Evolutionary Spectra for Exploratory Data Analysis

William S. Rea
University of Canterbury
Christchurch
New Zealand

February 17, 2009

Abstract

This paper gives a brief introduction to windowed Fourier analysis also known as evolutionary spectral analysis (ESA). ESA has largely fallen into a backwater because the type of analysis it was intended to perform is now usually done with wavelets. ESA is not identical to wavelet analysis and so could be a worthwhile analysis to routinely perform along side other more established methods. The paper presents some example time series of tree ring sequences and stock market volatilities and shows how evolutionary spectral analysis can be useful in the exploratory data analysis phase of analysing a time series.

1 Introduction

This brief note is intended to highlight the usefulness of evolutionary spectra in exploratory data analysis. Evolutionary spectral analysis (ESA), also known as windowed Fourier analysis, became computationally feasible as an exploratory technique about the same time that wavelet analysis became popular and was widely applied to time series. By taking a fixed width data window, sliding it across the data and making an estimate of the spectra at each point one obtained a series of overlapping spectra which could be used to study the both the time varying and time invariant properties of a time series. ESA was first used to try to understand the variation through time of both quasi-period phenomena which appeared in geophysical applications (the line components) and changes in the structure of the spectrum over time (the continuum component). Because of the large overlap between subsets of the data the ESA plots contain a great deal of redundant information.

ESA is often criticised as inferior to wavelets in the wavelet literature (Torrence and Compo, 1998). It is fair the say that much of this criticism is justified. Consequently it seems to have fallen into a backwater but could be a worthwhile analysis to routinely perform because, while similar to wavelet analysis, it does do different things.

It is well understood that in many time series the observed values are constrained to lie in some bounded range but the process which produces the series is either non-linear or changing over time or both. It is a hallmark of non-linear processes that their estimated spectrum depends on when you sample the process (Gipp, 2001). Thus a single spectral estimate covering the whole of the sample period may mask a great deal of change in the process over time.

In spectral analysis there are many difficult questions which need to be answered to obtain a useful spectral estimate. For example

1. Should the time series be centered? (i.e. the sample mean substracted)
2. How serious are the problems of leakage and overtones?
3. Should the data be tapered? If so, how much of the data should be tapered? Is a single taper sufficient or are multiple tapers required?
4. Does the data require pre-whitening? If so, how do we contruct a suitable pre-whitening filter?
5. Should we undertake any smoothing of the periodogram?
6. Is a parametric estimate worthwhile instead of a non-parametric estimate?
7. Is any part of the spectral estimate biased?

ESA does not seek to minimize the difficulty in, or importance of, answering these questions but if used as an exploratory technique less care can be taken because the purpose is to locate interesting features in the series which will be subjected to much more careful and detailed examination later as the process of analysis progresses. In fact, answers to the above questions may become clear by examining an evolutionary spectral plot which has had only minimal steps taken to deal with the above issues. That said, it is feasible to use multitaper estimates for the spectra but, of course, at a higher computational cost.

The remainder of the paper is organised as follows. Section (2) discusses the data sets analysed. Section (3) presents the methods used. Section (4) give the results the results. Section (5) contains the discussion and Section (6) concludes.

2 The Data Sets

There are two data sets used in this paper. The first is the Campito Mountain data of LaMarche (1974) which is a 5405 year sequence of annual ring thicknesses from bristlecone pines on Campito Mountain, California. A plot of the time series is presented in Figure (1).

The second data set is a set of 16 realized volatility series of large capitalization stocks which are part of the Dow Jones Industrial Averages. The period of the data was from January 3, 1994 to December 31, 2003. There are 2539 daily realised volatility values. Scharth and Medeiros (2007) give details on how the data set was produced. We use Exxon-Mobil as our example data set. A time series plot of the log volatility in Figure (2).

3 Method

The Campito data was zero padded to a length of 5500 data points and subjected to an evolutionary spectral analysis with 500 data windows each of a 1000 years in width. Between the generation of each spectral estimate the start point was moved nine years giving an overlap of 991 years. The spectral estimates were generated by centering the data, using a 10 percent cosine bell taper and a (5,5) modified Daniell smoothing window using `spectrum` in `R` (R Development Core Team, 2005). The spectral estimates were transferred to `Matlab` for plotting.

