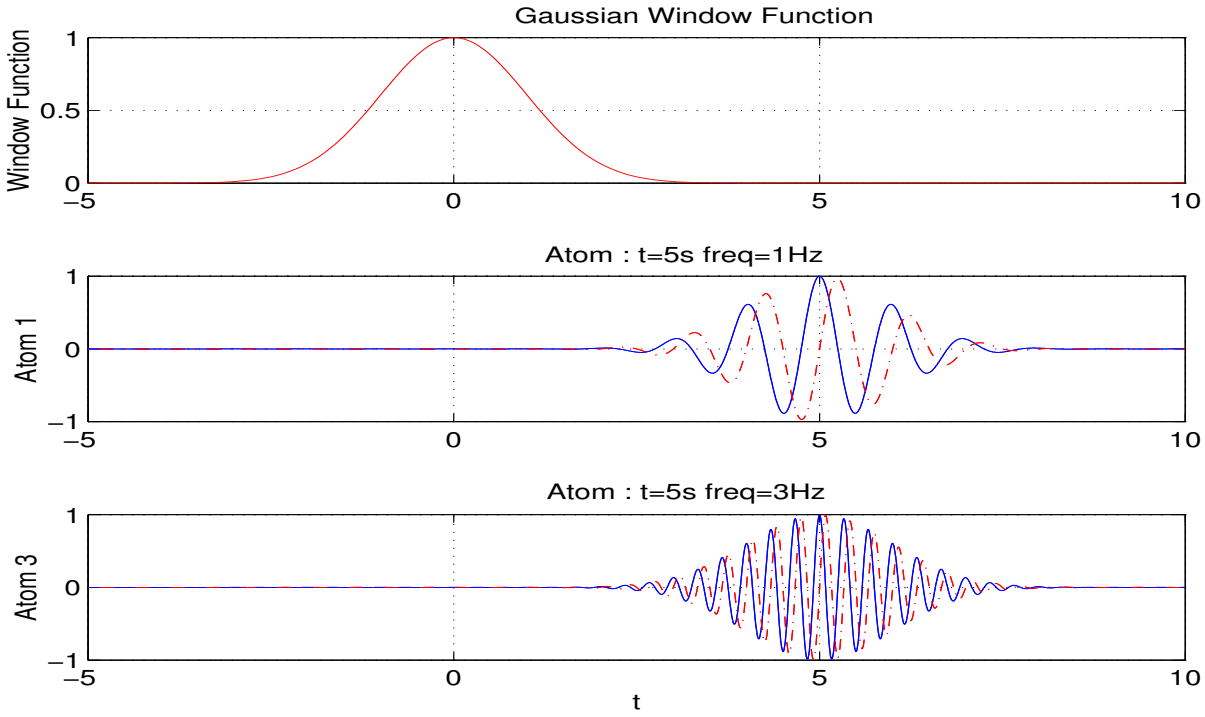


Figure 3: Atomic Decomposition.

$$F_x(t,\nu,h) = \int_{-\infty}^{+\infty} X(\xi)H^*(\xi-\nu)e^{j2\pi(\xi-\nu)t}d\xi \tag{15}$$

Where X and H are the Fourier transforms of x and h.

Equation 15 can be interpreted as the result of a band-pass filter whose frequency response is given by $H(\xi-\nu)$. As we have seen in Figure 3 where the atoms had a constant envelope variance (in the time domain), similarly in the frequency domain, the band-pass filters have constant bandwidth. Just like in the time domain, the bandwidth is determined by the *mother filter* $H(\xi)$. For that reason, in the signal processing fields, STFT is assimilated as a bank of band-pass filter with constant bandwidth.

Before discussing the issues of time-frequency resolution, we first examine an application on the competition time series.

Figure 4 represents the short-time-frequency tranform over the first 1024 samples of the competition time series. It seems that the energy density is maximum for frequencies in around 0.026 Hz, i.e. a period of 38 samples. Between $t = 400$ and $t = 600$, another spectrum corresponding to very low frequency, i.e. large periods, seems to become predominant while components such as 0.035 Hz coexists.

Before interpreting further the results, it is important to consider the issues of time-frequency resolution. In other words, it is clear that the use of windows truncate the information so that different spectra are not estimated with the same accuracy.

### 2.4.3 Time-Frequency Resolution

The use of windows to obtain localized Time-Frequency multicomponent at first a good approximation. The windowing can use basically two types of windows function : a box function where all the signal within the interval is considered evenly or *smoothing functions* such as Gaussian functions. With smoothing function, the calculation of the spectrum will emphasize the information contained in the signal inverse proportionally to the distance from the actual time index. In that case, the window function can be seen as a density probability function, the density of probability for a given sample to be used in the Time-Frequency evaluation [3].

In both case, the multi-frequency component are obtained on a restriction of the input signal whether the boundaries of the interval are crisp or fuzzy. This interval will be constant for all frequencies. Therefore, it is quite clear that frequencies associated with periods smaller that the restriction interval will be estimated with much better accuracy than the one associated with larger periods.

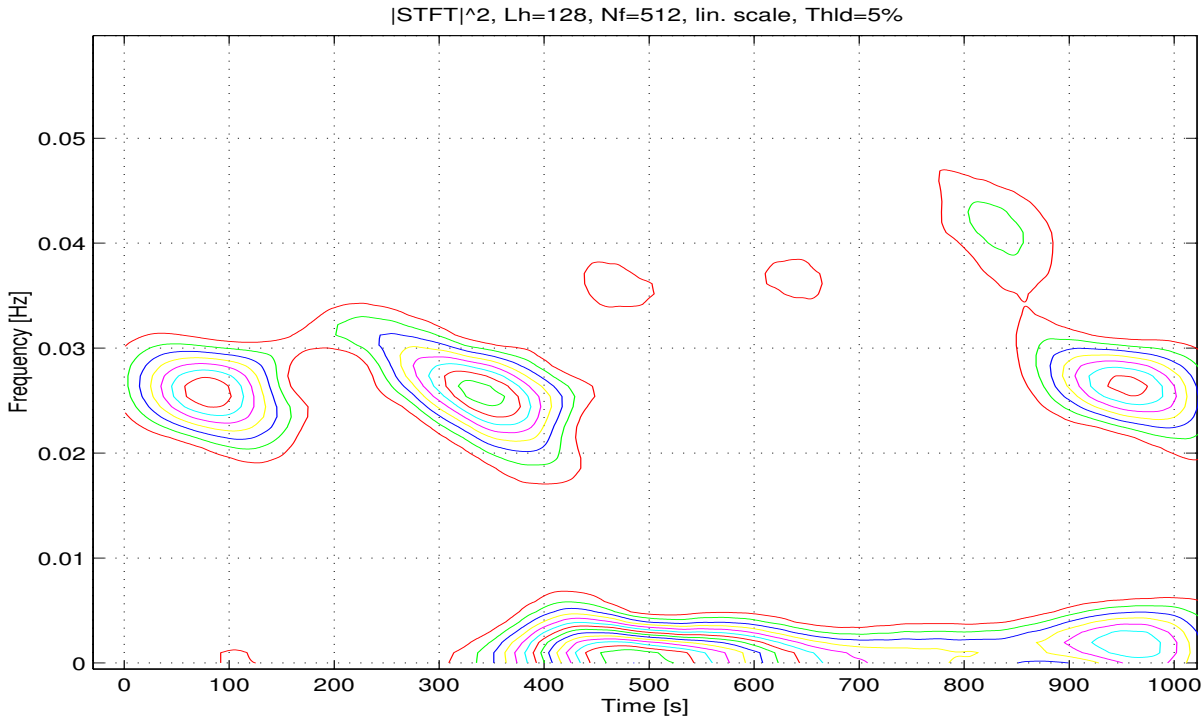


Figure 4: STFT of the Competition Time Series.

In order to quantify that issue, we first define these intervals. In the box function case, the interval is simply defined by the width of the box function. In the Gaussian function case, we will use an interval that will provides with a cumulated probability of membership of 95 %, i.e. an interval of width $2\sigma$ where $\sigma$ is the standard deviation parameter associated with the Gaussian function. As stated before, these intervals will be fixed for the entire frequency spectrum and will be called T.

To quantify the time resolution, we examine a simple case, the Dirac impulse. Let the signal x be a Dirac delta function also called impulse function.

$$x(t) = \delta(t - t_0) \tag{16}$$

One of the properties of the Dirac delta function is that its Fourier transform is the unity.

$$\int_{-\infty}^{+\infty} \delta(u - t_0) e^{-j2\pi\nu u} du = 1 \tag{17}$$

The impulse signal has a flat spectrum, i.e. all frequencies are present at unit strength. When can then examine the short-time Fourier transform of the Dirac delta function.

$$F_x(t, \nu, h) = \int_{-\infty}^{+\infty} \delta(u - t_0) h^*(u - t) e^{-j2\pi\nu u} du \tag{18}$$

The Dirac delta function is equal to zero in all point except in $t_0$, therefore, the integration gives :

$$F_x(t, \nu, h) = h(t - t_0) e^{-j2\pi\nu t_0} \tag{19}$$

In the ideal case, the short-time frequency function should be equal to 1 in $t_0$ and zero elsewhere. Therefore, it appears that the time resolution, i.e. the spread of the $F_x(t, \nu, h)$ around $t_0$ is proportional to the duration of the analysis window h.

In order to examine the frequency resolution issues, we consider another signal with some particular properties, a complex sinusoid.

$$x(t) = e^{j2\pi\nu_0 t} \tag{20}$$

Where $x(t)$ is a complex sinusoid of frequency $\nu_0$. The Short-time Frequency transform is given by :

$$F_x(t,\nu,h) = \int_{-\infty}^{+\infty} h^*(u-t)e^{-j2\pi(\nu-\nu_0)u}du \tag{21}$$

If H is the Fourier transform of the window function h, we have :

$$F_x(t,\nu,h) = e^{-j2\pi t\nu_0}H(\nu-\nu_0) \tag{22}$$

Similarly as in the time resolution case, the *width* of the window function in the frequency domain is the cause of the spreading in frequency. The *width in frequency domain* is the bandwidth of the band-pass filter corresponds to h. Here, it appears that the frequency resolution is proportional to the bandwidth of h.

In order to quantify further, we can consider the box window function. If the discrete signal is sample with time interval $\delta t$ and the width of the window is T, then the available frequencies will go from $T^{-1}$ to $(2\delta t)^{-1}$ (Shannon principle). This put limits on low and high frequencies. If the higher frequency boundary is fixed by the sampling rate, the lower frequency bound depends on T. The large T, the lower the frequency bound.

Therefore, a good frequency resolution requires a large window function while the good time resolution is obtained with smaller window functions. This demonstrate the problems associated with the Heisenberg-Gabor Inequality expressed in section 2.2.

In practice, several windows functions as well as widths will have to be used to find the best tradeoff for the analysis. Nevertheless, the inequality will remain so that for better accuracy, another technique must be used.

In the next section, we will examine a candidate transformation that will eliminate the tradeoff problem by using *windows* with widths varying with the frequency or *scale*.

# 3 Wavelet Transform

## 3.1 Continuous Wavelet Transform (CWT)

### 3.1.1 Definitions

We have seen in section 2 that the Time-Frequency localization problem could be addressed by deriving the Fourier approach for spectrum analysis. In the short-time frequency transform (windowed Fourier transform), a signal or a time series is expanded into a series of elementary wave called atoms obtained from the translation of an elementary function (window function or mother function).

One of the limitation of the STFT approach is the Time-Frequency resolution tradeoff due, in parts, to the fact that the atoms are obtained only by translation of the elementary function. As a consequence, the *atomic functions* have the same *width* for all frequencies [3, 9].

In the Wavelet transform, the atoms will be obtained by translation but also by dilation of the elementary function. In order to match the notations used in the wavelet field, the atoms will be called *Wavelets* while the elementary function associated with the window function will be called *mother function*.

Just like we have defined the Short-time Fourier transformation in Equation 11, we can give a similar formulation for the *Wavelet transformation* [7, 8].

$$W_x(t, a, \Psi) = \int_{-\infty}^{+\infty} x(u)\bar{\Psi}_{t,a}(u)du \tag{23}$$

Where

x is the time series.

$W_x(t, a, \Psi)$ is the Continuous Wavelet transformation of x.

$\Psi^*_{t,a}$ is the wavelet function $\Psi \in \mathcal{L}^2(\Re)$.

a is the scale factor.

In STFT, $h_{t,\nu}$ was obtained from the translation of the elementary $h$ window function as expressed in Equation 14. The wavelet function is defined by :

$$\Psi_{t,a}(u) = \|a\|^{\frac{1}{2}}\Psi(\frac{u-t}{a}) \tag{24}$$

In the definition of the wavelet in Equation 24, the role of the scale factor appears clearly as the dilation factor. With $a > 1$ the elementary wavelet or mother wavelet is scaled down while the values superiors to one lead to dilations [**?**]. For a unity scale, the wavelet is located around the single frequency $\nu_0$. In the wavelet field, we will no refers to scales rather than frequencies (in the Fourier analysis acceptance of the term). However, it is possible to define the relationship between scales and frequencies.

$$\nu = \frac{\nu_0}{a} \tag{25}$$

The continuous wavelet transform is said to belong to the affine transformation because it is covariant by translation in time and scaling.

For the transformation to be inverted, the wavelet $\Phi$ must satisfy the so-called *admissibility condition*. To define the admissibility criterion, we introduce the parameter $C_\Phi$.

$$C_\Phi = \int_{-\infty}^{+\infty} \frac{\|\hat{\Phi}(w)\|^2}{w}dw \tag{26}$$

Where $\hat{\Phi}$ is the Fourier transform of the wavelet $\Phi$.

The admissibility condition imposes that $C_\Phi$ is finite. That condition is responsible for much of the properties of the wavelet functions.

First, $\frac{\|\hat{\Phi}(w)\|^2}{w}$ must be define by continuity in zero which imposes :

$$\hat{\Phi}(0) = 0 \tag{27}$$

The immediate consequence of Equation 27 is that the wavelet functions must exhibit oscillatory behaviours.The second consequence is that to converge, the norm must be bounded by a function that converge when integrated.

$$\|\hat{\Phi}(w)\|^2 w < w \tag{28}$$

Equation 28 requires the wavelet function to decay. In other words, a wavelet function must have a zero mean and be localized both in time and frequency. As a matter of fact, the wavelet functions were given that name because they look like attenuated symmetrical oscillation, like small waves (*ondelette* in French).

When the parameter $C_\Phi$ is known, it is possible to give the inverse transformation definition.

$$x(t) = \frac{1}{C_\Psi} \int_{-\infty}^{+\infty} W_x(u, a, \Psi) \Psi_{u,a}(t) du \frac{da}{a^2} \tag{29}$$

In order to refine the mathematical definition, we have to distinguish the time t parameter used in the time series and the one used to define the wavelet function. As a matter of fact, even is a wavelet is defined by two parameters that can be associated to time and frequency, it is more accurate to used the terms of *nondimensional time parameter* and *scale parameter*. From now on, in the definition of the wavelets, the parameters $\eta$ and s will be substituted to the t and $\nu$ parameters respectively [14].

Although a wavelet will be a function these two parameters, the *mother wavelet* from which the others are defined is defined for the initial scale $s_0$. In that case, the mother wavelet is defined as a function of $\eta$. An example is the Marr or Mexican hat wavelet :

$$\Phi_{s_0}(\eta) = (1 - 2\eta^2) e^{-\eta^2} \tag{30}$$

To further simplify the notation $\Phi_0(\eta)$ will also be used to express the mother wavelet function.

The *Mexican hat* function is a particular case of a family of mother wavelet functions called DOG (Derivative Of a Gaussian) and is obtained for a second derivative order. This function is said to have *two vanishing moments*.

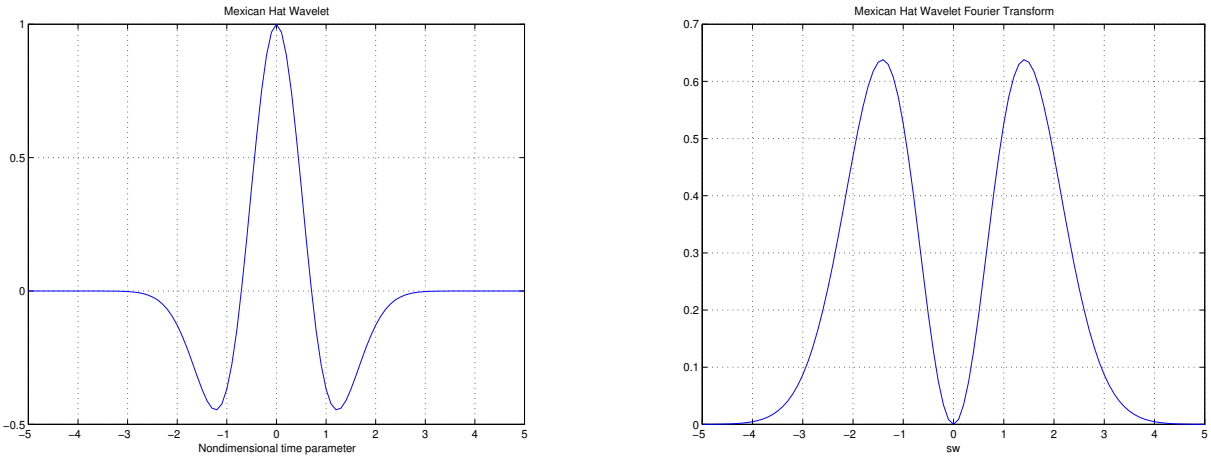


Figure 5: Mexican Hat Wavelet and its Fourier Transform

We will see later that the mother wavelet will define a basis and are therefore also called *wavelet basis functions*.

### 3.1.2 Continuous Wavelet Transform of a Discrete Sequence

When the continuous wavelet transform is applied on continuous signals, the integration are replaced by summations. Because this particular case is the most common case, we will detail the definitions.

First, we characterize the sampled signal or time series. The time series $x_n$ is defined by N samples $(n = 0 \cdots N-1)$ with a sampling period $\delta t$.

Equation 23 gave the relation for the continuous transformation on continuous signals.

$$W_{x_n}(n\delta t, s, \Psi) = \sum_{n\prime=0}^{N-1} x_{n\prime} \bar{\Psi}_{n\delta t, s}(n\prime\delta t) \tag{31}$$

Equation 31 give the definition of a wavelet transform with a discrete signal. Similarly as in Equation 24, we have :

$$\bar{\Psi}_{n\delta t, s}(n\prime\delta t) = (\frac{\delta t}{\quad})^{\frac{1}{2}} \Psi(\frac{(n\prime - n)\delta t}{\quad}) \tag{32}$$

In order to simplify the notation, the continuous wavelet W will be expressed as $W_n(s)$ and the wavelet function $\Psi(\frac{(n\prime-n)\delta t}{s})$.

In some case, it is faster to calculate the wavelet transform in th Fourier space. In the continuous limits, the transform of $\Psi(t/s)$ is $\hat{\Psi}(sw)$ (where $\hat{}$ is the Fourier transform operator. In that case, the wavelet transform is given by Equation 33.

$$W_n(s) = \sum_{k=0}^{N-1} \hat{x}_k \bar{\hat{\Psi}}(sw_k) e^{iw_k n\delta t} \tag{33}$$

Where the angular frequency is defined by :

$$w_k = \begin{cases} \frac{2\pi k}{N\delta t} & \text{if } k \leq \frac{N}{2} \\ \frac{-2\pi k}{N\delta t} & \text{if } k > \frac{N}{2} \end{cases} \tag{34}$$

We will give the relation for the reconstruction of $x_n$ in the section 3.2 as the most common case is the use of discrete wavelet transform on discrete signals.

Before introducing the notion of *frames*, we first examine the Time-Frequency resolution issues.

#### 3.1.3 Time-Frequency Resolution

The scaling is the responsible for the main differences between wavelets and windowed Fourier transforms : when the scale changes, the bandwidth and the duration of the wavelets are changed but unlike the STFT, the shape of the function does not change. If we look at the wavelets under the window angle, we could say that the transformation uses short windows for high frequencies while it uses large windows for low frequencies. In that case, the Time-Frequency resolution is almost constant through the Time-frequency space [15].

The Heisenberg-Gabor inequality expressed in Equation 7 no longer applies and is replaced by :

$$\frac{B}{\nu} = Q \tag{35}$$

Where

B is the bandwidth.

$\nu$ is the frequency.

Q is the constant.

It is possible to make an analogy with the windows used in short-time Fourier transform. We will consider that the wavelet function defines the window. Therefore, the center and the width or the window are defined by :

$$\bar{x} = \frac{1}{\|\Psi\|^2} \int_{-\infty}^{+\infty} x \|\Psi\|^2 dx \tag{36}$$

Equation 36 defines the center of the hypothetical window while Equation 37 defines its widths in the time domain.

$$\Delta_x^2 = \frac{1}{\|\Psi\|^2} \int_{-\infty}^{+\infty} (x - \bar{x})^2 \|\Psi\|^2 dx \tag{37}$$

Similar definition can be obtain to define the center and the width in the frequency domain ($\bar{s}$, $\Delta_s^2$).