The Exxon-Mobil (ticker symbol XON) stock from the financial data was

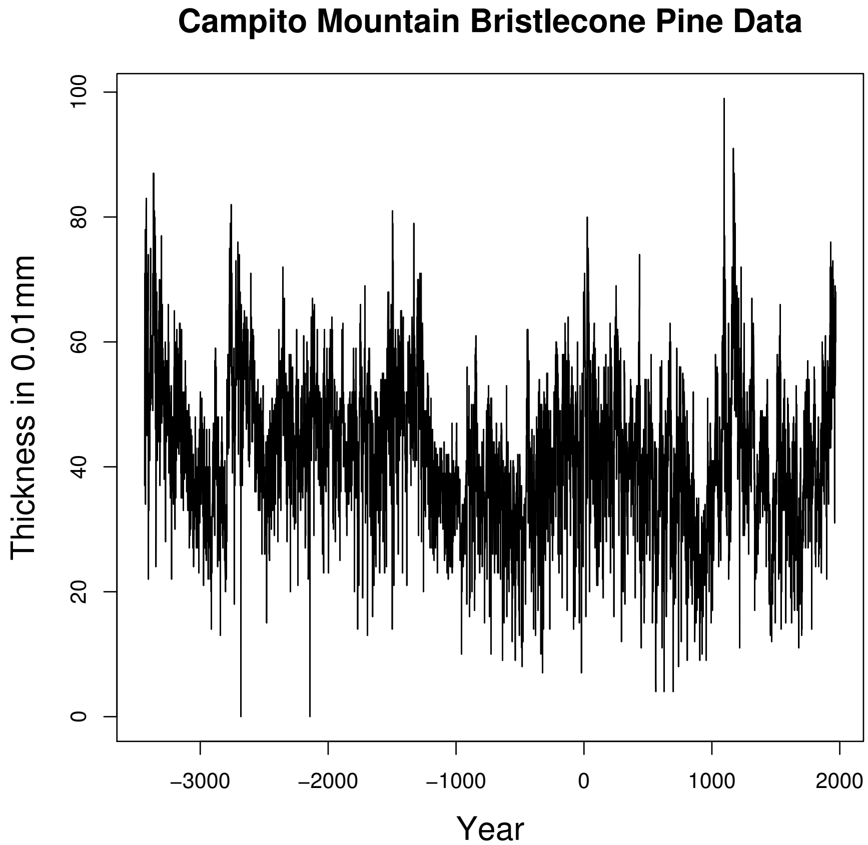

Figure 1: Plot of the Campito Mountain bristlecone pine tree ring time series.

subjected to an evolutionary spectral analysis with 250 data windows each of 500 trading days in width, Between the generation of each spectral estimate the start point was moved eight trading days giving an overlap of 492 trading days. This means that 2500 days of the 2539 days of data was analysed. Spectral estimates were generated and plotted as above.

The Campito and XON data was subjected to a wavelet analysis using the software of Torrence and Compo (1998). Scales between two and 256 years or trading days as appropriate were used.

The multitaper analyses were performed with the SSA-MTM toolkit of Ghil et al. (2002).

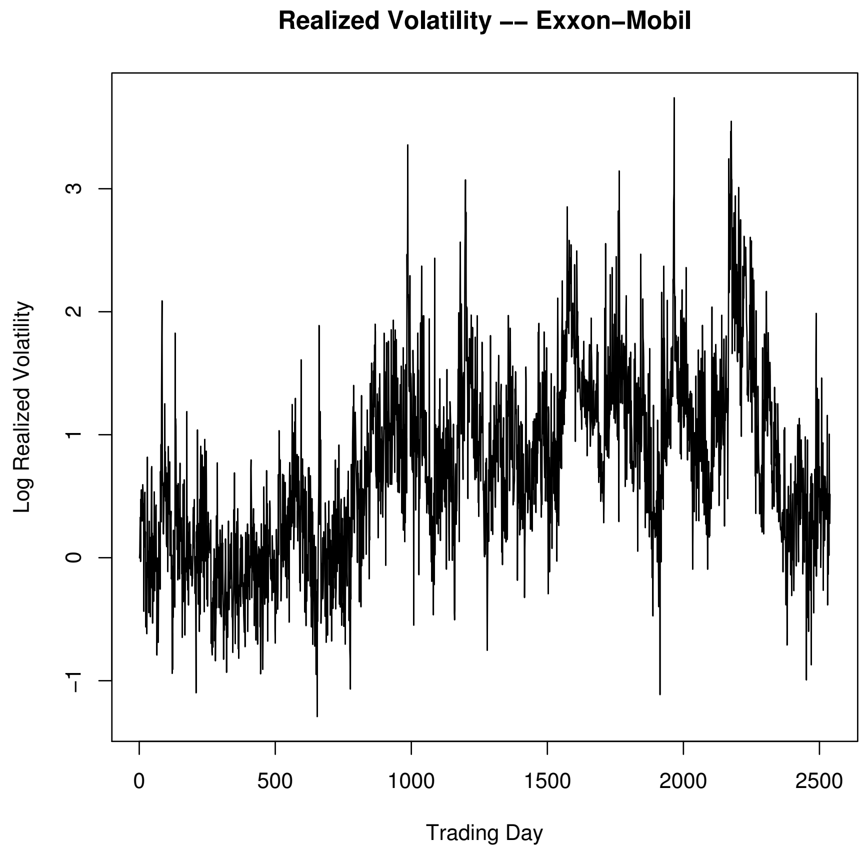

Figure 2: Plot of the log of the Exxon-Mobil realised volatility time series.

4 Results

Figure (3) presents an evolutionary spectrum for the Campito data using a 1000 year wide data window. The two horizontal axes are the start date with negative numbers being the years BC and positive numbers being years AD, and the frequency in cycles per year. The vertical axis is in decibels (dB), which is a logarithmic scale.

The broad features of the spectrum are immediately obvious. The "ripples" are due to the use of the Fejer kernel in the spectral estimate. The fact that there are a large number of "ripples" present in the plot indicates

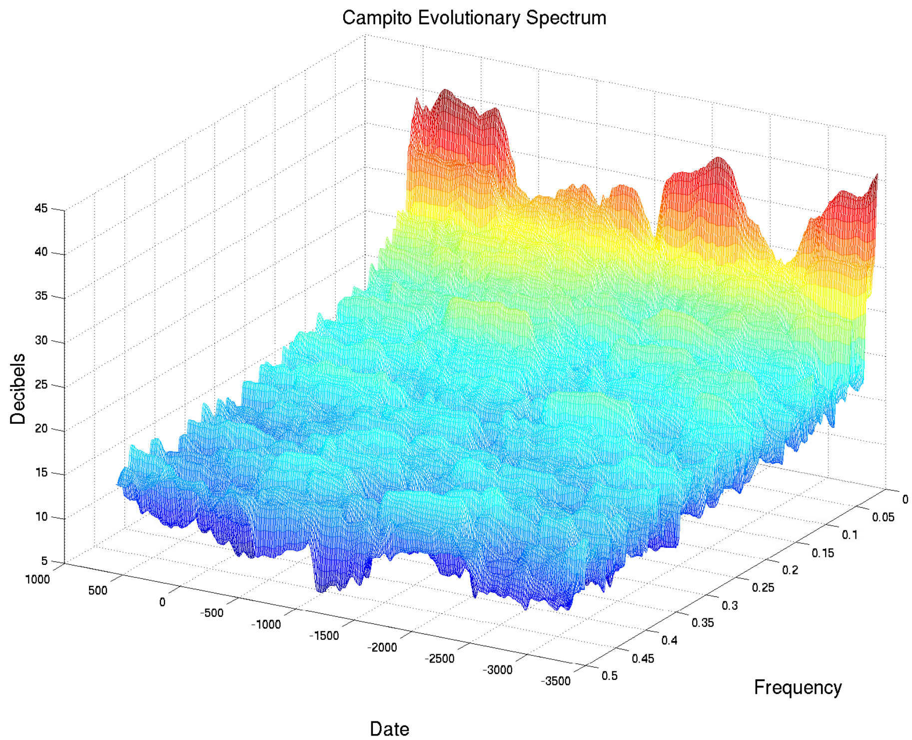

Figure 3: Evolutionary spectral estimate for the Campito Mountain data with 1000 year window width.