An important property is that the area corresponding to the wavelet window is constant and is given by $4\Delta_x\Delta_s$.

### 3.2 Discrete Wavelet Transform (DWT)

#### 3.2.1 Definition

In the continuous wavlet transform, the parameters $\eta$ and s are defined on continuous intervals. However, we when they are restricted to a set of discrete values, the continuous wavelet transform is called *Discrete Wavelet Transform* [12].

We can define the sampling of the time-frequency plane, or $(\eta, s)$ plane, by :

$$(\eta, s) = (nt_0 s_0^{-m}, s_0^{-m}) \tag{38}$$

Where

$\eta_0 > 0$

$s_0 > 0$

$n, m \in Z$

In that context, the discrete wavelet transform is given by :

$$W_x(n, m, \Psi) = s_0^{m/2} \int_{-\infty}^{+\infty} x(u)\bar{\Psi}_{n,m}(u)du \tag{39}$$

The wavelet function is then defined by :

$$\Psi_{n,m}(u) = \Psi(s_0^m u - n\eta_0) \tag{40}$$

The field of *Theory of Frames*, provides with an insights on the choice of the *optimal* discrete values of s and $\eta$ (see [16]). In short, a can be a set of vectors that defines a non-orthonormal basis.

A very popular frame is the so-called *dyadic grid* where the scale and nondimensional parameters take the following values.

$$\text{Dyadic Grid } \Rightarrow \begin{cases} s = s_0 2^{-j} \\ \frac{\eta}{s} = l \end{cases} \tag{41}$$

Where j and l $\in Z$.

For discrete signal or time series, the Shannon theorem indicates that the minimum useful period is twice the sampling period. For that reason, a popular choice are $\eta_0 = 1$ and $s_0 = 2$.

The use of such a grid is also referenced as a dyadic sampling of the time-frequency plane. In order to make some comparisons, we can examine the sampling plane for four transformations, the Discretization, the Fourier transform, the Short-Time Fourier Transform and the Discrete Wavelet Transform.

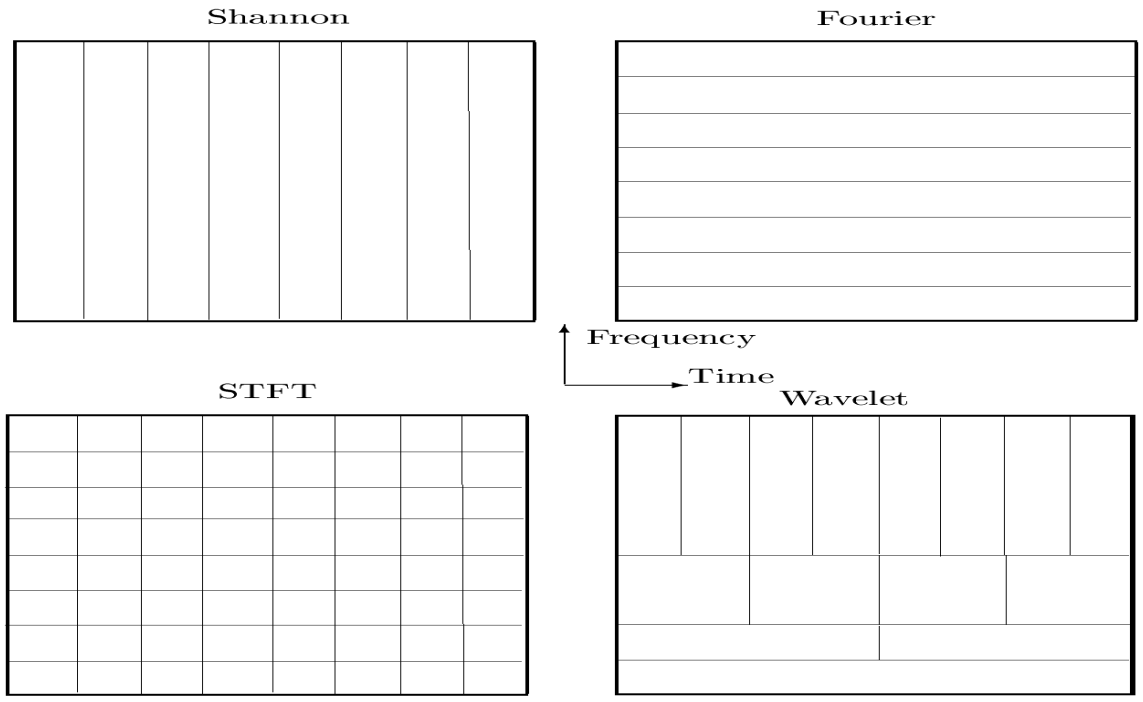


Figure 6: Sampling in Time-Frequency Plane.

Figure 6 represents the four basic grids used in time series analysis. The first one represents called *Shannon* corresponds to the discretization of the time series. The continuous plane is divided into time intervals. In the Fourier transform, the time-frequency plane is divided into frequency intervals (defined by multiples of the basic frequency $\frac{1}{2\delta t}$) and the spectrum is evaluated on the entire time series. In the Short-Time Fourier Transform, the plane is divided into square areas of equal dimensions, the windows. For the wavelets, we can see that the windows have different sizes for different frequencies and that

the sizes follow a geometric progression. The lower frequencies (longer periods) are evaluated on large (time direction) windows while the high frequencies are evaluated on smaller (time dimensions) windows. However, it is important to remember that the *surface* defined by the windows are constant.

Almost as important as the definition of the transformation, the inverse transformation is also very important when the wavelet are used not only for time series analysis but also for feature extraction or multiresolution modelling. However, before addressing these issues, we will first examine some particular wavelets.

### 3.2.2 Nonorthogonal Wavelets

There are many wavelet function used in the wavelet field so that it is not possible to perform an exhaustive enumeration. However, the reasons we will detail in section 3.2.5, the use of nonorthogonal wavelet function seems to be more adapted for time series modelling [14].

It is possible to classify the wavelet as orthogonal or non-orthogonal, real or complex, defined analytically or by recursive definitions *dilation equation*. Here, we will first examine a set of wavelets defined analytically.

| Name | $\Psi_0\eta$ | $\hat{\Psi}_0(sw)$ | $e$-folding times | Fourier Wavelength |
|---|---|---|---|---|
| Morlet | $\pi^{-1/4}e^{jw_0\eta}e^{-\eta^2/2}$ | $\pi^{-1/4}H(w)e^{-(sw-w_0)^2/2}$ | $\sqrt{2}s$ | $\frac{4\pi s}{w_0+\sqrt{2+w_0^2}}$ |
| Paul | $\frac{2^m j^m m!}{\sqrt{\pi(2m)!}}(1-jn)^{-(m+1)}$ | $\frac{2^m}{\sqrt{m(2m-1)!}}H(w)(sw)^m e^{-sw}$ | $s/\sqrt{2}$ | $\frac{4\pi s}{2m+1}$ |
| DOG | $\frac{(-1)^{m+1}}{\sqrt{\Gamma(m+\frac{1}{2})}}\frac{d^m}{d\eta_m}(e^{-\eta^2/2})$ | $\frac{j^m}{\sqrt{\Gamma(m+\frac{1}{2})}}(sw)^m e^{-(sw)^2/2}$ | $\sqrt{2}s$ | $\frac{2\pi s}{\sqrt{m+\frac{1}{2}}}$ |

Table 1: Nonorthogonal Wavelet Basis Functions

Table 1 represents three families of nonorthogonal wavelet functions defined analytically. The Morlet and Paul wavelet functions are complex while the DOG (derivative of Gaussian) is real. In the third column representing the Fourier transform of the functions, $H(w)$ represents the Heaviside step function.

In each family, a particular wavelet function is obtained by choosing one parameter, the frequency $w_0$ for the Morlet wavelet, the order m for the Paul family and the derivative order m for the DOG. it is important to notice that in the case where $m = 2$ in the DOG family, the resulting wavelet function is called the *Marr* or *Mexican Hat* wavelet.

The *width* of the wavelet is defined by the *e-folding time* of the wavelet amplitude. As seen already, the resolution is a tradeoff between the width in Fourier and Time space.

Figure 7 represents two wavelet basis examples, the Morlet and the Paul wavelets. On the left column, the complex wavelet function are represented. The solid line represents the real part while the dash line represents the complex part of the function. The right column represents the Fourier transform of these function.

A similar representation of the DOG wavelet can be found on Figure 30.

### 3.2.3 Orthogonal and Fractal Wavelet Functions

Another historically important wavelet function is the so-called Meyer wavelet [16]. The Meyer wavelet is the function $\Psi \in \mathcal{L}^2(\Re)$ defined by :

$$\hat{\Psi}(s) = e^{js/2}\omega(\|s\|) \tag{42}$$

$$\omega(\|s\|) = \begin{cases} 0 & \text{if } s \leq \frac{1}{3} \text{ or } s \geq \frac{3}{4} \\ \sin\frac{\pi}{2}v(3s-1) & \text{if } \frac{1}{3} \leq s \leq \frac{2}{3} \\ \cos\frac{\pi}{2}v(\frac{3}{2}s-1) & \text{if } \frac{2}{3} \leq s \leq \frac{3}{4} \end{cases} \tag{43}$$

Where the function v is defined by :

$$v = \begin{cases} 0 & \text{if } s \leq 0 \\ 1 & \text{if } s \geq 1 \\ v(s) + v(1-s) = 1 & \text{if } s \in [0,1] \end{cases} \tag{44}$$

With some computation, it can be verified that $\|\Psi\| = 1$ and $\sum \|\hat{\Psi}(2^n s)\|^2 \equiv 1$. Such equalities finally implies that the wavelet is an orthonormal basis.

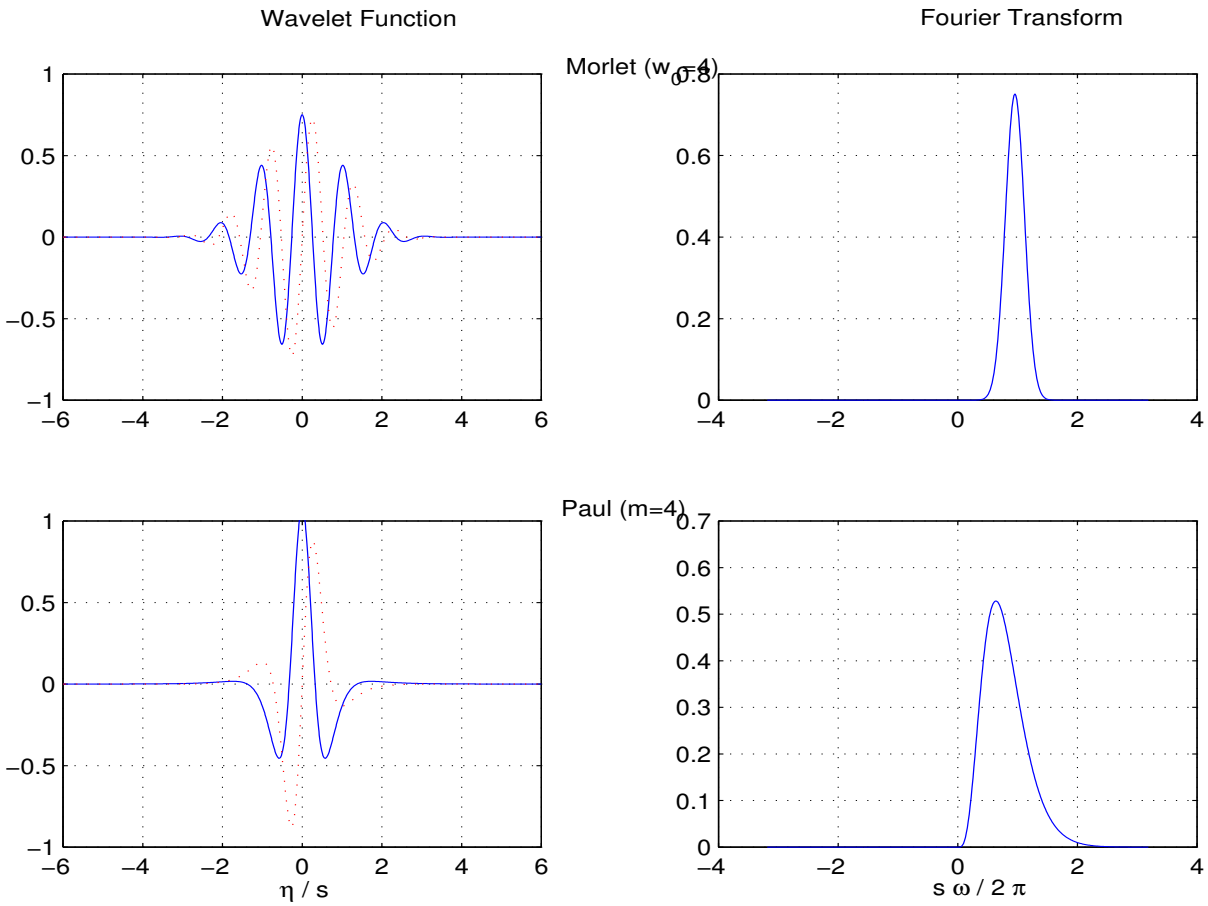


Figure 7: Two Wavelet Bases Examples.

The recursive definition of v in Equation 44 enable to construct families of orthonormal basis function.

These families of recursively defined functions, also called *Fractal wavelets* are all issued from a basic definition called the *dilation*, *refinement* or *two-scale difference equation* of a *scaling function* $\phi$ (also called father wavelet).

$$\phi(x) = \sum_{k=0}^{M-1} c_k \phi(2x - k) \tag{45}$$

Where M is the number of nonzero coefficient and is referred as the *order of the wavelet*.

The values of the coefficients $c_k$ are not arbitrary but determined by the constrains of orthogonality and normalization.

$$\sum_k c_k = 2 \tag{46}$$

Equation 46 ensures the uniqueness of $\phi$, the scaling function.

The wavelet function is defined from the scaling function by :

$$\Psi(x) = \sum_k (-1)^k c_{1-k} \phi(2x - k) \tag{47}$$

This class of wavelet, called *Fractal Wavelet* is constrained by definition to be equal to zero outside a small interval [8, 9, 11]. This property called *compact support* enable to deal with finite samples. Because of the recursion, the wavelet is nondifferentiable *everywhere*.

Practically, the functions are defined for specific set of parameters $c_k$

The Haar wavelet is obtained from the box function with a second order regression, using $\phi(x)$, $\phi(2x - 1)$.

Similarly, the Hat wavelet is obtained with a third order regression.

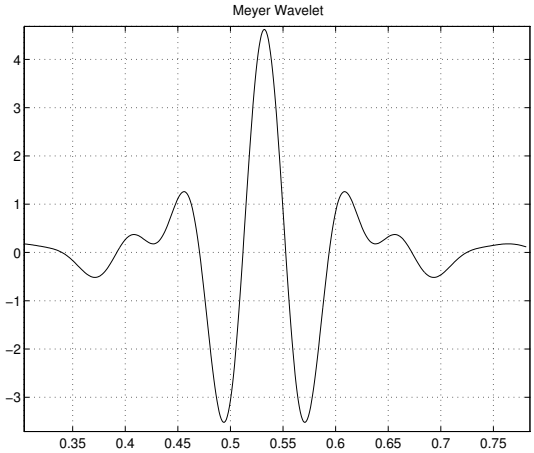


Figure 8: Meyer Wavelets

| Wavelet | $c_0$ | $c_1$ | $c_2$ | $c_3$ | $c_4$ | $c_5$ |
|---|---|---|---|---|---|---|
| Haar | 1 | 1 | | | | |
| Dubechies-4 | $\frac{1}{4}(1+\sqrt{3})$ | $\frac{1}{4}(3+\sqrt{3})$ | $\frac{1}{4}(3-\sqrt{3})$ | $\frac{1}{4}(1-\sqrt{3})$ | | |
| Daubechies-6 | 0.332671 | 0.806891 | 0.459877 | -0.135011 | -0.085441 | 0.035226 |

Table 2: Coefficient of Three Fractal Wavelets

The fractal wavelet defined by the hat function is :

$$\Psi(x) = \frac{1}{2}\phi(2x) + \phi(2x-1) + \frac{1}{2}\phi(2x-2) \tag{48}$$

### 3.2.4 Reconstruction

With orthogonal wavelet, there is a unique decomposition on the basis while with nonorthogonal wavelets such as those expressed in Table 1, the degrees of freedom allow to use arbitrary sets of scales. This can be a very important property we one want to *extract* or emphasize certain scales within the time series [14, 15].

Using a dyadic grid, it can be convenient to write the scales as a fractional powers of two.

$$s_j = s_0 2^{j\delta j} \tag{49}$$

Where $s_0$ is the smallest scale wavelet (should not be smaller period than $2\delta t$ according to Shannon theorem).

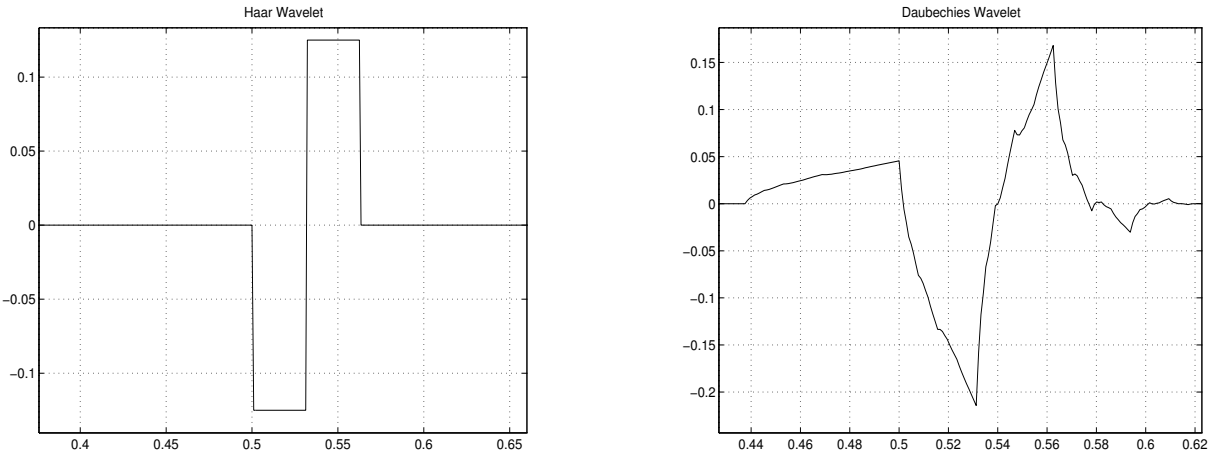


Figure 9: Haar and Daubechies-4 Wavelets

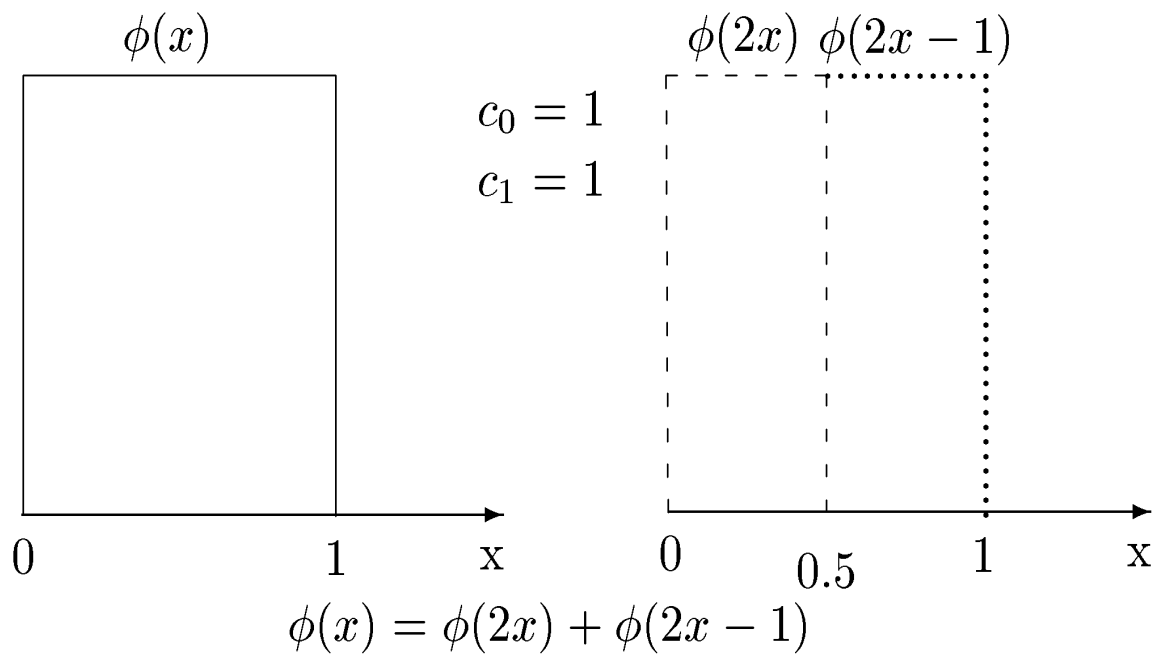


Figure 10: Construction of the Haar Wavelet.

The index j defines the scale index and varies between two values, 0 and J ($J\delta j$ determines the largest possible scale given the limited time series).

$$J = \frac{Log_2(\frac{N\delta t}{s_0})}{\delta j} \tag{50}$$

Broadly speaking, J represents the decomposition in power of two of the sample size, or the number of time the time series can be *folded* on itself.

The choice of the increment $\delta j$ defines the resolution of the wavelet. For Morlet wavelets, $\delta j = 0.5$ is the largest value that still will gives good scaling.

As the transformation with arbitrary scales and nonorthogonal wavelets gives more degrees of freedom, the inverse transformation is not unique. The redundancies generated by the nonorthogonality makes also possible to reconstruct the time series with another wavelet function, the easiest being the *Dirac impulse* or *Delta* function.

The reconstruction of the time series is then obtained with a formula similar to the inverse Fourier transform.

$$x_n = \frac{\delta j \delta t^{1/2}}{C_\delta \Psi_0(0)} \sum_{j=0}^{J} \frac{\Re\{W_n(s_j)\}}{{s_j}^{1/2}} \tag{51}$$

The parameter $c_\delta$ comes from the reconstruction of the Dirac impulse $\delta$.

$$C_\delta = \frac{\delta j \delta t^{1/2}}{\Psi_0(0)} \sum_{j=0}^{J} \frac{\Re(W_\delta(s_j))}{{s_j}^{1/2}} \tag{52}$$

Where $W_\delta$ is given by :

$$W_\delta(s) = \frac{1}{N} \sum_{k=0}^{N-1} \hat{\psi}(s\omega_k) \tag{53}$$

For standard, it is possible to use a table of $C_\delta$. Table 3 represents this parameter for the Wavelet function shown in Table 1.

Where

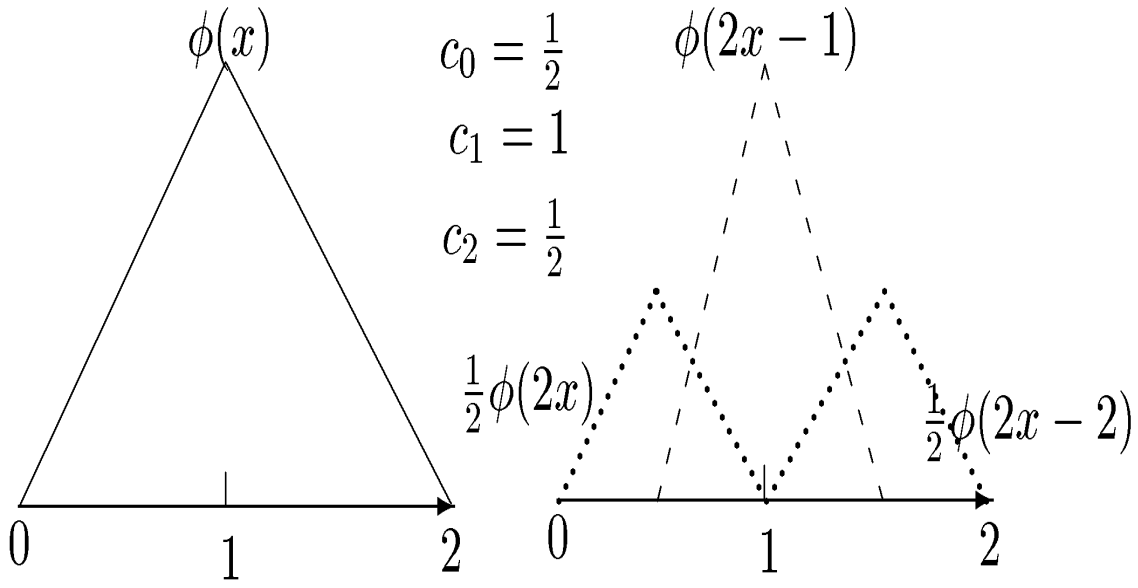


Figure 11: Construction of the Hat Wavelet.

| Name | $C_\delta$ | $\gamma$ | $\delta_{j_0}$ | $\Psi_0(0)$ |
|---|---|---|---|---|
| Morlet ($w_0 = 6$) | 0.776 | 2.32 | 0.60 | $\pi^{-1/4}$ |
| Paul (m=4) | 1.132 | 1.17 | 1.5 | 1.079 |
| Marr (m=2) | 3.541 | 1.43 | 1.4 | 0.867 |
| DOG (m=6) | 1.966 | 1.37 | 0.97 | 0.884 |

Table 3: Empirically Derived Factors.

$C_\delta$ is the reconstruction factor.

$\gamma$ is the decorrelation factor for time averaging.

$\delta j_0$ is the factor for scale averaging.

### 3.2.5 Properties and Choices

We have seen that wavelets can be built from various *scaling functions* (father wavelets) and that the discretization of the scaling can be arbitrary defined. Therefore, just like the general issues involving system identification, the problem is to choose the right wavelet for the right time series [14, 16].

Based on the experience of [14], it is possible to give some general guidelines on the properties of the wavelets. The first choice involves the determination of the characteristics of the wavelet functions.

The main characteristics concerns the orthogonality or the nonorthogonality of the basis defined in the wavelet space. Orthogonal wavelet will define the most compact representation of the signal which also mean less freedom for the representation. For time series, it is more recommended to use nonorthogonal wavelets functions (see Table 1) when the series are smooth with continuous variations in the wavelet amplitude. The main drawback is the arbitrary freedom lead to redundancies that will have to be extracted using other methods (PCA).

The nature of the wavelet, real or complex, is also important. A complex wavelet function provide information by the means of the amplitude but also the phase of the wavelet and is therefore particularly adapted for oscillatory time series. On the other hand, real valued wavelets will be seen as *more robust* when dealing with peaks or discontinuities in the time series.

The shape of the wavelet will be correlated to the type of features that can potentially be extracted or highlighted. For *crisps* time series with jumps or steps, box or Haar like wavelet will perform better. For more continuous time series, smooth functions such as damped cosine (Meyer wavelet) will be appropriate.

When the wavelet function is determined (even through a *trial and errors* search process), and if nonorthogonal wavelets have been chosen, the issues of the choice of scales must be addressed. For orthogonal wavelets, the scales are imposed while with nonorthogonal, the possibility to arbitrarily chose them is given.

As already expressed in section 3.2.4, the minimal scale should be determined so that the Shannon theorem on sampling is respected. Basically, if the sampling period is $\delta t$, the minimum scale $s_0$ should correspond at least to the equivalent Fourier wavelength $2\delta t$. Although the Fourier wavelength and the wavelet scales have the same order of magnitude, it is still possible to have an exact expression (see Table 1.

In Equation 49, we chosen to express the scale as a power of two, given the Dyadic space. The choice of $\delta j$ define the scale increment and therefore the frequency resolution. Here again, the choice of $\delta j$ will have to be a trade-off between resolution (small $\delta j$) and better significance (large $\delta j$. Here again, choices will have to be made, involving trial and errors but also understanding of time series behaviour. For example, $\delta j = 0.5$ for Morlet wavelets gives satisfactory results.

# 4 Wavelets and Time Series Analysis

## 4.1 Introduction

The wavelets can be used principally for a the analysis of the time series analysis by providing more accurate time-frequency insights [15, 17]. In turn, the analysis of the time frequency features can be used in the modelling itself by using harmonic model, by helping in the choice of regression orders or bootstraps characteristics.

In a second phase, the analysis of the time series will help to define an optimal transformation of the time series. The results of the transformation can be used to extract frequency related information and reduce the complexity of the input space (compression) but can also be used to perform a multiresolution modelling. We will detail the issues of the multiresolution modelling as it has several meaning in the wavelet field.

First, we will examine the application of the wavelet transform to the time series analysis illustrated by the competition time series example.

## 4.2 Time-Frequency Analysis with Wavelets

### 4.2.1 Wavelet Power Spectrum

In section 2.4.2, we have performed the Time-Frequency analysis of the competition time series using short-time Fourier transform. Figure 4 represents the square modules of the Fourier transform. It appears that the main frequency components occurs for frequencies around 0.026 although the representation of the time series on Figure 1 would indicate more oscillatory behaviour throughout the time series.

In order to have a better resolution, we turn to wavelet transform. In order to make a broad comparison, we perform the wavelet transform with a Morlet wavelet function. For better resolution, we choose the initial scale $s_0$ to be 2 on a dyadic grid. The scale arbitrary discretization are defined by Equation 49. The scale increment $\delta j$, in a first trial, is chosen to be 0.5.

Given the time series number of samples and the initial scale, Equation 50 defines J, the maximum number of scales. Here, with $N = 2000$, we obtain $J = 20$. Therefore, the time series will be expanded into 20 scales.

Figure 12 represents the number of scale (resolution) and their distribution on the dyadic grid for different values of $\delta j$, 0.25, 0.5 and 0.75.

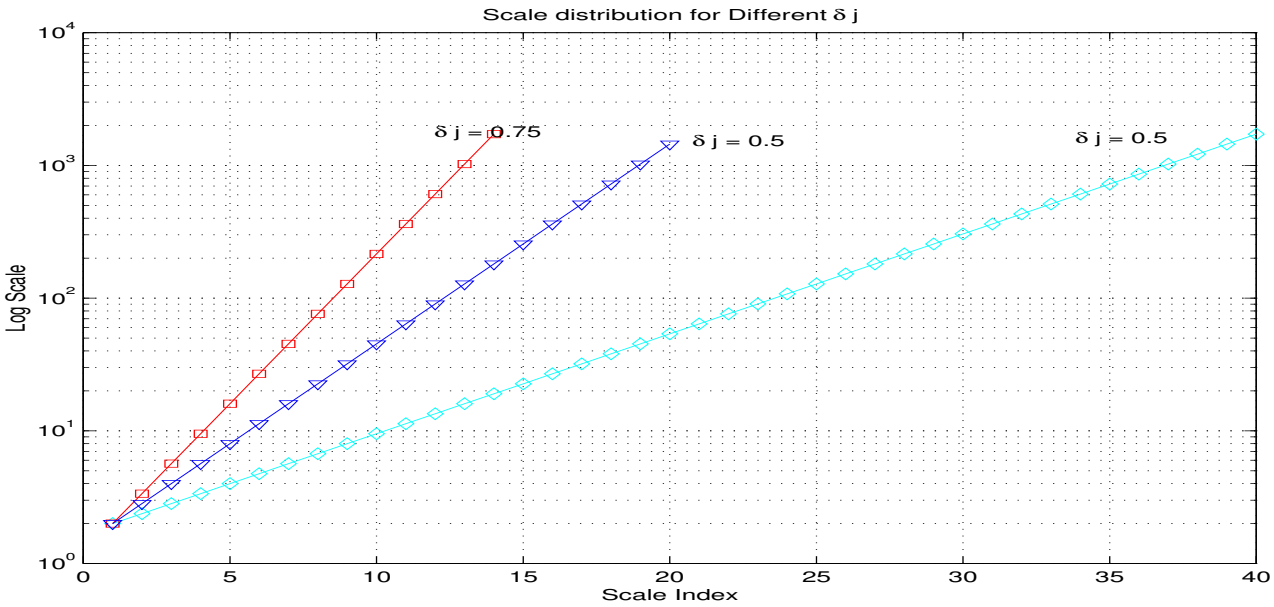


Figure 12: Distribution of the Scales.

The initial scale as well as the highest scale are defined by the time series, i.e. the sampling period and the number of samples. It is obvious that the maximal scale has an upper bound due to the limitation of

the time series. The equivalent Fourier wavelength corresponds to the size of the time series, here 2000 samples.

When the scale and the wavelet have been chosen, we must determine the criterion for the analysis. In order to continue the comparisons with Fourier analysis, we now introduce the concept of *Wavelet spectrum*.

The Morlet wavelet function is complex and so is the wavelet transform of the time series $W_n(s)$. The *Wavelet Power Spectrum* is then defined to be :

$$\|W_n(s)\|^2 \tag{54}$$

In order to make significance analysis and interpretation, it is required to normalize the power spectrum. If $\sigma^2$ is the variance of the time series and N is the number of samples, the *normalized wavelet power spectrum* (NWPS) is defined by :

$$NWPS_n(s) = N\frac{\|W_n(s)\|^2}{\sigma^2} \tag{55}$$

Equation 55 is the normalized power spectrum relative to white noise. Figure 13 represents the contour plot of the normalized wavelet power spectrum for the competition time series.

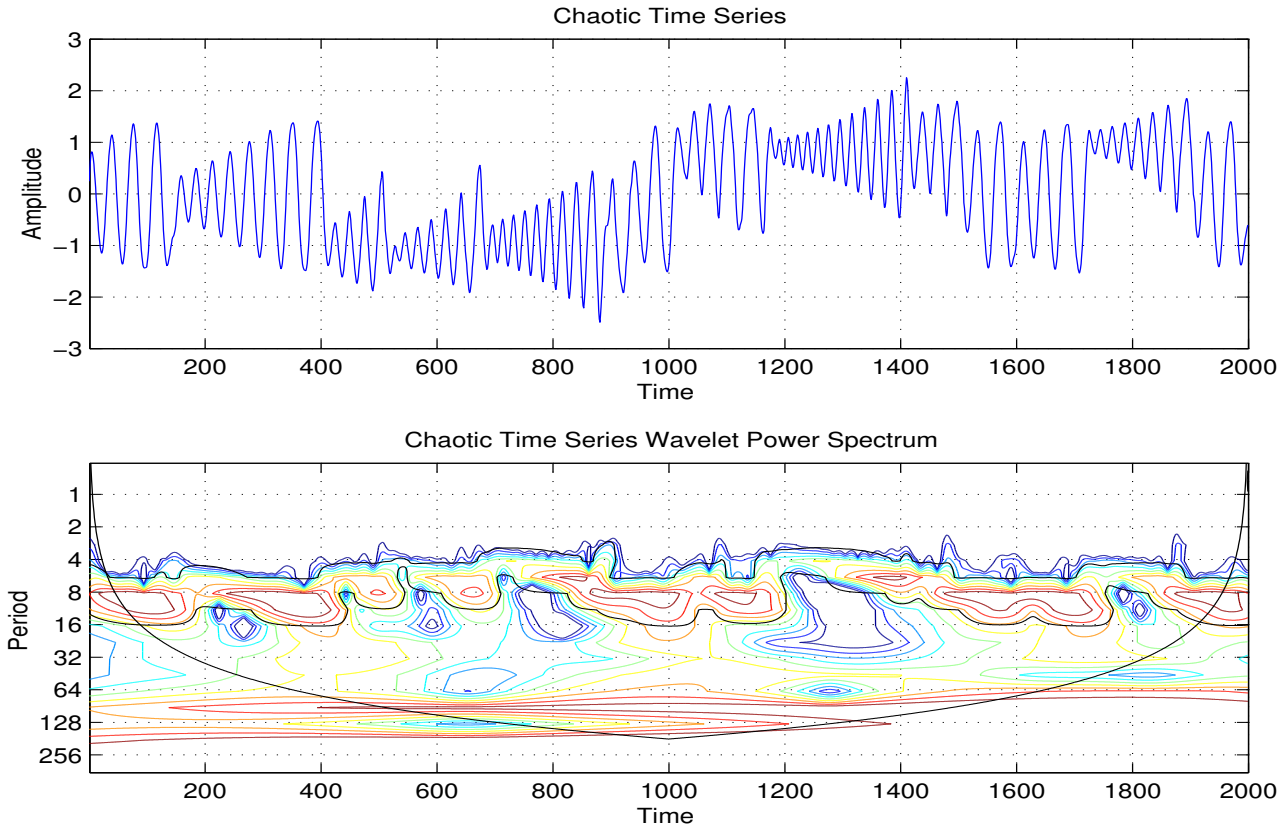


Figure 13: Chaotic Time Series and Local Wavelet Power Spectrum .

The maxima of the power spectrum are indicated in red while the minima are in blue. It appears that there are no oscillatory features associated with wavelength inferior to four. This is more that the two required by the Shannon theorem so that with have a good significance level for that assertion. On the other extreme of the scale, the maximum period that can be identified with at least a 95 % confidence limits is inferior to 128. The dark curve across the graph is called the *Cone of Influence* and expresses the uncertainty due to the limited data set (edge effects). only the maxima inside the cone of influence can be considered as reliable. For details on the calculation of the cone of influence, see [14].

A closer examination indicate that there are two significant scale (periods) *bands*. The first one is n the 8-16 area while the second one includes scales superior to 64. Unfortunately the limited data set does not allow accurate estimation of that oscillatory behaviour. We will see how the local power spectrum can be transformed in global power spectrum to obtain that missing information.

On the other hand, we have plenty of data to analyze the 8-16 band feature. We can start by examining a close-up in the time-period space.

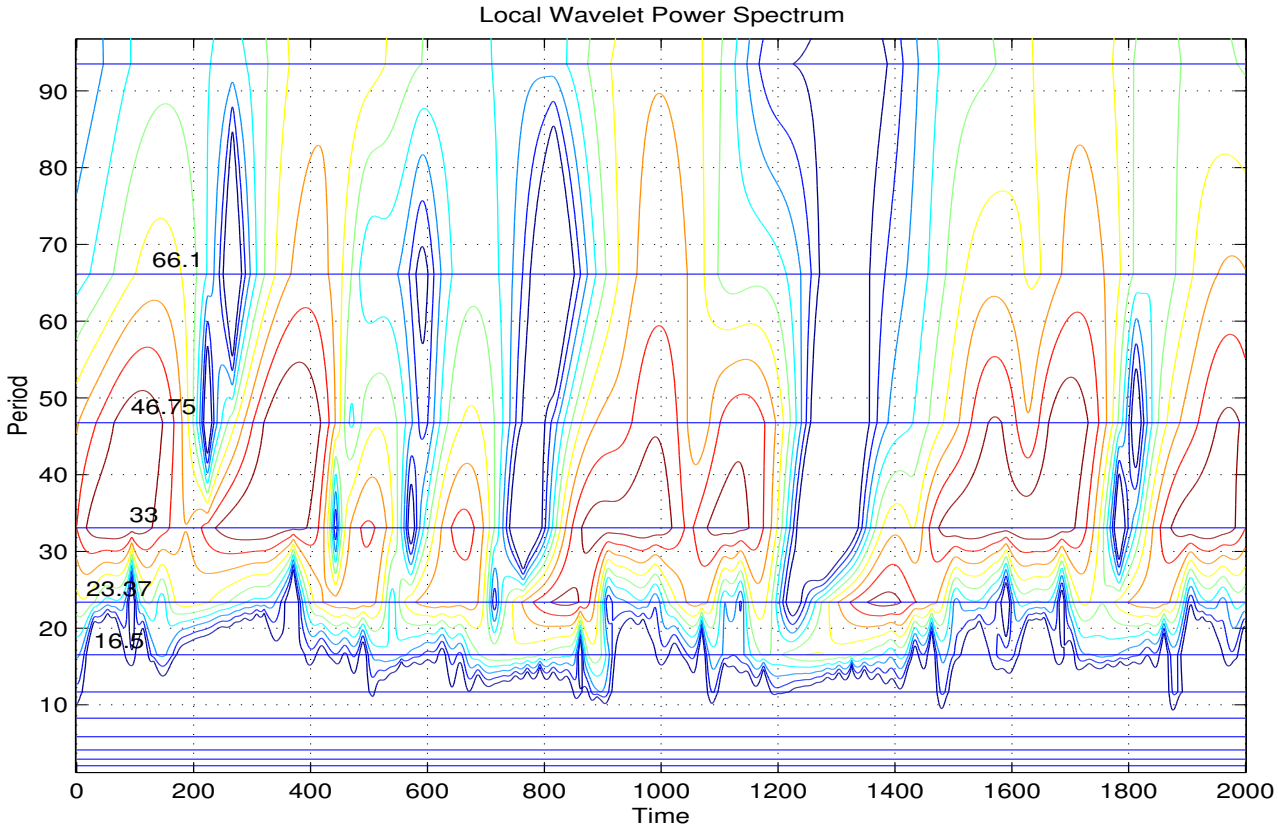


Figure 14: Close-up on Local Power Spectrum.

Figure 14 represents a close-up of the local wavelet power spectrum. The horizontal line indicates the dyadic grid in the frequency domain on which the power spectrum is interpolated.

### 4.2.2 Other Wavelet Functions

We have already mentioned that the choice of the wavelet function and its parameters can be critical for obtaining satisfactory results. In section 4.2.1, we have used the Morlet's wavelet with $w_0 = 6$ (see Table 1) [14]. The Morlet wavelet function characteristics (width and shape) can be tuned precisely. As the time series exhibited chaotic behaviour, it was decided to use a function that does not provide with the best time resolution (relatively large in time domain) but with a better frequency resolution (the Fourier transform of the wavelet give a narrow band-pass filter).