that we have significant leakage from the higher power low frequencies regions. This indicates that a single taper may not be sufficient to obtain a good spectral estimate. There is some very long term evolution of the spectrum which can be seen in changes in the very lowest frequencies. There are three "peaks" which reach the 40 dB level. There is a period around 1200 BC where the variability is reduced compared to the remainder of the series. This manifests itself as a "valley" of dark blue colour running up the center of the plot. By contrast there are a number of quite sharp yellow-capped "bumps" in the medium frequencies on either side of the valley feature.

Of the interesting features in the data we single out one clear "bump" which appears somewhere close to the 1000 BC in the frequency region between 0.1 and 0.2 cycles per year and extends close to the year 0. This is subjected to a multitaper analysis.

Figure (4) presents a multitaper (Thomson, 1990) spectral estimate using three discrete prolate spheriodial sequence tapers. To keep the window size the same as in the ESA we chose a single 1000 year period in the Campito data starting at 484 BC. An AR(1) model has been fitted to the data to estimate the continuum component of the spectrum and against which the 90, 95 and 99 percent significance levels are estimated. The AR(1) estimate seems to model the continuum reasonably well except the very lowest and highest frequencies. The particular periodicity which prompted us to look at this subset of the data appears as a 7.1 year cycle which is significant at the 99 percent level.

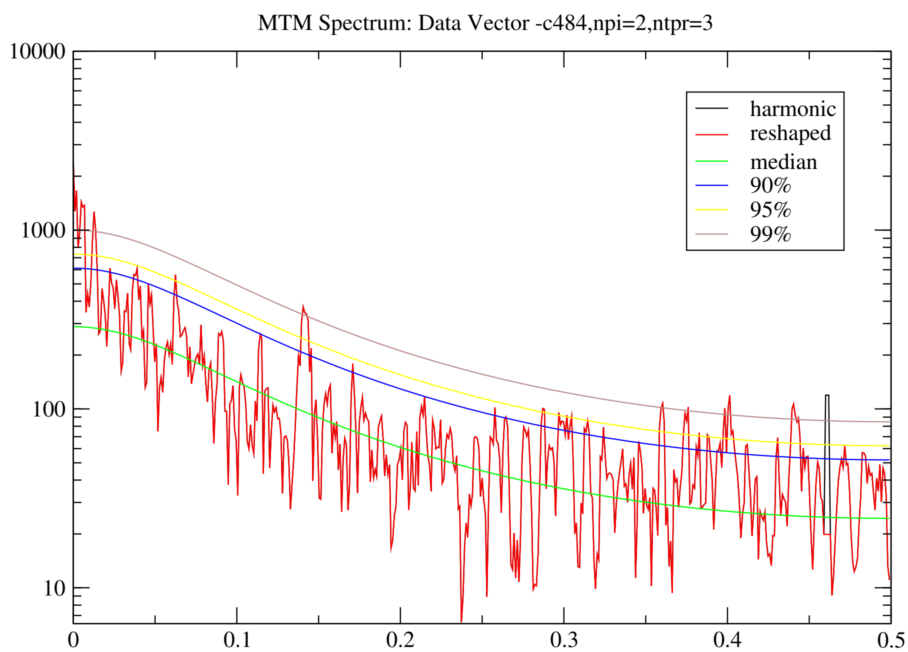

Figure 4: Multi-taper spectral estimate for a 1000 year window width starting at 484BC for the Campito Mountain data.