Other nonorthogonal wavelet functions such as Paul's wavelet and the Derivative of Gaussian (DOG) wavelets can be tried. Figure 15 represents the power spectrum of the competition time series using Paul's wavelet.

Comparisons between Figure 15 and 14 indicate that Paul's wavelet lead to better time resolution but worse frequency distribution. The peaks are spread vertically in the frequency domain. With such a deformation (lack of resolution), it is difficult to characterize the variation of the frequency in the oscillatory features. On the other hand, that transform gives clear indication on when the *switches* occurs.

Another important wavelet is the Derivative of Gaussian. Here, we examine the Marr wavelet (Mexican hat) corresponding to a DOG with the derivative parameter m set to 2 (see Table 1. That wavelet has a much wider time base so that the time resolution is not expected to give very accurate results. On the over hand, it is a narrower band-pass filter than the Morlet's wavelet so that better frequency resolution is expected. Figure 16 represents the power spectrum of the competition time series obtained with the Marr wavelet.

Figure 16 shows clearly the frequency characteristics of some oscillation modes. For the modes in the

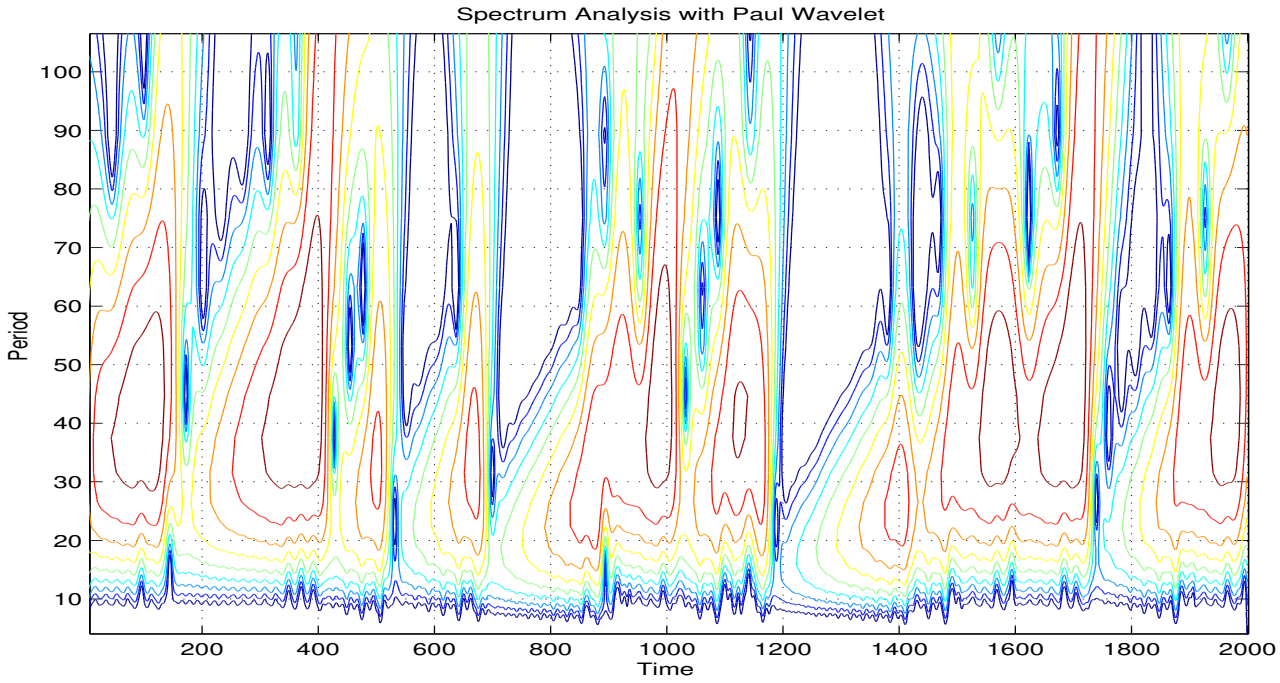


Figure 15: Power Spectrum with Paul Wavelet.

sample interval $[900, 1100]$ and $[1500, 1700]$ especially, things cannot be discerned efficiently. A close-up on the more stable areas shows valuable information on the frequency variations.

Figure 17 represents the close-up on the power spectrum on the time interval $[1100, 1500]$. During that time, the period of the oscillation are increasing almost linearly from a period of 15 to 25. Similar variations, although with different period interval, appears on the over stable oscillatory intervals. Such information could be used to create a modulated harmonic model of the time series. As already mentioned, such accurate estimation would not be possible with Short-Time Fourier transform.

From the two examples studied, it is clear that the choice of the relevant wavelet is important to emphasize some characteristics in the time series. However, it will be again a trade-off between time and frequency resolution. For analysis purposes only, it is recommended to use various wavelet function to find out different feature. However, intermediate wavelets such as Morlet's should be used for a first examination or even a non-destructive preprocessing of the data.

### 4.2.3 MultiResolution Analysis

It appears that the main oscillatory features occurs throughout the time series with a vertical elongation (frequency resolution) [15]. It also seems that the *period* of the oscillations are increasing between *switches* to modes. If the time resolution is satisfactory, the analysis can be improved by emphasizing the scales in the neighborhood of the oscillatory feature. The degree of freedom allowed by nonorthogonal wavelets enable to target features more precisely. It will be referred as *multiresolution analysis*.

The optimization is carried out by decreasing the scale increment $\delta j$. $\delta j$ is set to 0.05 instead of 0.5. The dyadic grid becomes much thinner with an overall 200 scales (instead of 20).

Figure 18 represents the resolution optimization perform to emphasize the features in the 20-25 (period) area. It appears that the division in the peaks are not artifacts due to estimates but relates to real discontinuities in the time series. Moreover, the peak areas can be all considered oblique, indicating an increase in the period of the pseudo-oscillation. The angle seems to be the same throughout the time series although the spectrum does not indicate any *cause* for the switches.

The overlaid horizontal lines represents the scales at which the time series is expanded. It is clear that the narrow gap between two scales will lead to overlapping and redundancies. Using numerous scales increases the computation load and does not allow to use the wavelet expansion directly. As we will see in section 5, there will be two ways for using the expansions.

We have used the degrees of freedom allowed by nonorthogonal wavelets to refine the frequency resolution. There is another way to perform multiresolution analysis, this time in the time domain.

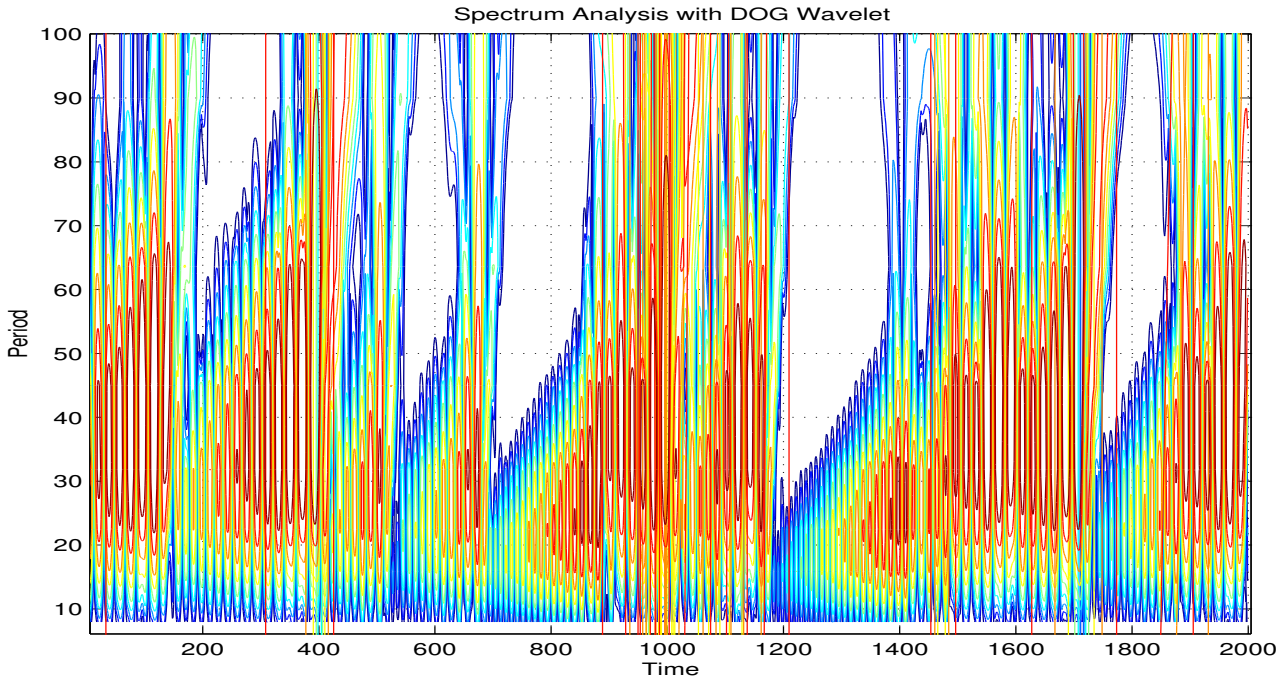


Figure 16: Power Spectrum with DOG Wavelet.

The Fourier transform can perform a spectral analysis on the full time series. The accuracy is obtained by the numerous periods available and provides a global spectrum. We have seen that for large period feature, the cone of influence lead to dismiss some patterns. It is possible to now consider the wavelet transform on a large resolution, in the time domain. Practically, this lead to obtain the *Time Averaged Wavelet transform*. By interfering the amplitude of the different scale components for each sample, it is possible to obtain statistics on the spectrum and then better accuracy for the global spectrum.

The *Global Wavelet Transform*, is defined by :

$$\|W(s)\|^2 = \frac{1}{N}\sum_{n=0}^{N-1}\|W_n(s)\|^2 \tag{56}$$

The *Normalized Global Wavelet Transform* (N is then defined by :

$$NGWT(s) = \frac{\|W(s)\|^2}{\sigma^2} \tag{57}$$

It is interesting to compare the global spectrum obtained by the wavelet transform and the Fourier transform. On Figure 19, the two spectrum are compared.

With the global wavelet transform, it is possible to obtain a spectrum significance function. The significance limit is obtain by comparing the power spectrum to the power spectrum of a red noise, i.e. the spectrum obtained from the residuals of a Markovian process. It is possible to obtain similar curve for the Fourier analysis but it accuracy does not make it practical.

The two global spectrum indicate more or less the same thing with a main component having a period of about 35 while another, much longer, is just above the significance threshold and correspond to a period of about 350. The power spectrum issued from the wavelets is smoother because of the limited scales available around the peak.

## 4.3 Scale Averaging

The wavelet transform enable to get Time-Frequency representation of the time series not only more accurate than the Short-time Fourier transform but also provide with degrees of freedom to optimization the local resolution [14]. We have seen that the price of that freedom is redundancy in the wavelet expansion. It is possible to obtain better quantitative results for the spectrum by examining a monocomponent (one frequency at a time). however, unlike in Fourier transform, it is possible to obtain the monocomponent on a limited frequency band so that the redundancies operate in a constructive way.

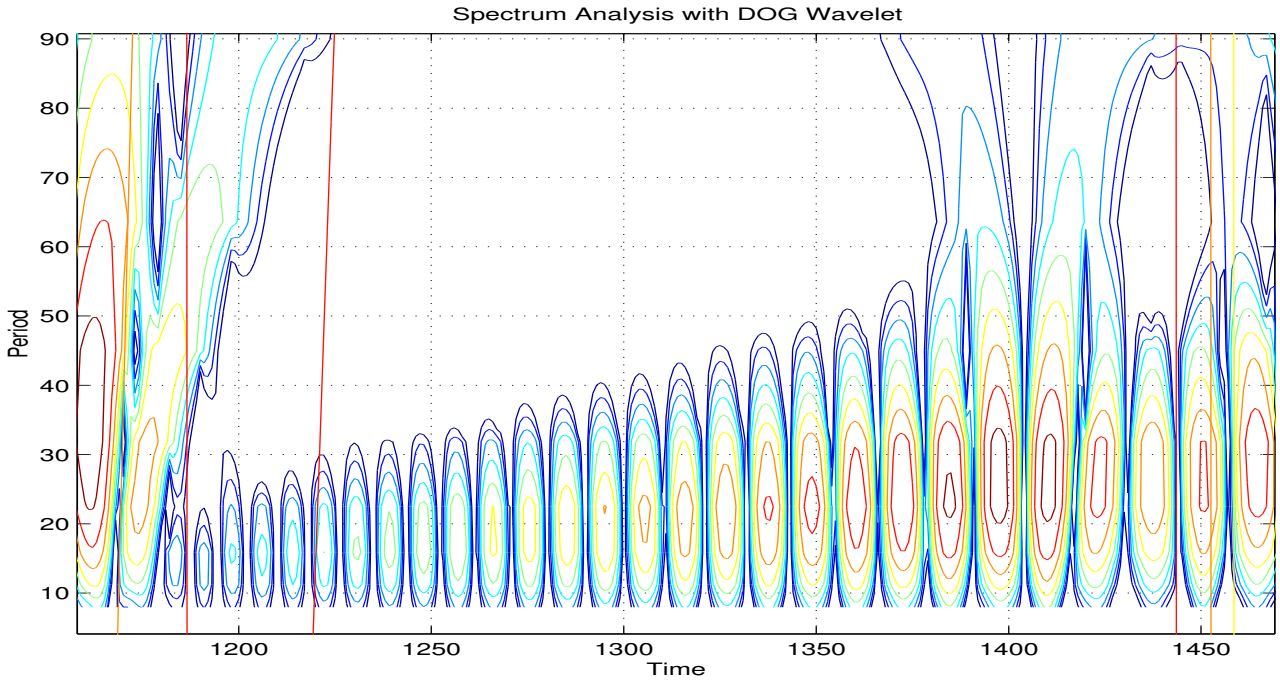


Figure 17: Close-up on Power Spectrum with DOG Wavelet.

The technique involving to combine the results of a user defined frequency band is called *scale averaging*. The resulting power spectrum can be defined by :

$$\|W_n^{j_1,j_2}\|^2 = \frac{\delta j \delta t}{C_\delta} \sum_{j=j_1}^{j=j_2} \frac{\|W_n(s_j)\|^2}{s_j} \tag{58}$$

Where

$j_1$ is the index of the first scale to be averaged.

$j_2$ is the index of the last scale to be averaged.

$\|W_n^{j_1,j_2}\|^2$ is the averaged in scale wavelet.

The scale averaging technique was applied to the competition time series. The wavelet function used was the Morlet function with $\omega_0 = 6$. In accordance with the spectrum analysis, the scales corresponding to periods between 11 and 60 were averaged. The results can be seen on Figure 20.

On Figure 20, the peaks in the monocomponent spectrum corresponds to the *switches* in the time series from one oscillation mode to another. Although these events occurs abruptly, the time resolution of the wavelet transform tend to *spread* their effect. Moreover, since the scaling average was made on the 10-60 period interval, it can be considered that the spectrum is significant only in that band and that the features occurring elsewhere are artifacts.

Figure 21 represents the monocomponent spectrum analysis using a better time resolution wavelet function, Paul's wavelet. The switching is better localized than with Morlet's wavelet and the modulation can be quantitatively estimated. The switch correspond to the rising front while the modulation is the decreasing one.

We will see in section 5 that the same technique of scale averaging can be used to further extract information.

Finally, when comparisons are made with the time-frequency analysis performed on the competition time series in section 2.4, it is clear that the wavelet provide with a better resolution both in time and frequency domain. Furthermore, the flexibility gained by the arbitrary defined scales enable to perform target analysis in the frequency domain. The drawbacks generated by the redundancies can be easily addressed as we will see now in the section concerning the signal preprocessing with wavelets.

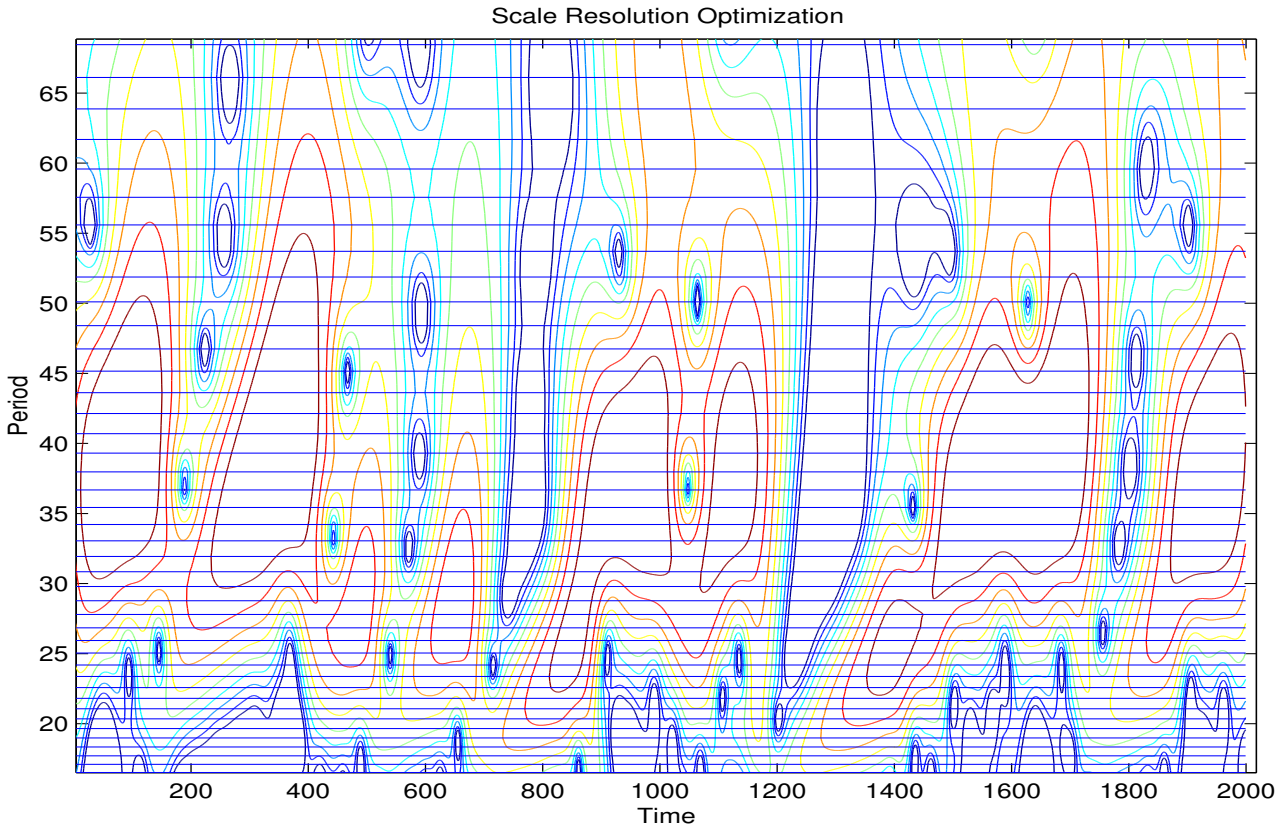


Figure 18: Optimization of the Resolution

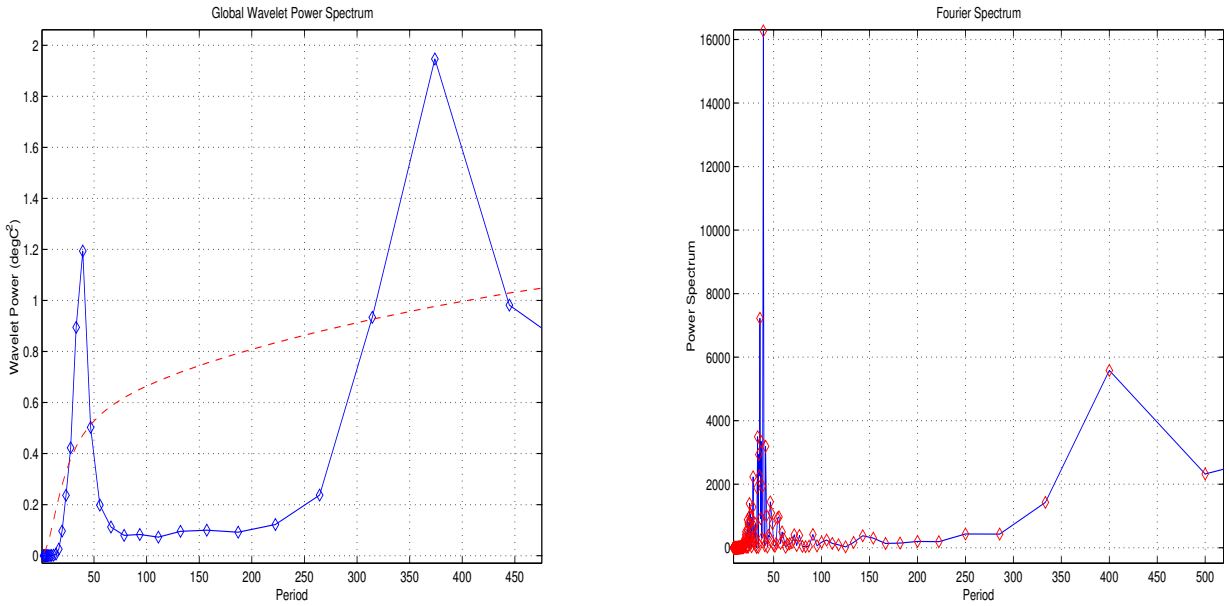


Figure 19: Global Power Spectrum in Wavelet and Fourier Spaces

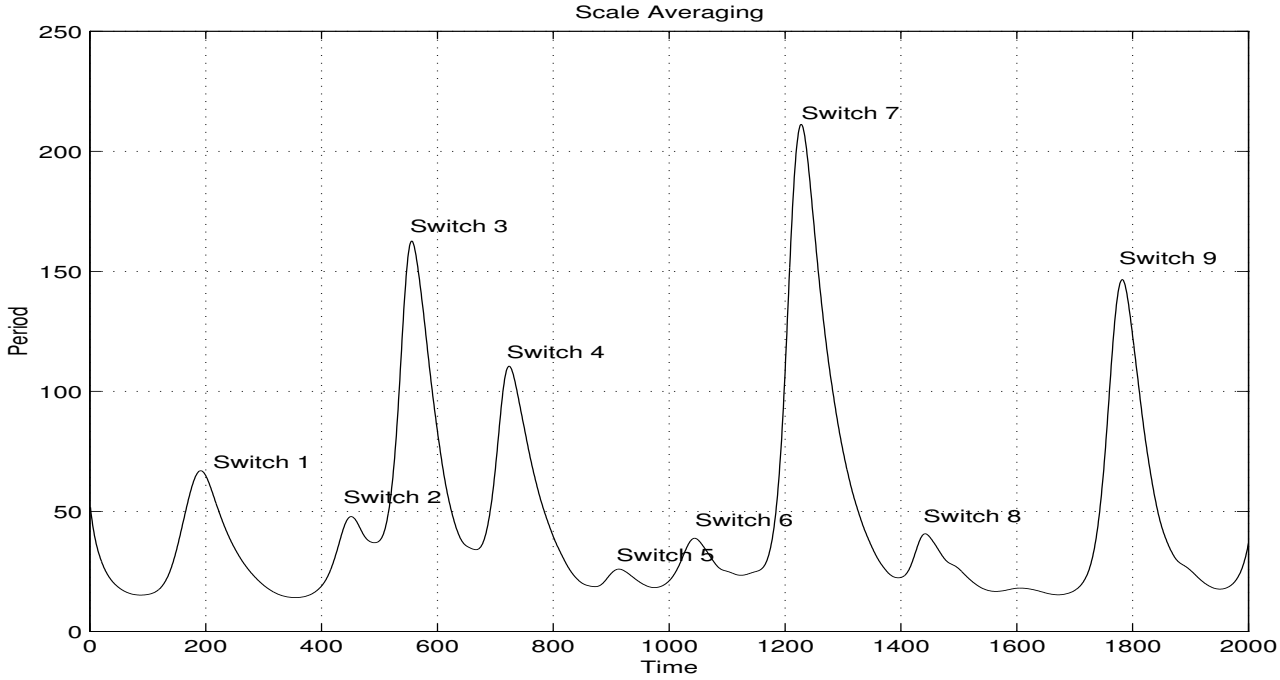


Figure 20: Monocomponent Spectrum

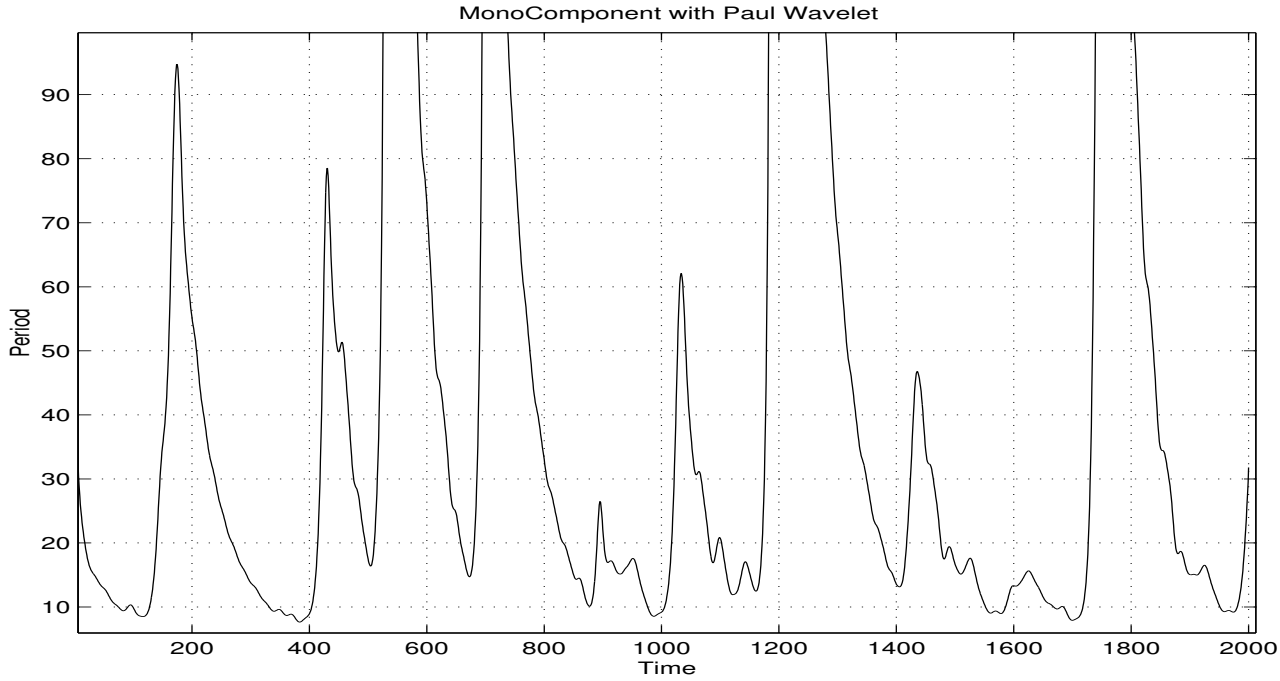


Figure 21: Monocomponent Spectrum with Paul Wavelet

# 5 Wavelets and Preprocessing

## 5.1 Introduction

In section 4, we have seen how the wavelet transform, and particularly the wavelet power spectrum, can be used to perform a Time-Frequency analysis of a time series. Although wavelets improve greatly the analysis, wavelets are mostly used for other characteristics. As a matter of fact, their are particularly adapted to signal processing or preprocessing. As a consequence, wavelets are mostly studied and developed in the communication field.

They are used to expand the signals (uni or bidimensional) into distinct features ordered by scales. These features are then sorted by their relevance to the signals. Some features are associated with noise and are therefore removed while the others defines *resolutions* of the signal. A set of waves expansions is defined to be a compressed equivalent of the original signal. The level of compression is inverse proportional to the number of scales transmitted through the channels. The wavelet compression, also called fractal compression because of the use of recurrent wavelet function, is one of the most promising compression technique available.

The purpose of that study is not to examine the issues proper to the communication field but rather to examine how these techniques can be applied to time series modelling. The main foreseable features are the possibility to perform feature extraction (source extraction ), noise cancelling and more importantly multi-scale modelling. Before examining the methods for selecting the components, we first examine the issues involved in time series expansion with wavelets.

## 5.2 Wavelet Expansion

The basis of wavelet signal processing is the expansion of the signal on a set of waves associated to scales or frequency band. In that respect, wavelet signal processing is often considered as operating a filter bank, each wave being the results of a band pass filter. The corresponding bandwidth and the central frequency depends on the wavelet function and its parameters.

We have already mentionned that for time series analysis, nonorthogonal wavelets were the most suitable. The main draw back in the redundancies occurring and materialized by larger number of waves associated with overlapping filters.

By breaking down or expanding the time series, it is possible to extract frequency related feature. For example, oscillatory modes and the different type of noise have different *signatures* that can be isolated. Similarly, several feature can be overlap in the time series and if they have different signatures, it may be possible to separate them. Another factor due to the use of nonorthogonal wavelets is that the wavelet basis is not orthogonal so that it will not generated an evenly distributed input space for implementing gradient based parameter optimization for a model of the time series.

In order to illustrate the wavelet expansion, we can examine the expansion of the competition time series. As for the analysis, it is important to determine the optimal wavelet and scale functions. We start with the Morlet wavelet function with a scale increment $\delta j = 0.5$. Figure 22 represents the wavelet expansion into 20 components.

The expansion was performed using a dyadic grid. Therefore, each wavelet or component correspond to a *central frequency* (period). Here, from top to bottom, the scales and periods are given by Table 4.

| Name | Index | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Scales | 2 | 2.82 | 4 | 5.64 | 8 | 11.3 | 16 | 22.6 | 32 | 45 | 64 | 90.5 | 128 | 181 | 256 | 362 | 512 | 724 | 1024 | 1448 |
| Period | 1.97 | 2.79 | 3.95 | 5.58 | 7.9 | 11.1 | 15.8 | 22.34 | 31.6 | 44.6 | 63.2 | 89.39 | 126.4 | 178.8 | 252.8 | 357.5 | 505 | 715 | 1013 | 1430 |

Table 4: Scales and Periods of the Expansion

On Figure 22, the wavelet expansions are called W1, W2, ... W20. The first four wavelets associated with small scales (high frequency) do not have an amplitude significant compared to other components. Moreover, they do not exhibit particular feature. As a consequence, they will be considered as noise and will be removed from the input space. Form W5, the wavelet expansion becomes more significant, as already shown by the wavelet time series analysis. Most interesting of all is the fact that the different features are not in phase, a consequence of the modulation pointed out during the analysis. From W12 to W20, the wavelet expansion becomes more regular in time and probably associated with a longer period feature.

Once the expansion is available, it is possible to carry out the processing itself. This will involve the noise cancellation by removing some scales, the multiresolution construction by scale averaging technique and finally the local orthogonalization using principal component analysis.

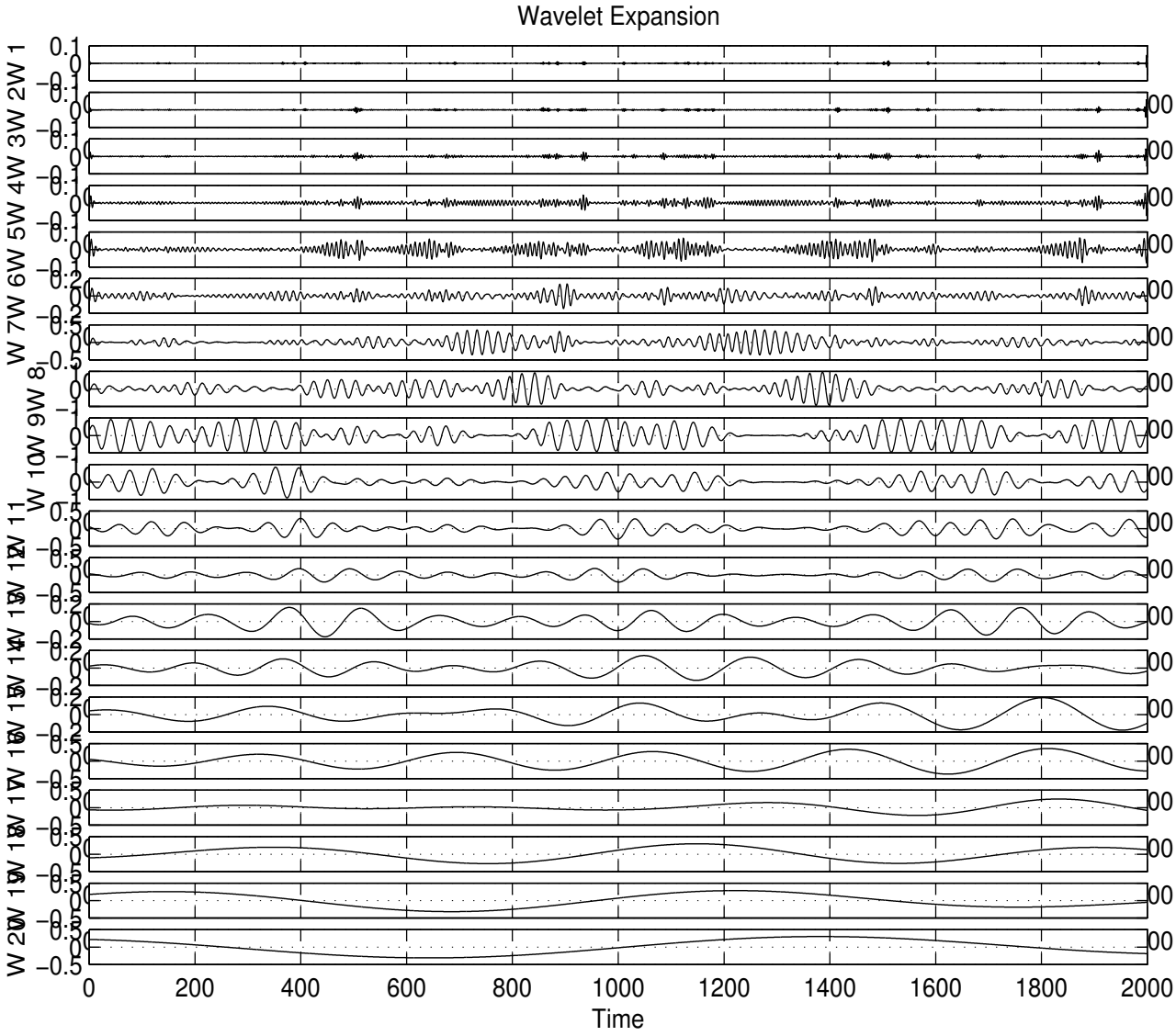


Figure 22: Wavelet Expansion.

## 5.3 Noise Filtering

The measurement noise characteristic is usually high frequency. To a more general definition of the noise, we will see it as a non-significant or even better non-causal feature of the input space. The wavelet expansion enable the frequency clustering and removal. Therefore, it is possible to remove some high frequency noise by removing a set of small scales.

To determine which scale can be removed without affecting the overall coherence of the time series, we can examine the variance of each wavelet decomposition on Figure 23.

Figure 23 indicates that the main scales have their index between 6 and 14, i.e. periods between 11 and 178.

To illustrate that simple filtering technique, we can examine the reconstruction of the wavelet expansion without the first 4 wavelets on Figure 24.

The cumulative variance explained by the first 4 wavelets is only 0.2 % of the overall variance. Therefore, these scales can be removed without loss for the time series. This seem to indicate that the competition time series had already undergone a noise cancellation process since higher noise level as normally expected.

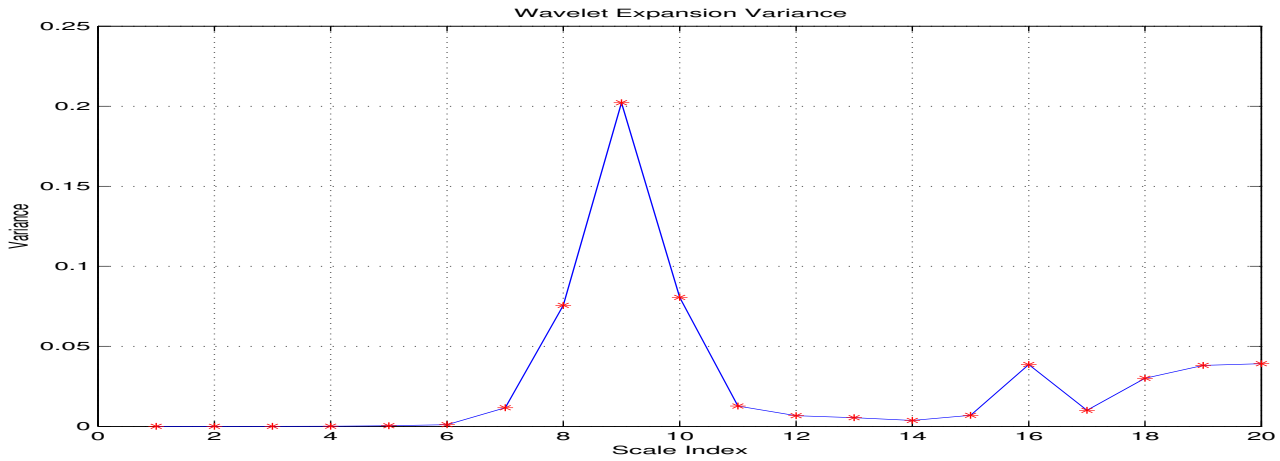


Figure 23: Wavelet Variance

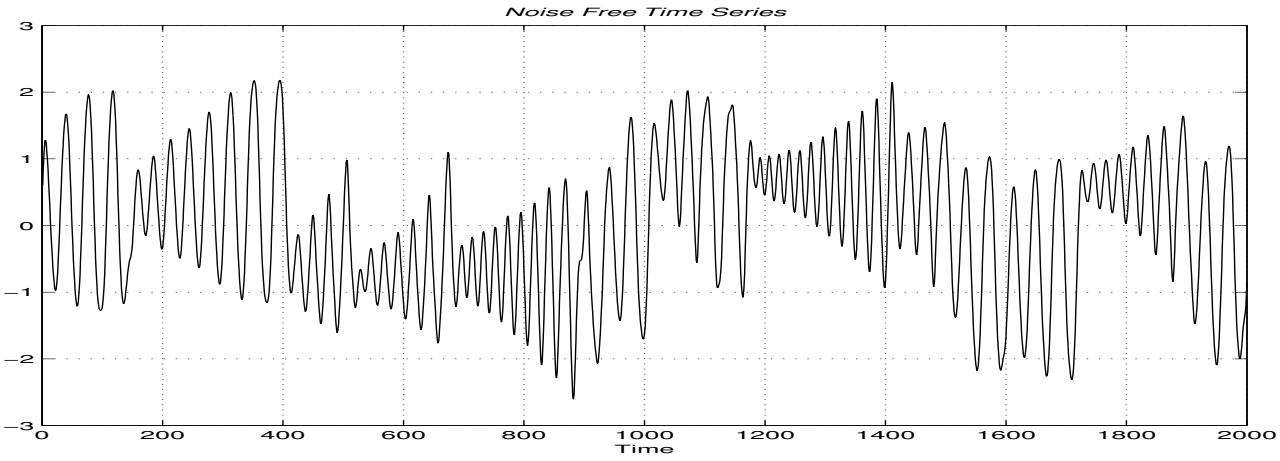


Figure 24: Filtered Time Series.

## 5.4 Frequency Clustering

When the filtering is completed, it is possible to examine the most interesting feature enabled by the wavelet expansion, the multi-scale modelling based on *Frequency Clustering*. We will define overlapping frequency clusters composed of wavelet expansions and perform a scale averaging.

This can be useful for non-oscillatory feature or ultra-large features that are extensively decomposed on a set of scales. To illustrate the gain of such clustering, we will examine the *large scale cluster*, including scales W12 to W20.

Figure 25 represents the averaged wavelet obtained on the last 8 scales and corresponding to the largest scales. This very interesting component enable to *extract* a critical information : The steady values from which the oscillation are modulated. The four steady state values from which the time series switches are -2, -1, 0.5 and +1.75. Using that information, it could be possible to synthesize a hybrid model made of the four steady state values and a modulated oscillator. However, if the values of the steady state can be well estimates, their occurrence and eventually the cause remains elusive.

To determine coherent cluster on other scales, it is required to have a quantitative criterion. In the case of the competition time series, the modulation of the oscillation tend to spread the oscillations on different scales, making them appearing as distinct features (or sources).

For the determination of the intermediate scales, we will use the cross-correlation factor between the wavelets.