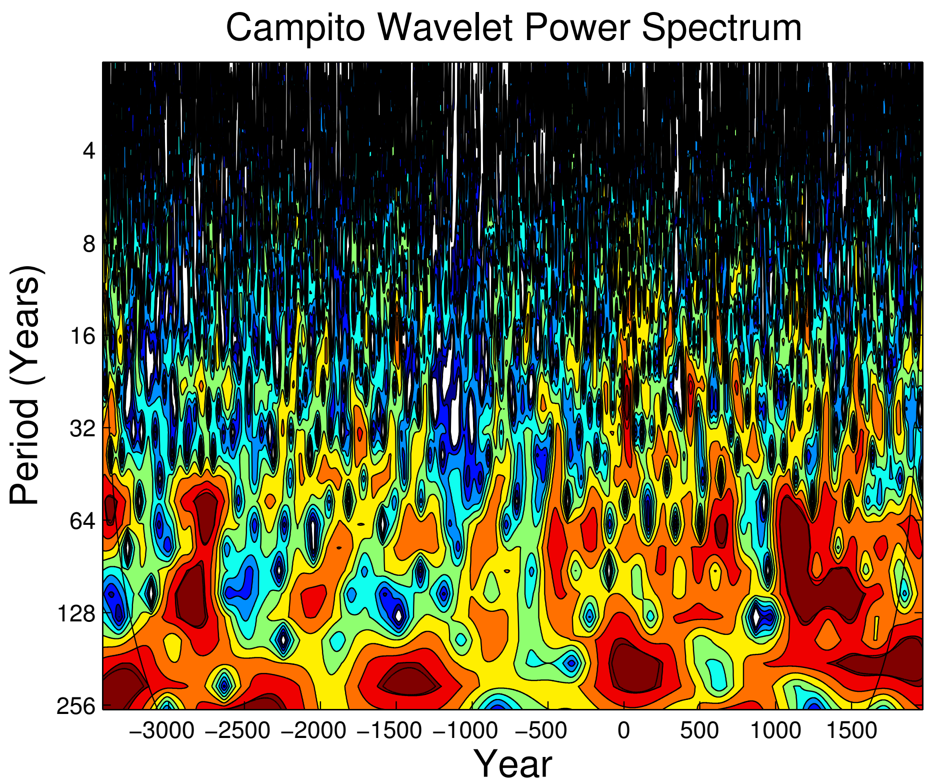

Figure 5: Plot of a wavelet power spectrum for the Campito time series.

Figure (5) presents a wavelet power spectrum for the Campito data. There are several ways of presenting a wavelet analysis, but this two dimensional contour plot is one of the common ones. The vertical axis uses a $\log_2$ scale and is the period in years. The horizontal axis is the time in years, again with negative number being years BC and postive numbers years AD. The two dark curved lines on either side of the plot are the cone of influence (COI) and are the regions of where the spectrum is unreliable because of the inclusion of zero padding at the ends of the series. Some of the features mentioned for the evolutionary spectra are evident here. The "valley" feature seen around 1000 BC is evident in the same place as an intrusion of the blue into the periods of up to 64 years. A cyclic phenomena would manifest itself as a horizontal band of high power in the spectrum.

Figure (6) presents an evolutionary spectrum for the realized volatility series. The general features of the evolutionary spectra are immediately obvious at a glance. There is a slow increase in the power over time for cycles longer than 0.1 (i.e. periods of 10 trading days and longer.) The “ripples” in the foreground of the plot are usually attributed to leakage and not regarded as significant. Thus subseries taken from these periods would require efforts to be made to control leakage to obtain a useful spectral estimate. There are a couple of periods in the data where the problem of leakage seems neglible, particularly at the start of the series and again after trading day 1000.

Figure (7) presents a multitaper spectral estimate for a single 500 trading day period starting at trading day 1750. This period was selected because in the evolutionary plot there appears to be a “bump” in the spectrum at around 0.2 cycles/day (i.e. a five day cycle) starting shortly after trading day 1500 and continuing until the end of the data.

For comparison purposes we present a wavelet power spectrum in Figure (8). The two dark curved lines on either side of the plot at the cone of influence (COI) and are the regions of where the spectrum is unreliable because of the inclusion of zero padding at the ends of the series.

In the wavelet spectrum the change from low variability on time scales from 64 to 256 tradin days to high variability appears to be quite abrupt and occuring about trading day 1000. This could indicate the presence of a structural break in the data. There also appears to be a cycle in the data at scales between 128 and 256 trading days starting shortly after trading day 1000 and continuing to approximately trading day 2200.

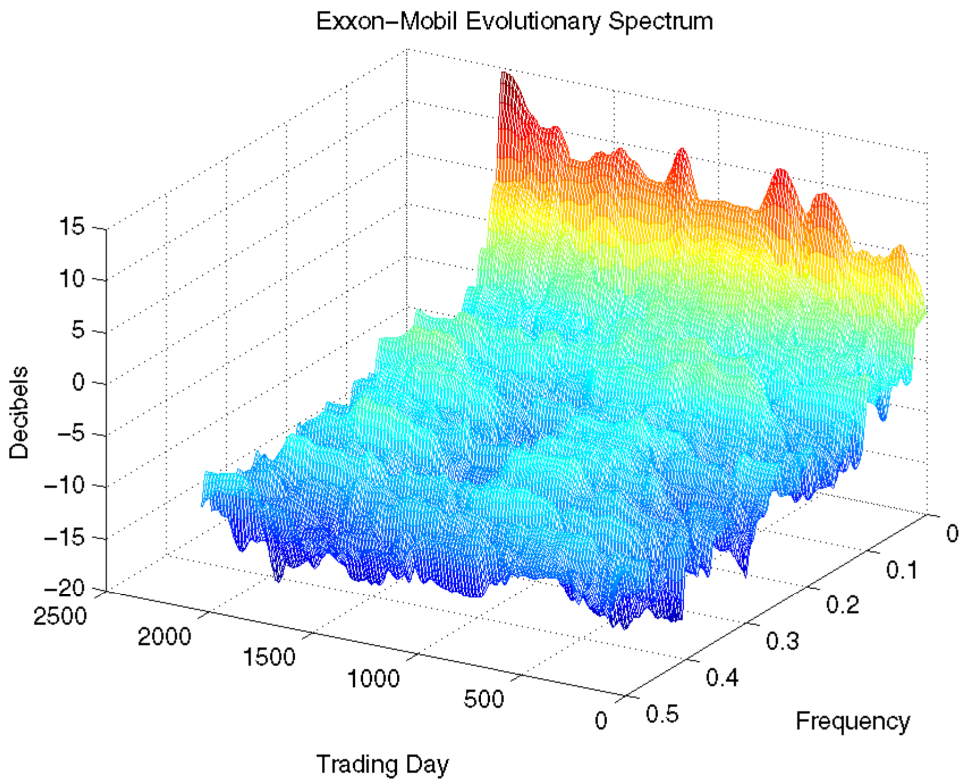

Figure 6: Evolutionary spectral estimate for the Exxon-Mobil realized volatility data with 500 trading data window width.