Figures 26 represents the cross-correlation function between the wavelets. On the right graph, the inner area, the 95 % correlation area, indicates a division in the wavelets. Wavelets W1 to W10 seem to be correlated to two adjoint wavelets. Wavelet W11 is not highly correlated to either W10 or W12,

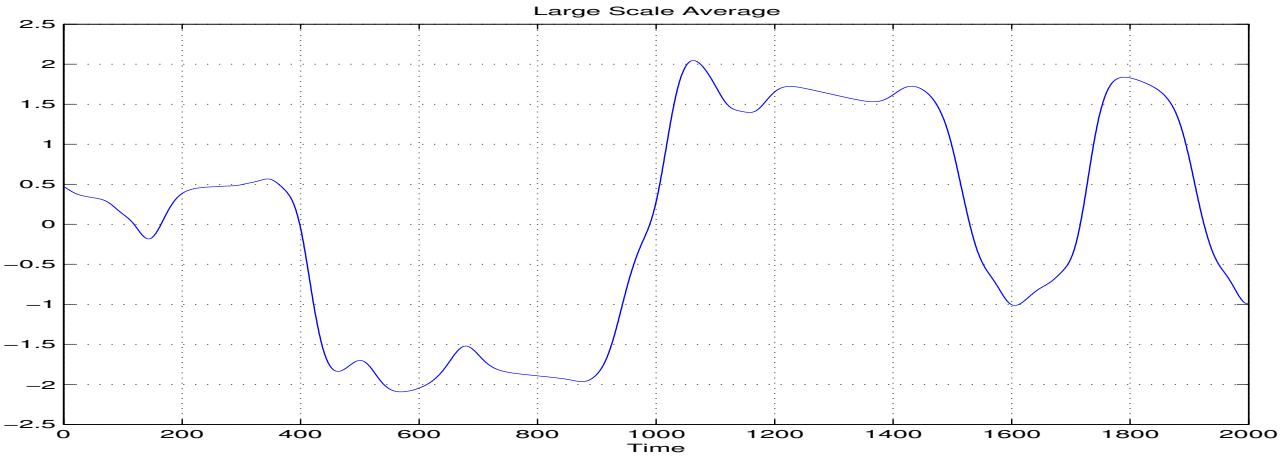


Figure 25: Scale Averaging on Large Scales Cluster.

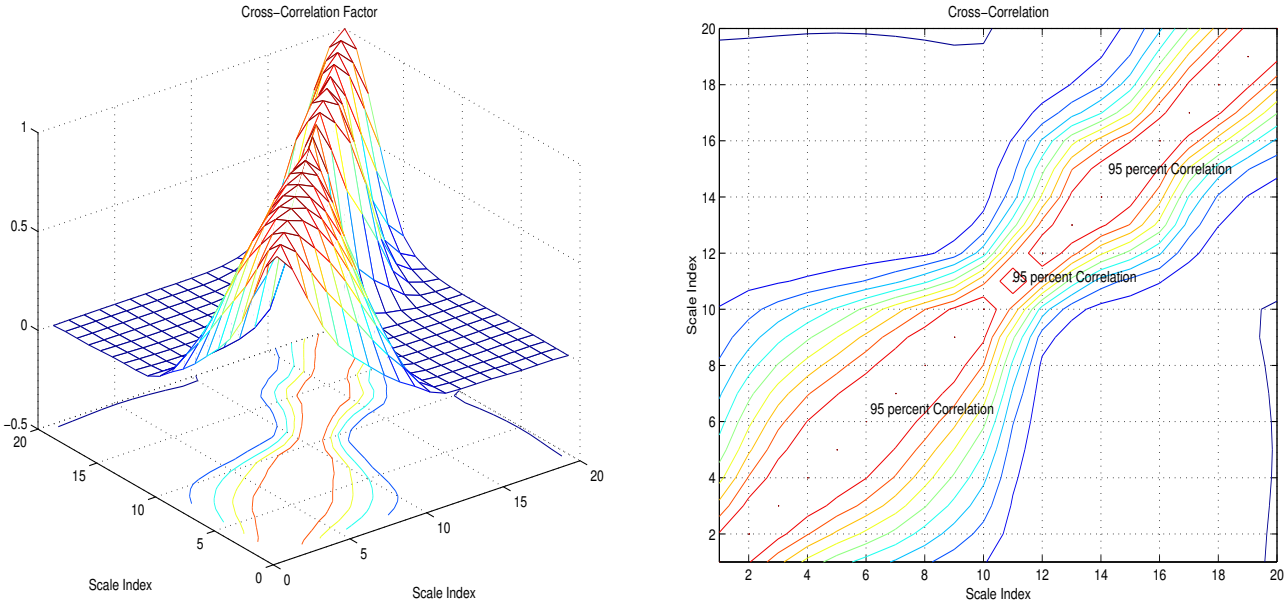


Figure 26: Cross-Correlation Analysis on the Wavelet Expansions

indicating a pivot role.

In the lower area, the maximum cross-correlation occurs around W6. The correlation was expected as we used nonorthogonal wavelet functions. Therefore, if we want to extract distinct feature in the low frequency area, we need to built orthogonal features. This will be achieved using an optimal linear combination based on Singular Value Decomposition. More precisely, we will use the *Principal Component Analysis* [21].

## 5.5 Orthogonalization

The classical statistical analysis methods aims at determining correlation relationships between the different input candidates and the outputs. The results are indications on the most suitable inputs to be incorporated in a model. If these *passive*, or observation variable directed methods can be useful to understand the system behaviour, they have limited effects on the system identification performance. It is therefore possible to use *active* methods that will transform the input spaces to make it more suitable for the numerical system modelling.

One of these methods is the so-called *Principal Component Analysis* (PCA). It is a multivariate analysis technique that was first introduced by Pearson in 1901 in a biological context to recast linear regression analysis in a new form. Later, the method have been improved by Karhunen and Loève in the probability theory field.

It is a statistics based methods that aims to find new variables which are linear combination of the observed variables so they have maximum variations and are orthogonal. This method can be used either to re-arranged the input data without lost of information leading to *orthogonal rotation*, or can also be used to perform a dimensionality reduction. Because this methods and its derivatives are used in various fields, the denominations and the interpretations of the transformations vary a lot. For example, the orthogonal rotation can be interpreted as a feature extraction in the field of multivariable data analysis and the dimensionality reduction can be called compression in the communication field.

The goal of this study is system identification. An any identification approach, many similar problems have to be overcame. Among these, the structure and the size of the information used to perform the identification is crucial. Because most of the methods are numerically based, the computation problem can be tackled only if the number and the structure of the information to process is optimal.

In linear identification, this problem is corollary to the choice of the number of inputs and then of the parameters. With neural networks identification the problem is to determine the optimal set of inputs that will enable the network to converge quickly as possible to a global minima of a multi dimensional error surface. The higher the dimensional, the higher the probability to end in a local minima. Whatever the modelling approach chosen, the first concern must be the determination of the optimal set of inputs. This can be the choice of the inputs, the size and the number of the sets, or even the choice of the transformation of the input sets. In this section, we will examine the choice of this transformation.

In any identification approach, it is always important to provide all and only the relevant information. Any redundant information such as correlated but not causal information or noise will overload the numerical methods and will downgrade the identification performance. The elements of the principal component analysis can effectively used to understand the behavior of the system by extracting and exhibiting its main features but it can also improve the identification procedures using the principal component loadings.

The principal components loadings can be interpreted as the transformed inputs in which the main features have been extracted. If all the loadings are used, the information within the data sets is preserved. This means that features such as outlier data or noise are also fed into the identification process. The only gain is that by dividing these features in distinct variables, the identification process can treat them more appropriately. This procedure is designated as *orthogonal rotation* or *data expansion*.

A more dramatic transformation would be to reduce the input dimensional by truncation. In that case, only a subset of principal components loadings is preserved. The rational is that this enable to perform a *negative feature filtering*, freeing the input signals from noise and outliers. Moreover, because of the orthogonal based used, the reduction of the number of loadings to be used lead to reduce the number of input variables and then lighten the task of the identification process. The approach has different denomination according to the field of application. It can be designated as *subspace reduction*, *subspace decomposition*, *feature extraction* or *signal compression*. However prospectful this method seems, some important issues still have to be addressed. For example, it is crucial to distinguish the *positive features*, i.e. the one that do represent a feature of general behavior of the system, from the *negative features* that corresponds to particularities of the data set, noise and outliers. Unfortunately, there are no definitive answer to that problem so that trade off have to be made. Moreover, the use of subspace reduction methods may also have effects on the generalization properties of the models identified.

To the critical question of the number of principal components loadings to be used, the associated eigenvalues can be used and interpreted as an index for the *proportion of information carried* by the principal components. Some general rules can be used to determine a god trade off in the selection of the number of principal components :

- Use the subset of principal components which are responsible of at least 80 % of the input data variance.
- Exclude principal components whose eigenvalues are less than the average of the eigenvalues.
- Use the scree graph to find a *bend* or an inflexion point.
- Test the significance of the retained subset of principal components.

The different rules can be used as stand alone or can be interfered but none has any theoretical better ground. However, we can attempt to explain the rational behind these rules.

The eigenvalues are calculated on the variance covariance data matrix. For that reason, their mean is 1. It is possible to consider the mean as a decision threshold for dimensionality reduction. The small eigenvalues can be associated to either noise, outliers or small variance information. To reconstruct the data without the small eigenvalued components is similar to noise filtering.

The *proportion of variance explained* is an index of the conservation of the integrity of the data. It measure how the overall information, expressed by the variance, is divided into the different principal components loadings. The cumulative proportion can be used to determine the best compromise between lost of information, a quantitative lost rather than qualitative lost, and the reduction of the input space dimension. However, we have to remember that quantitative information does not mean qualitative information and it is possible that the principal component loadings associated with the smaller eigenvalues carries valuable qualitative information. The cumulative proportion of variance will then be an index of the degradation of the data.

The two last approaches are not clearly defined since they rely on case by case determination. Although sensible, they are difficult to express quantitatively. However, in order to compare the efficiency of the different rules on our case problem, we will evaluate the efficiency of the principal components using a model based criterion. These rules are not the only possible approach to answer the issue of the number of principal components, other statistical methods such as the study of the distribution of the eigenvalues or the so-called factor analysis can be used but they will not give definitive answers either.

The three first rules indicate clearly that the three first principal components are the most suitable to be preserved for subspace reduction. To get a quantitative criterion, we will use the forecasting performance of linear models using subsets of the principal components as inputs.

The PCA technique can be applied to reconstructed wavelets especially when nonorthogonal wavelets functions are used. When used in a *scale averaged* context for multi-scale modelling, the PCA is applied only to a subset of waves. Figure 27 represents the first 6 Principal Components (PC) obtained on the 6 selected waves.

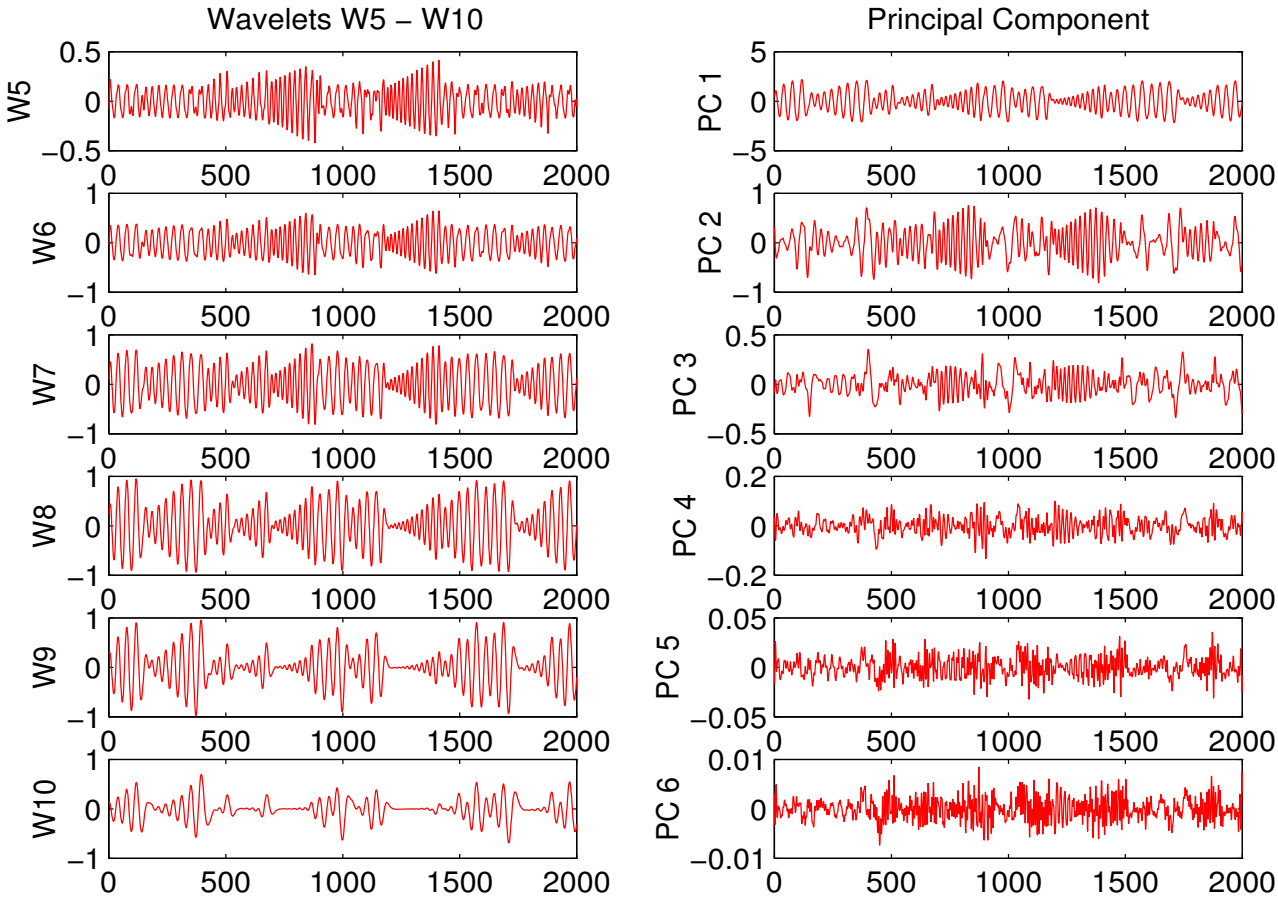


Figure 27: Principal Component Analysis of W5-W10 Input Space

The issue of the selection of the principal component used to implement the modelling is addressed by a simple variance analysis as the threshold is clearly identified. Figure 28 shows the cumulated variance explained by the first PC.

According to the variance analysis, the two first principal component contribute for more than 99 % of the W5-W10 space variance. Therefore, when implementing into a modelling process, these will be the

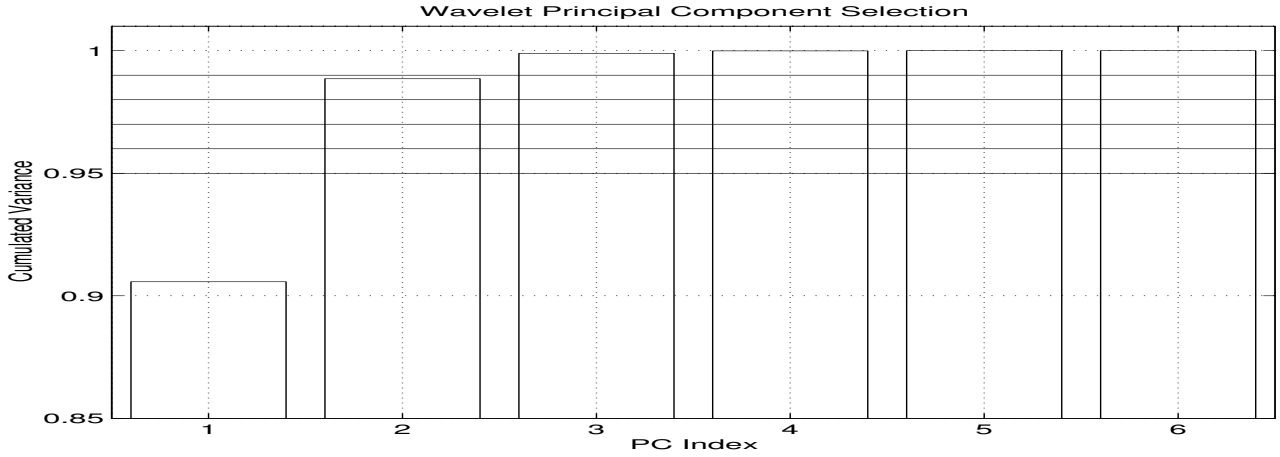


Figure 28: Principal Component Analysis

two inputs for the scaled model.

Although PCA is considered to be a time domain technique, it does not downgrade the performance of Time-Frequency techniques such as wavelet transform. On the contrary, by enabling a multicriteria preprocessing, it is possible to optimize the input space by mean of optimal scale and orthogonal processing.

In order to find out about the performance of the use in conjunction of wavelets and PCA, we can examine a comparison with the preprocessing using only PCA.

## 5.6 Feature Extraction Techniques Comparison

We have seen that wavelets expansions could be combine on a set of scales to produce a signal partly freed from redundancies issued from the nonorthogonal wavelets and the arbitrary discrete scaling. Therefore, the *scale averaging* technique can be seen a refinement of the feature extraction process in the frequency domain. In previous studies [5, 21] we used only PCA to perform the feature extraction.

To evaluate the advantage of the wavelet preprocessing, we first examine the input space issued from the method presented in sections 5.4 and 5.5.

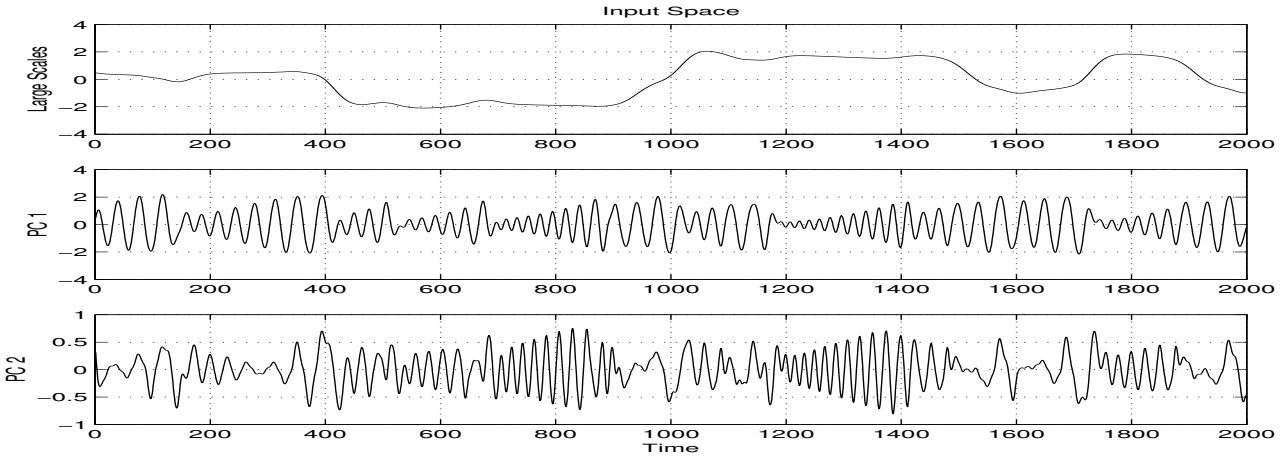


Figure 29: Preprocessed Input Space

In [5], where the competition time series modelling was attempted, PCA was used as the only preprocessing method. The main problem was the multi-scale feature encountered in the time series. As a matter of fact, a first input space made of regressive input was selected. In order to extract the features related to the oscillation mode, a relatively large regression order, 150, had to be used. If it provided a good extraction of that feature, very similar to the first input of the wavelet-PCA extraction technique,

the variance of the remaning signals was spread out on the remaining 149 principal components. In order to incorporate a 95 % cumulated variance, a much large number of components had to be used, up to 25. Of course, the poor extraction did not performed well with the higher frequency features. The modulation feature was faded so that the input space, although orthogonal was not as regularized as could be.

It was then decided to go for an in-between solution, to use a lower regression order, at the risk of loosing the long period, the steady state values. A regression order of 40 was used and the PC containing 95 % of the initial input space variance are shown on Figure 30.

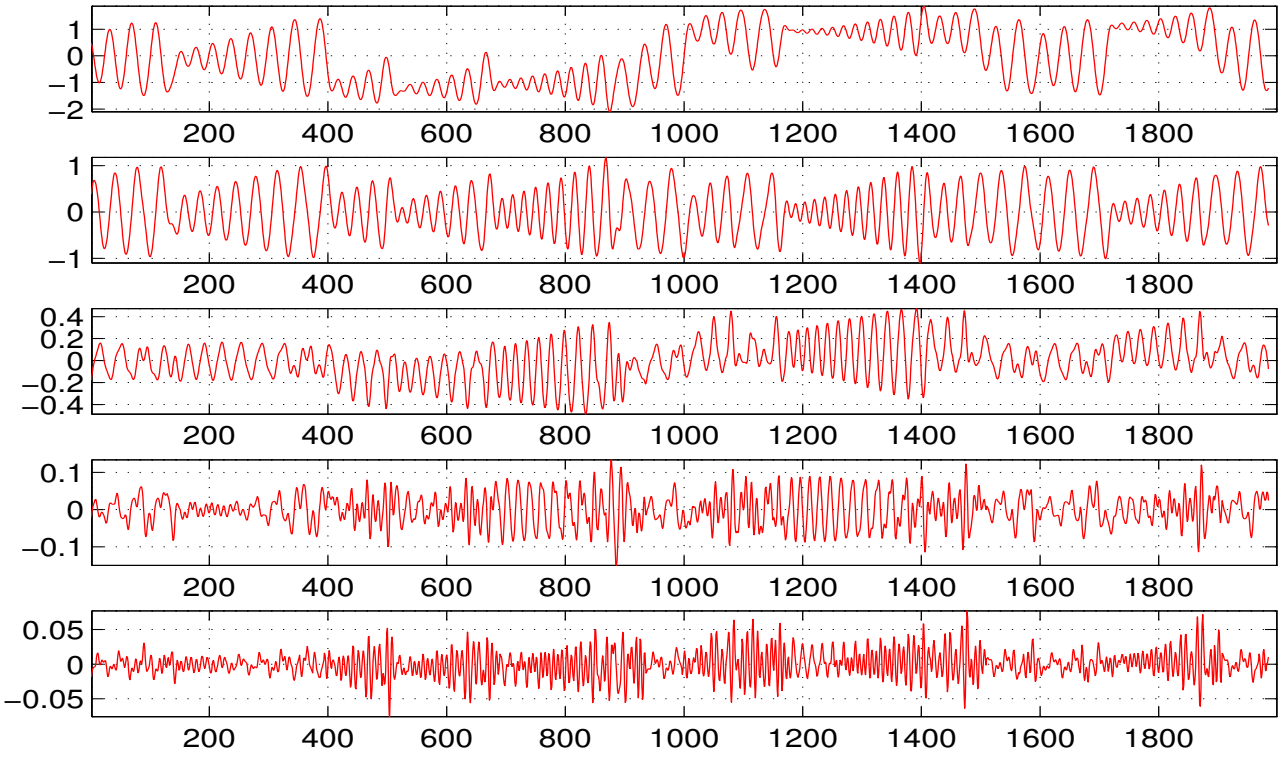


Figure 30: PCA Preprocessing

On Figure 27, we can see that if the 95 % variance is spread out only on 5 components, the extraction of the different frequencies is not well performed. The first component is obviously related to the long period feature but is modulated by a higher frequency feature. The comparisons with the Wavelet-PCA extraction and multi-scaling techniques are striking. The composite technique enable to perform scale extraction thanks to wavelets while the use of averaged scaling and PCA enable to keep the input space compact and well regularized.

In that respect, if the wavelet-pca input space was used in a modelling processing, the performance of that model might be better than the one obtained in [5] although that application might be left to further works.

We have seen how the Wavelet-PCA preprocessing technique performed well with the competition time series. We can now examine the results of the same technique on another chaotic time series, the Sunspot time series.

# 6 Application Example : Sunspot Time Series

## 6.1 Introduction

The sunspot series is composed of the yearly mean of an arbitrarily defined sunspot number $R_I$ that expresses the number of spots or group of spots on the surface of the sun. Although such phenomena have been observed since 1700, there is still no physical explanation. The resulting time series is thought to be chaotic with a pseudoperiod of 12 years with some dynamical system behaviour. It has been used as a benchmark problem for various time series prediction methods [21].

Modelling of the sunspot series must observe several constrains. The first is the very limited number of data with regard to the 12 years pseudoperiod. Most of the statistical analysis and identification methods rely on statistical coherence of the data set, i.e. data sets on which it is possible to carry out generalization assessments. Another problem associated the limited number of data and the lack of understanding of the physical phenomena is the question of the time invariance of the sun considered as a dynamical system. The reduced observation time scale, with regard to astronomical time scales and the important variance of the pseudoperiodic signals both in amplitude and frequency (the *cycle* varies between 6 and 14 years), increases the uncertainty on the nature of the system studied.

We will apply the techniques exposed in the previous sections to find out how the wavelet transform can be used to improve the analysis, the preprocessing and finally the sunspot time series modelling.

## 6.2 Analysis

The classical analysis was performed in [21] and the Fourier analysis indicated a pseudoperiodic time series of about 12 years period as already mentioned in section 6.1. Figure 31 represents the sunspot time series (upper graph), and the wavelet power spectrum.

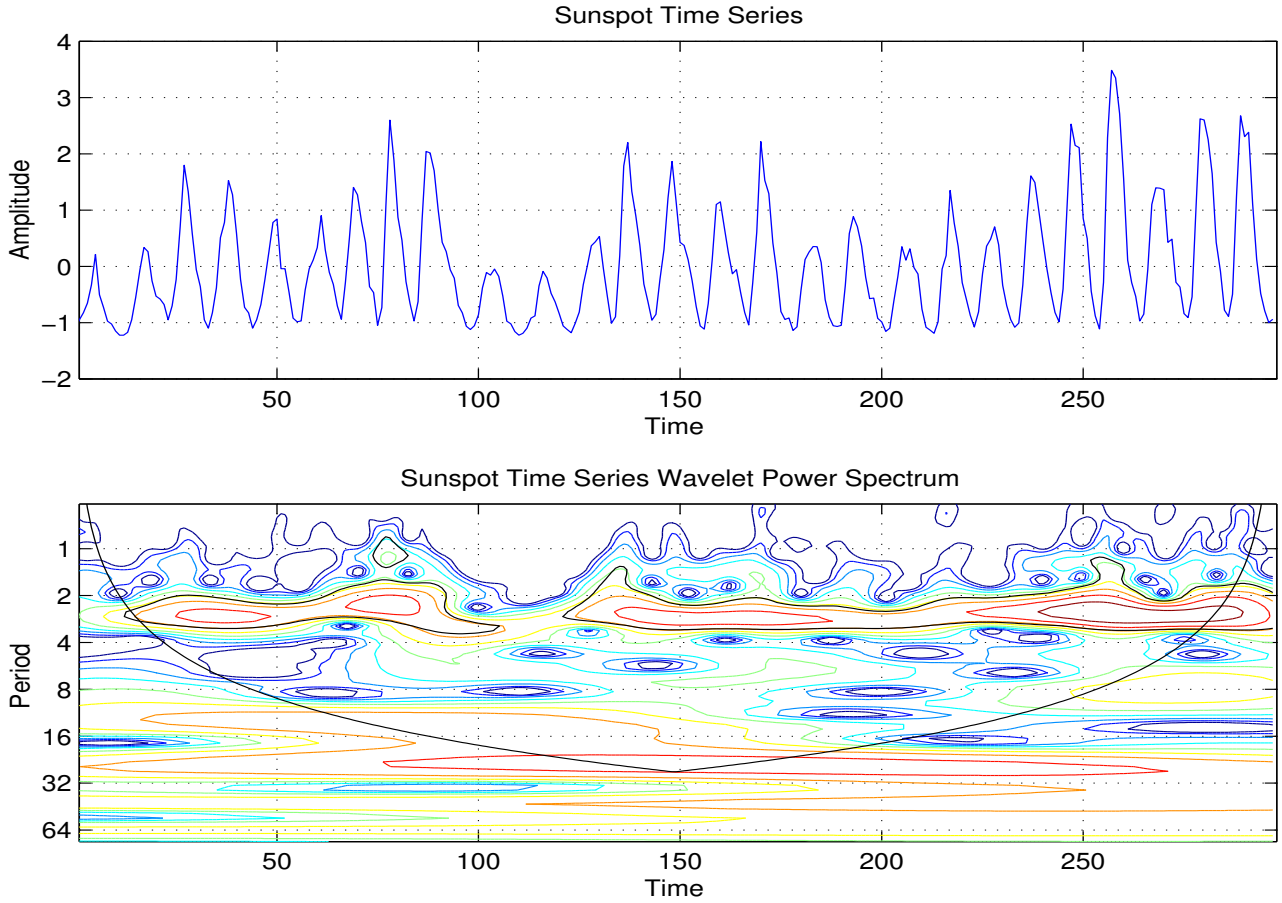


Figure 31: Sunspot Time Series and Local Wavelet Power Spectrum .

The wavelet function used to perform the spectrum analysis was the Morlet wavelet, chosen for its resolution trade-off characteristics. The arbitrary scales were expressed as a power of two, following

the definition given in Equation 49. The scale increment $\delta j$ was set to 0.5 so that the time series was expanded into 20 scales (see Table 4).

There are two main significant features, one situated around *period* = 12 and another one around 80, the *superperiod* feature. However, the cone of influence indicates that the limited data set make the identification of the latter uncertain. We can have a closer look at the 12 year period feature on Figure 32.

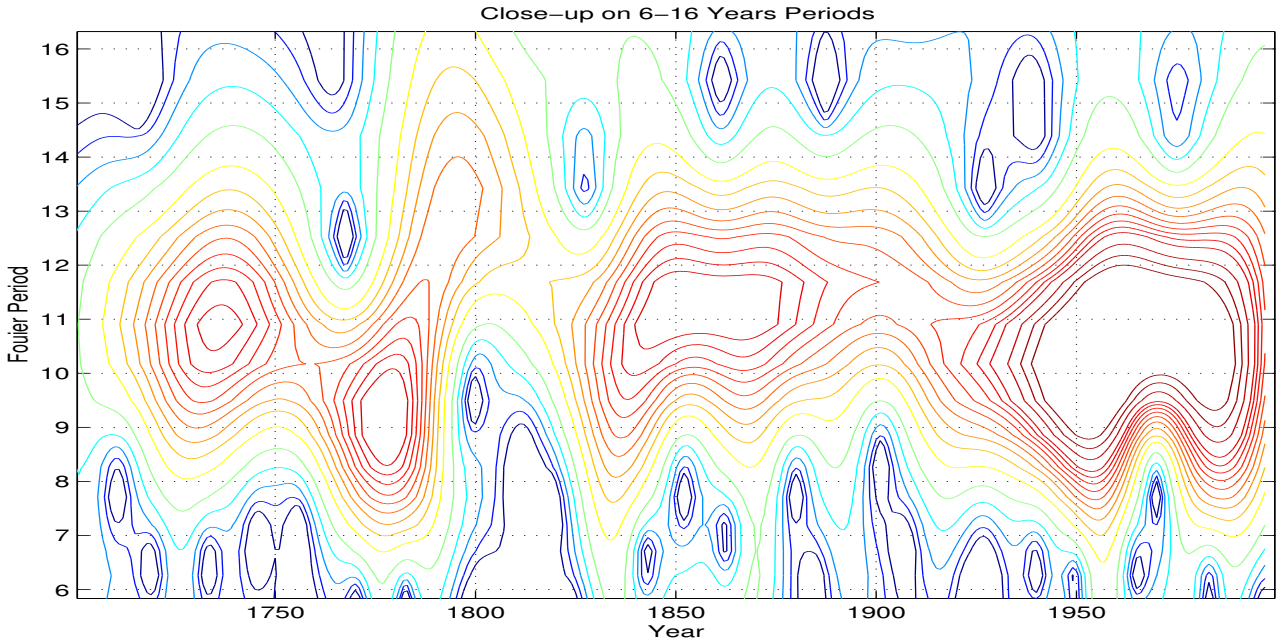


Figure 32: Local Wavelet Power Spectrum .

The resolution is not ideal since the feature peaks are interpolated from the wavelet expansion at periods 8.2, 11.7 and 16.5. For better frequency resolution, we set $\delta j = 0.1$.

The *stability* of the period is ensured for measurements posterior to 1830, a time when more rigorous and targeted experiment were implemented. The combination of the about-to-be-invented photography as well as the combination of observation from different observatories produced certainly better estimates. The measurement improvement may explain the variation in frequency before 1830 although the large variation that occurred between 1790 and 1820 (almost no sunspots) must have some reality.

In order to make a more reliable quantitative estimate of the period of oscillation, we examine the global wavelet power spectrum as well as the Fourier power spectrum. These graph are on Figure 33.

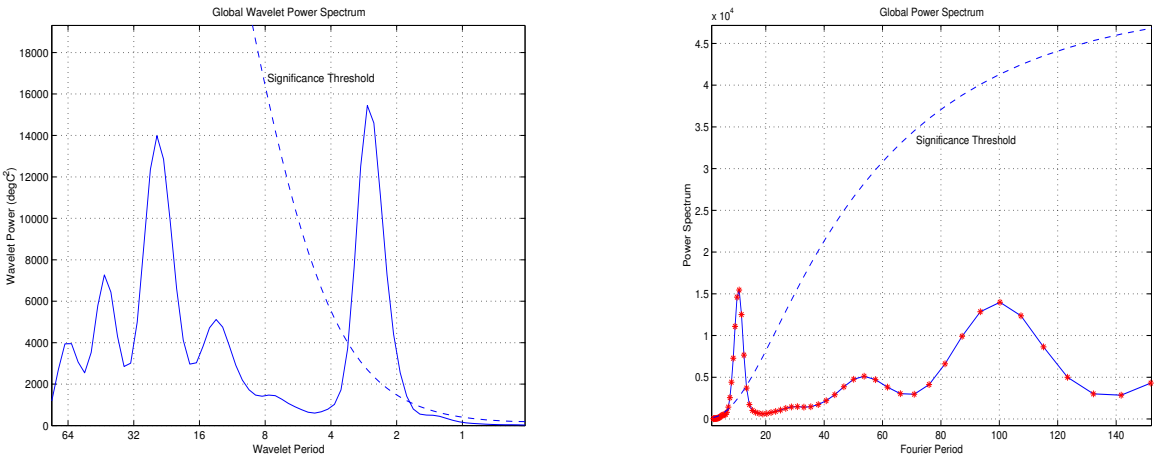


Figure 33: Global Power Spectrum in Wavelet and Fourier Spaces

The global power spectrum on Figure 33 both indicate that only the 12 years period feature can be identified with a satisfactory confidence. Therefore, later in the modelling, we will perform the denoising

of the time series and that superperiod feature will be considered as noise. These issues will be addressed in section 6.3 on wavelet preprocessing.

## 6.3 Preprocessing

### 6.3.1 Wavelet Expansion

Once the time series analysis is completed, it is possible to carry out targeted preprocessing techniques. We first implement the wavelet expansion on 29 scales in order to have a good frequency resolution and being able to extract precisely the useful feature. The 29 scales are obtained with a scale increment $dj = 0.25$.

We are mainly interested in the wavelets corresponding to periods between 6 and 17. This lead us to look at wavelet W7 (period=5.8 years) to wavelet W13 (period = 16.8 years).

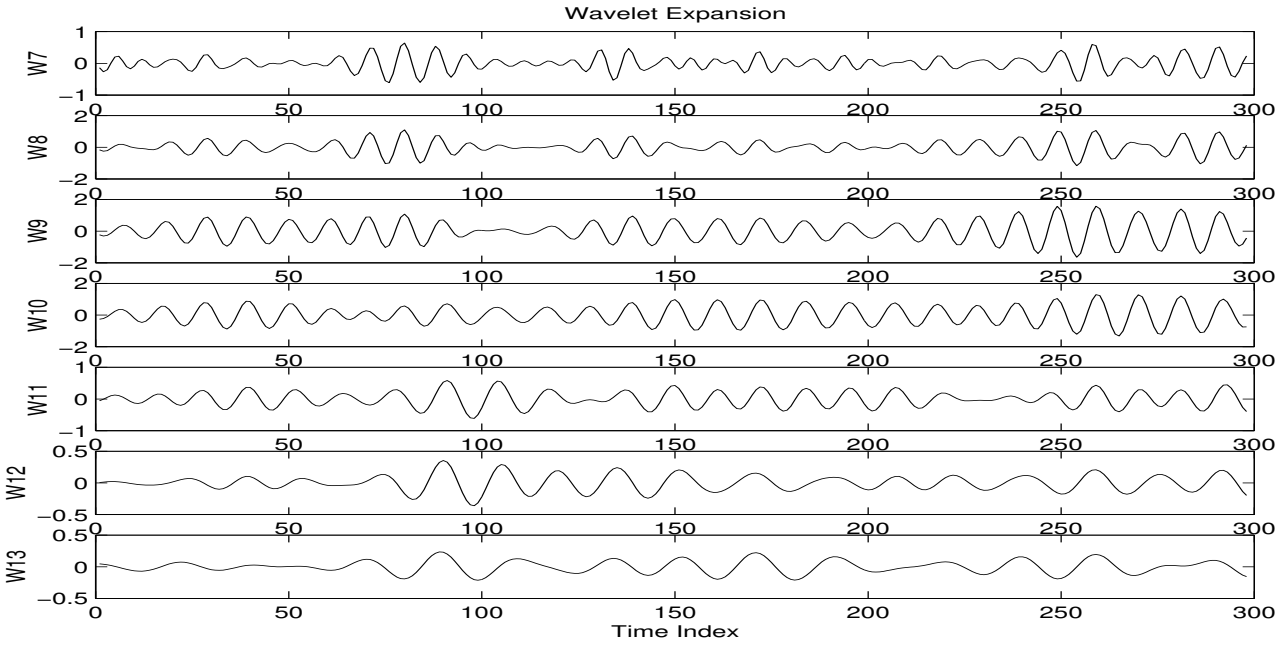


Figure 34: Wavelet Decomposition.

### 6.3.2 Filtering

We have selected the wavelets W7 to W13 to be the effective components we want to use in the modelling. This selection can be seen as a band-pass filtering where the period retained are in the interval 6 to 17 years. If we perform a scale averaging on that interval, we obtained the *filtered sunspot number* expressed on Figure 35.

Once the time series has been filtered and the undesired feature are extracted, it is possible to examine at the orthogonalization of the remaining wavelet.

### 6.3.3 Principal Components

The PCA is performed on the 7 wavelets represented on Figure 34. The resulting 7 components are expressed on Figure 36.

Figure 37 represents the cumulated variance explained by the successive principal components. The 97% variance explained is obtained for the first three components. Therefore, the input space will be composed of the first three components.

When we performed the preprocessing using only PCA, we obtained three principal components for 95 % variance explained [21]. These components, as well as the original time series are represented on Figure 38.