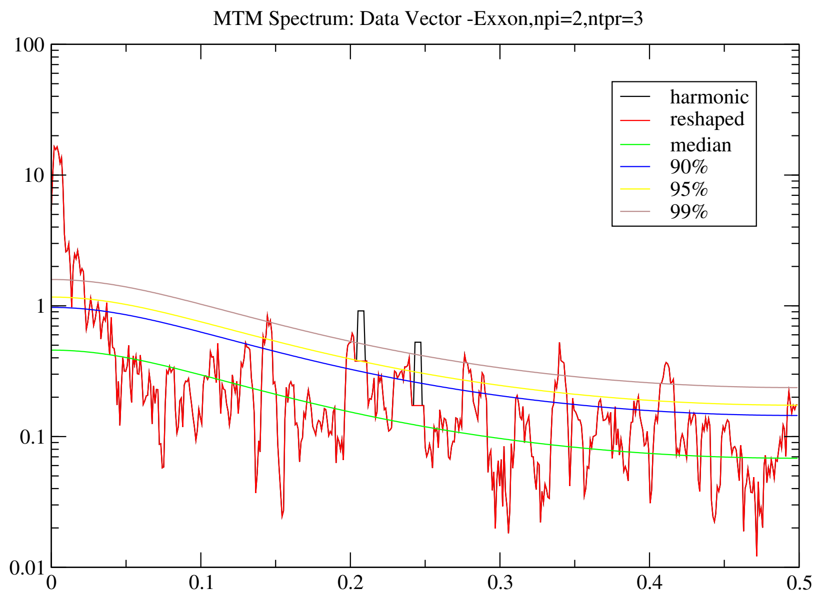

Figure 7: Multi-taper spectral estimate for a 500 trading day window width starting at the 1750th trading day for the Exxon-Mobil realized volatility data.

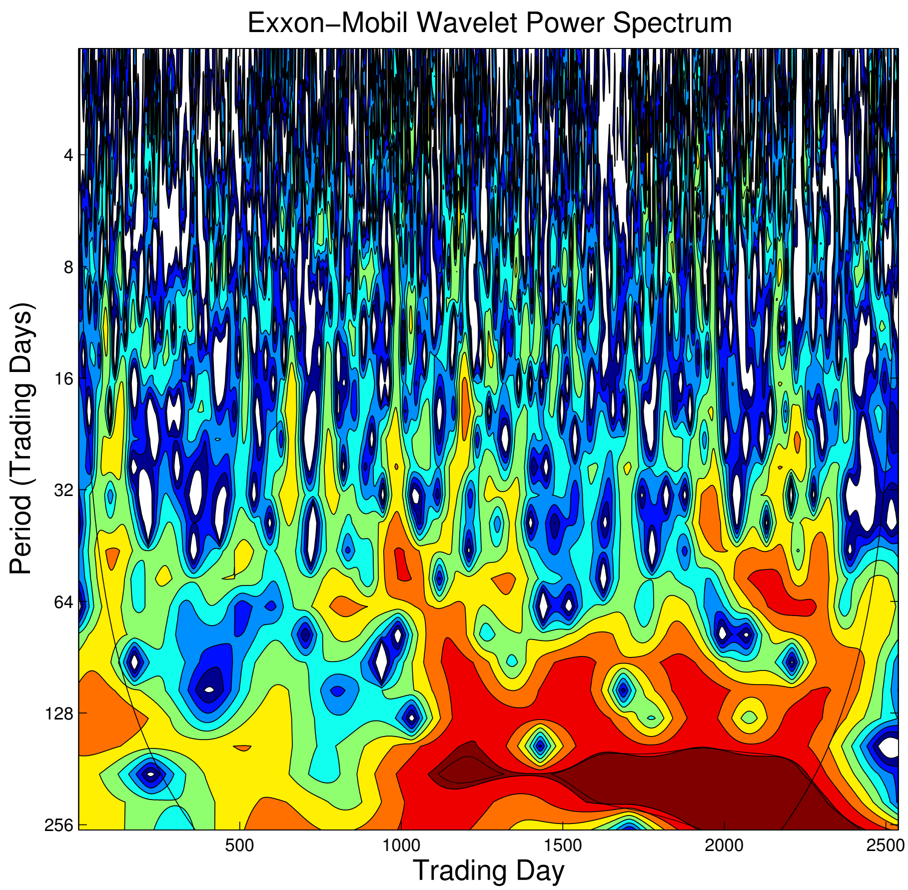

Figure 8: Wavelet plot using a Morlet wavelet as the mother wavelet.

5 Discussion

The Camptio data was presented as a case study because it is easier to understand what insights an evolutionary spectral analysis might afford us over a single spectral estimate in tree ring data than for the financial data. The process which generated the tree ring sequence, which includes climate, does not appear to have been entirely stable over the 5000 years we examined of the data. Some of the changes can be easily seen, particularly the slowly evolving changes in the lowest frequencies (longest time scales). There is a small "valley" of dark blue near 1500 BC in which the variability of the data is much lower than in periods which lie to either side of this feature. This period of time also seems not to suffer as much from leakage. There are a number of relatively prominent "ridges" running parallel to the time axis indicating the presence of periodic components in the data which do not extend for the whole period of the sequence.