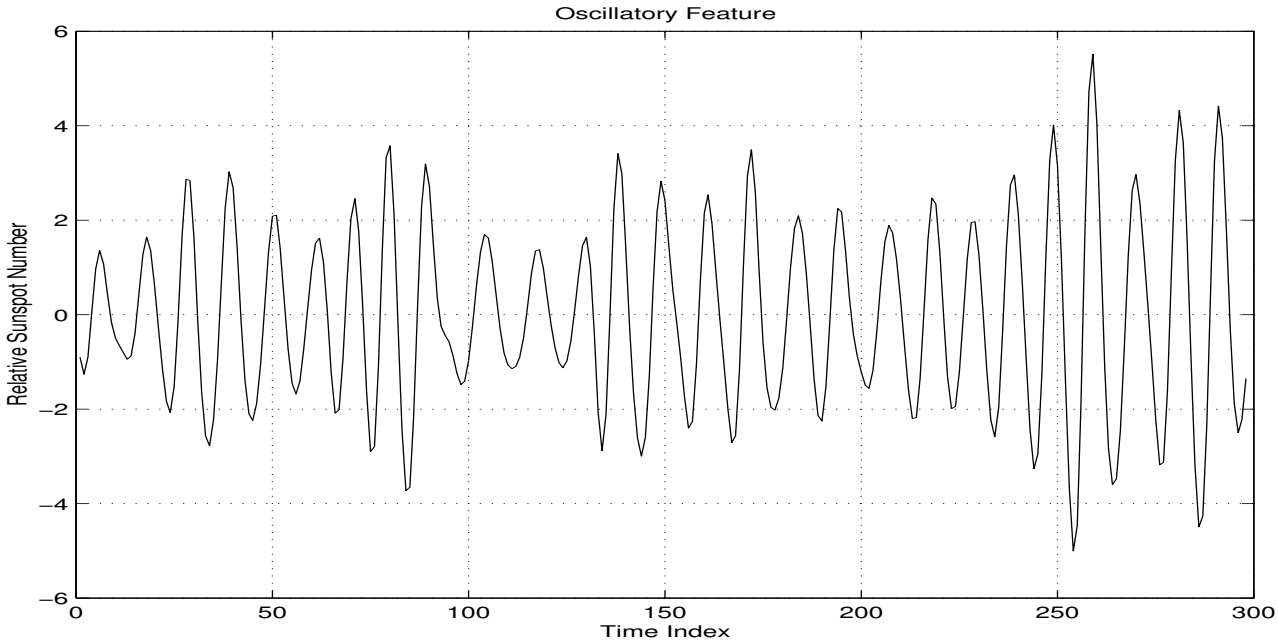


Figure 35: Filtered Time Series

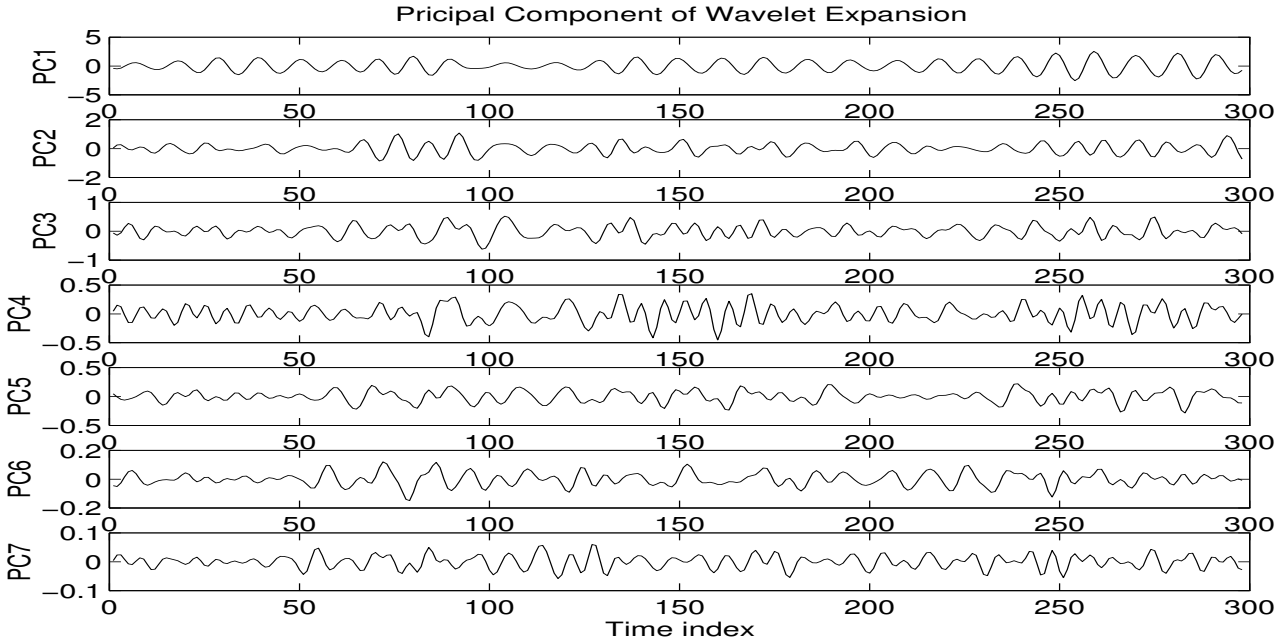


Figure 36: Principal Component of the Sunspot Time Series Expansion

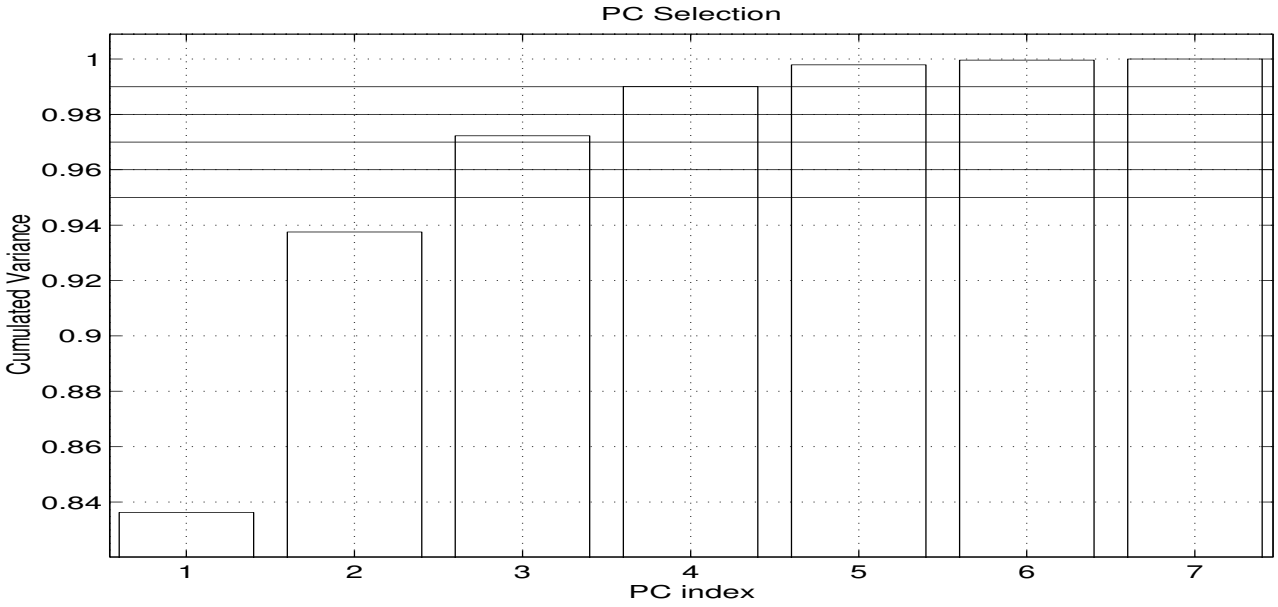


Figure 37: Statistical Analysis of the Principal Components.

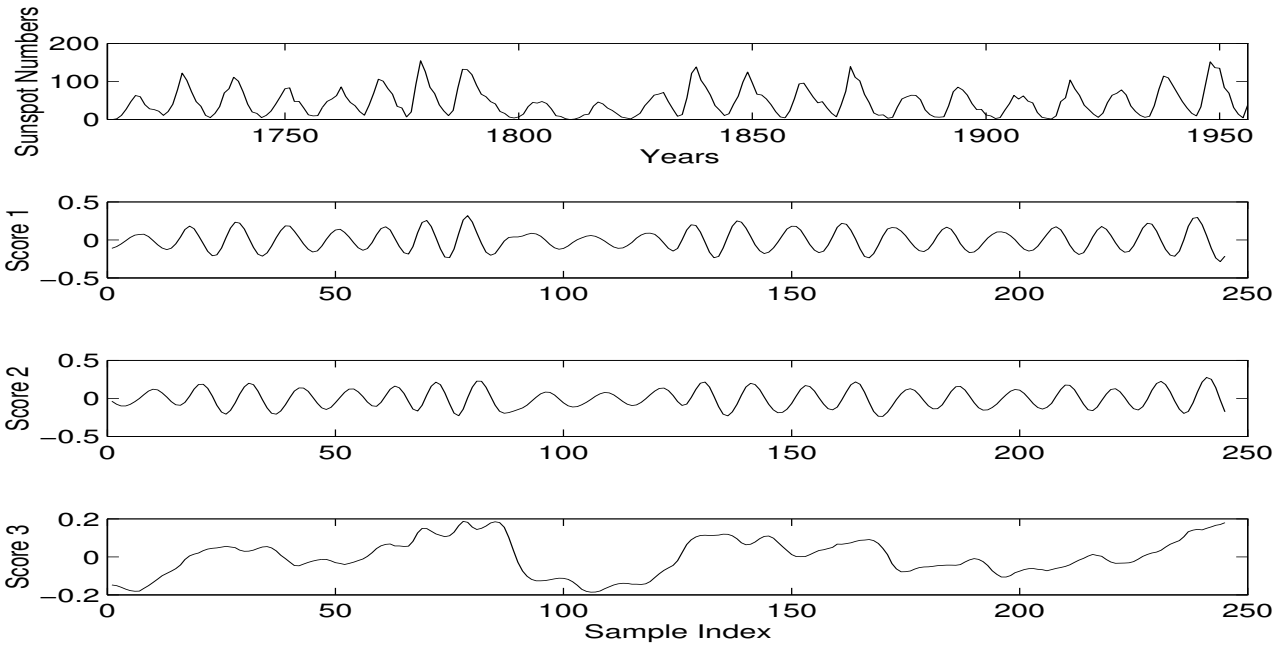


Figure 38: Input Space of Sunspot Time Series Modelling.

# 7 Toolboxes for Wavelets

## 7.1 Wavbox : The Mathworks Wavelet Toolbox

`http://www.wavbox.com/`

WavBox Software is the original MATLAB Wavelet Toolbox, the first available as free software in 1991, and the first available as commercial software in 1994.

A brief summary of WavBox Software follows below:

WavBox 4.5 - A Software Toolbox for Wavelet Transforms and Adaptive Wavelet Packet Decompositions with FirWav Filter Library

WavBox Software provides both a function library and a computing environment for wavelet transforms and adaptive wavelet packet decompositions. The software package contains 317 MATLAB m-files (approximately 1.4 MB uncompressed) implementing an extensive collection of these wavelet transforms, expansions, decompositions, wavelet filters, wavelets, and related functions. This function library performs multiresolution analyses of 1-D multichannel signals and 2-D images for arbitrary length and size data. MATLAB is a technical computing environment available from The MathWorks Inc.

WavBox 4.5 Software includes: continuous wavelet transform valid for all wavelets including the complex Morlet, real Gabor, Meyer, and Mexican hat wavelets; overscaled pyramid transforms, discrete wavelet transforms, and adaptive wavelet and cosine packet decompositions by best basis and matching pursuit (Mallat, Coifman, Wickerhauser, and other authors); satisficing search algorithms for the selection of near-best basis decompositions with either additive or non-additive information costs (Taswell); the systematized collection of FIR wavelet filters computable by spectral factorization of the Daubechies polynomial (Taswell); empirical tests for evaluating the parameters of multi-rate filter banks (Taswell); wavelet shrinkage denoising methods (Donoho and Johnstone); the fast wavelet based numerical methods (Beylkin et al) for the implementation of fast matrix multiplication and solution of linear systems in the wavelet domain. These latter methods can be used to implement pseudo-differential and integral operators.

Various choices of filter classes (orthogonal, biorthogonal, etc), filter families (least-asymmetric, most-symmetric, etc), and convolution versions (interval, circular, extended, etc) exist for each transform, expansion, and decomposition. The software has been designed for efficient automated computation, interactive exploratory data analysis, and pedagogy. Essential features of the design include: perfect reconstruction for multiresolution decomposition of data of arbitrary size not restricted to powers of 2; both command line and graphical user interfaces with a comprehensive set of plots and visual displays; an object property expert system with artificial intelligence for configuring valid property combinations; heirarchical modules and switch-driven function suites; vector-filter and matrix-operator implementations of convolutions; extensibility for the inclusion of other wavelet filters, convolution versions, and transforms; optional arguments with built-in defaults for most m-files; and extensive on-line help and self-running tutorial demos.

## 7.2 Rice Wavelet Toolbox

`http://www-dsp.rice.edu/software/RWT/`

Rice-Wlet-Tools (RWT) is a collection of MATLAB M-files and MEX-files implementing wavelet and filter bank design and analysis. In addition to the design tools the toolbox provides code for wavelet applications for both 1D and 2D denoising as well as code for processing of SAR images. The toolbox (version 1.1) was first officially released on Aug 30, 1993 but have since undergone several significant changes and enhancements.

## 7.3 Wavelab

WaveLab .701 is a library of MATLAB routines for wavelet analysis, wavelet-packet analysis, cosine-packet analysis and matching pursuit. The library is available free of charge over the Internet. Versions are provided for Macintosh, UNIX and Windows machines.

WaveLab has been used in teaching courses in adapted wavelet analysis at Stanford and at Berkeley. It is the basis for wavelet research by the authors, and may be used to reproduce the figures in their published articles, and to redo those figures with variations in the parameters.

WaveLab has over 800 files which are documented, indexed and cross- referenced in various ways. MATLAB .mex files are used extensively to increase throughput, and *fat* mex files are provided with the Macintosh version so Power Macintosh users may take advantage of the increased speed of their machines.

In addition to routines implementing basic wavelet transforms for finite data sets (both periodic transforms and boundary-corrected transforms), wavelet-packet analysis, cosine-packet analysis and matching pursuit, the library contains scripts which the authors believe will assist in learning the practical aspects of wavelet analysis:

Scripts that reproduce the figures in the authors' published articles, including the de-noising articles of Donoho and Johnstone.

"Workouts" that give a quick guide to wavelets (1-d and 2-d); wavelet analysis; wavelet synthesis; wavelet and cosine packets, including the Coifman-Wickerhauser best-basis methodology; matching pursuit; and applications such as data expansion, progressive data transmission, image compression, speech segmentation, de-noising, fast matrix multiplication in wavelet bases, etc.

WaveLab also offers:

A library of datasets that are easily accessible to the user. Besides artificial signals that have scientific or pedagogical appeal, real data ranging from an image of Ingrid Daubechies to a recording by Enrico Caruso are included.

A point-and-click browser that allows the user to select data, perform various transforms or de-noising operations, and then see the results without using the MATLAB command-line interface. Extensive documentation, including on-line documentation for each function, Contents files for each subdirectory, a Reference Manual, an Architecture Guide and an overview document, About WaveLab, that introduces the software to a first-time user.

Wavelab 0.701 can be obtained for free at

`http://www-stat.stanford.edu/~wavelab/`

It is only compatible with Matlab 4.x as the next version, Wavelat 0.8 developed to be the official Wavelet toolbox provided by The MathWorks.

## 7.4 Wavekit

The WAVEKIT-toolbox is a collection of functions for Matlab that implement the following wavelet and wavelet packet algorithms:

One- and two-dimensional (periodic) fast wavelet and wavelet packet transforms and the best basis algorithm for wavelet packets. An implementation of the fast matrix multiplication algorithm of Beylkin, Coifman, and Rokhlin for both wavelets and wavelet packets. Various demonstrations on visualizing wavelets, signal analysis, and the multiplication algorithm.

The programs in this toolbox are not optimized other than that the inner-most loops are written in C. They are intended for learning about wavelets, not for serious applications. The toolbox grew out of my desire to teach myself about wavelets and now I feel it has matured to the point where it might prove useful to others. I think some of the visualization tools and demos are rather unique.

`http://www.math.rutgers.edu/~ojanen/wavekit/`

## 7.5 Wavelet Analysis Software with Significance and Confidence Testing

`http://paos.colorado.edu/research/wavelets/software.html`

The toolbox is being developed by Christopher Torrence from the Advanced Study Program, National Center for Atmospheric Research, and Gilbert P. Compo, NOAA/CIRES Climate Diagnostics Center, Boulder, Colorado

## 7.6 TIME-FREQUENCY TOOLBOX for Matlab

`http://www.physique.ens-lyon.fr/ts/tftb`

The Time-Frequency Toolbox is a collection of about 100 M-files , developed for the analysis of non-stationary signals using time-frequency distributions. It consists in groups of signal generation files, processing files and post-processing files (including visualization tools).

The toolbox contains numerous algorithms for time-frequency signal analysis, with special emphasis on quadratic energy distributions within Cohen's class and the affine class, including most recent developments based on reassignment.

As usual under Matlab, each function of the toolbox has a help entry and is documented in a reference manual. Several demonstration M-files are also available, showing the possibilities of the toolbox, and following the contents of a user's guide.

The toolbox is primary intended for researchers and engineers with some basic knowledge in signal processing. It requires Matlab v4.2c (or later version) and the Signal Processing Toolbox, v3.0 or later. In end of May 1998, a the toolbox was updated to be compatible with Matlab 5.x.

The Time-Frequency Toolbox has been developed by Francois Auger, Olivier Lemoine, Paulo Goncalves and Patrick Flandrin.

The Time-Frequency Toolbox results from a research effort conducted under the auspices of the French CNRS (Centre National de la Recherche Scientifique) within its successive "Groupements de Recherche" Traitement du Signal et Images (O. Macchi) and Information, Signal et Images (J.-M. Chassery). Parts of the Toolbox have also been developed at Rice University, when one of the authors (PG) was visiting the Department of Electrical and Computer Engineering, supported by NSF. Supporting institutions are gratefully acknowledged, as well as M. Guglielmi, M. Najim, R. Settineri, R.G. Baraniuk, M. Chausse, D. Roche, E. Chassande-Mottin, O. Michel and P. Abry for their help at different phases of the development.

# 8 Conclusion

In this report, we have explored the field of the so-called *Wavelets* with an emphasis on the issues of time series modelling. It involved examining the various Time-Frequency domain analysis techniques and compare them with the Wavelet transform. A theoretical basis for wavelets as well as the set of ideas that lead to them was given. Issues such as wavelet function determination and scale parameter selection was discussed in the context of time series analysis and preprocessing. Some important issues such as resolution and statistical inference were *approached* but further work must be carried out as the study of the application of wavelet to time series are scarce.

A signal preprocessing technique based on wavelet transform and Principal Component Analysis (PCA) were also presented. The proposed method enable to improve the feature extraction process, especially for multi-scale characteristics but also improve further the compression of the input space regularization.

The theories and techniques were illustrated by two chaotic time series examples, namely the *Competition Time Series* and the *Sunspot Time Series*. We showed how the wavelets could be a valuable tool to analyze and process such time series.

Further works should include a close analysis on the determination of the optimal wavelet function with respect to Time-Frequency resolution and statistical significance. A more detailed study on how exploiting the wavelet transform in term of preprocessing and multiscale modelling may lead to improvements in time series modelling.

Another aspect related to time series analysis and preprocessing would be the generalization and complexity issues. The wavelet analysis could be used to produce an estimate of the complexity of the input-output space mapping. The wavelet preprocessing could be used to actually reduce the complexity of the input space by compressing it and (or) extracting causal inputs. Wavelet would produce a Time-Frequency complexity estimate so that time, frequency or amplitude clustering could be used in ensemble models. Individual model could be trained to fit local features in either the frequency or time domain as well as the amplitude space. As mentioned already in [20], reducing the complexity of the mapping by dividing the problem and them recombine the individual prediction may very well produce far better modelling performance.

## References


[1] P. Naidu, *Modern Spectrum Analysis of Time Series*, CRC Press, 1996.

[2] D.B. Percival, A.T. Walden, *Spectral Analyis for Physical Applications : Multitaper and Conventional Univariate Techniques*, Cambridge University Press 1993.

[3] F. Auger, P. Flandrin, P. Gonçalvès, O. Lemoine, *Time-Frequency Toolbox, Tutorial*, CNRS, Groupements de Recherche *Traitement du signal et Image* and *Information, Signal et Images*, http://www.physique.ens-lyon.fr/ts/tftb.

[4] W.L. Briggs, V.E. Henson, *The DFT, An Owner s Manual for the Discrete Fourier Transform*, Society for Industrial and Applied Mathematics (SIAM) Philadelphia, 1995.

[5] G-M. Cloarec, J.V. Ringwood, *Probabilistic Approach for Hybrid Neural Network Modeling*, to appear at the International Workshop on Advanced Black-Box Techniques for Nonlinear Modeling: Theory and Applications, Katholieke Universiteit Leuven, Belgium, July 8-10, 1998.

[6] D. Gabor, *Theory of Communication*, J. Inst. of Electr. Eng., vol. 93, pp. 429-457, 1946.

[7] A. Haar, *Zur Theorie der Orthogonalen Funktionensysteme*, Mathematics Annal,, vol. 69, pp. 331-371, 1910.

[8] I. Daubechies, *Orthonormal Bases of Compactly Supported Wavelets*, Comm. Pure Applied Mathematics, vol. 41, pp. 909-996, 1988.

[9] I. Daubechies, *The Wavelet Transform Time-Frequency Localization and Signal Analysis*, IEEE Trans. Inform. Theory, vol. 36, pp 961-1004, 1990.

[10] I. Daubechies, *Ten Lectures on Wavelets*, Society for Industrial and Applied Mathematics, pp 357, 1990.

[11] G. Strang, *Wavelets and Dilation Equations : A Brief Introduction*, SIAM Review, Vol. 31, No. 4, pp 614-627, 1989.

[12] T. Edwards, *Discrete Wavelet Transforms : Theory and Implementation*, SIAM Review, Junr 1992.

[13] D.L. Donoho, I.M. Johnstone, *Ideal Spatial Adaptation by Wavelet Shrinkage*, Biometrika, vol. 81, pp 425-455, 1994.

[14] C. Torrence, G.P. Compo, *A Practical Guide To Wavelet Analysis*, Bull. Amer. Meteor. Soc.,vol. 79, pp. 61-78, 1998.

[15] B. Jawerth, W. Sweldens, *An Overview of Wavelet Based Multiresolution Analysis*, SIAM Review, Vol. 36, No. 3, pp 377-412, 1994.

[16] C.E. Heil, D.F. Walnut, *Continuous and Discrete Wavelet Transform*, SIAM Review, Vol. 31, No. 4, pp 628-666, Dec. 1989.

[17] F. Murtagh, A. Aussem, *Using the Wavelet Transform for Multivariate Data Analysis and Time Series Forecasting*, Proc. IFCS'96, Kobe, Springer-Verlag, Data Science, Classification and Related Methods, C. Hayashi, H.H. Bock, K. Yajima, Y. Tanaka, N. Ohsumi and Y. Baba, eds., Springer-Verlag, 617-624, 1998.

[18] G.P. Nason, T. Sapatinas, A. Sawczenko, *Statistical Modeling of Time Series using Non-Decimated Wavelet Representation*

[19] Wavelet software was provided by C. Torrence and G. Compo, and is available at URL: http://paos.colorado.edu/research/wavelets/

[20] G-M. Cloarec, *Statistical Methods for Neural Networks Prediction Models*, Research Report EE/JVR/97/2, Control Systems Group, School of Electronic Engineering, Dublin City University, 1997.

[21] G-M. Cloarec, *Multistep Ahead Prediction of the Sunspot Series using Hybrid Neural Network Models*. Research Report EE/JVR/97/3, Control Systems Group, School of Electronic Engineering, Dublin City University, 1997.