In this case we selected a single segment of the data starting at 484BC and subjected it to a multitaper analysis. This was motivated by the presence of the yellow coloured "ridge" starting around 1000 BC. The multitaper spectral estimate with three tapers is presented in Figure (4). While a few lines with 99 percent or higher significant could be by chance (there are 500 individual frequencies estimated) this line appears in a region of the spectra where the continuum seems well fitted by the AR(1) model. Recall that each estimate covers a period of 1000 years. Even allowing for the high over lap between estimates, this periodicity appears to have persisted for at least 1500 years before dieing out. This increases our confidence that it is reflecting some real phenomena, probably climatic, and is worth further investigation.

The wavelet spectrum (Figure 5) does not show any clear evidence of a long term periodicity in the data. This is somewhat surprizing given that both the ESA and the mutlitaper analysis indicate the presence of some periodic phenomena at several time scales. The seven year cycle could simply be obscured by the method of data presentation.

In the Exxon-Mobil data (Figure 6) there is a general increase in the power in the lower frequencies with time. (This is not as marked as in the some of the other volatility series.) The "ripples" seen in the foreground of the ESA would usually be attributed to leakage and not regarded as being of significance. There are periods when the problem of leakage seems to be of no practical consequence hence if these subseries were analysed more closely little or no tapering would be required. However, on the left hand side of the plot there is one high "ridge" feature with a frequency of 0.2 (5 trading days). Because of its presence we present a multitaper estimate in Figure (7). There is a statistically significant line at the 5 day period and, if real, would indicate some systematic change between high and low volatility on a weekly basis. One should not read too much into the continuum estimate because it clearly is not a good fit at the low frequencies.

Figure (8) presents an estimate of a wavelet power spectrum. There are some clear advantages with the wavelet spectrum. It can localize the periods and scales within the series where there is high or low variability. Two examples are the period between trading days 400 and 500 and, as noted above the abrupt shift near trading day 1000. In the first of these periods the variability in the data at scales of about 128 trading days is quite low as indicated by the blue colour. In the second there is a period of high variability on scales of between 64 and 256 trading days indicated by the rapid change from low variability (blue regions) to high variability (orange, red, brown). These types of small features or rapid changes would only appear in the EVA plot if the window size was correctly selected.

6 Conclusion

ESA offers some advantages over wavelet analysis.

1. It is easier to understand the broad changes in the continuum structure of the spectra in ESA, particularly in the higher frequency regions. In wavelet analysis these changes are often obscured in a mass of detail.

2. It can be easier to identify cyclic components, again particularly in the high frequency regions, in an ESA than with a wavelet analysis.

3. It is easier to transfer the knowledge gained from an ESA to advanced spectral analysis tools such as multiper analysis.

4. ESA should be considered as an additional tool available to the time series analyst.

References

Ghil, M., Allen, M. R., Dettinger, M. D., Ide, K., Kondrashov, D., Mann, M. E., Roberteon, A. W., Saunders, A., Tian, Y., and Yiou, P. (2002). Advanced Spectral Methods for Climatic Time Series. *Reviews of Geophysics*, 40(1):1–39. doi:10.1029/2000RG000092.

Gipp, M. R. (2001). Interpretation of climate dynamics from phase space portraits: Is the climate strange or just different? *Paleoceanography*, 16(4):335–351.

LaMarche, V. C. (1974). Paleoclimatic Inferences from Long Tree-Ring Records. *Science*, 183:1043–1048.

R Development Core Team (2005). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Scharth, M. and Medeiros, M. C. (2007). Asymmetric Effects and Long Memory in the Volitility of Dow Jones Stocks. working paper.

Thomson, D. J. (1990). Time series analysis of Holocene climate data. *Philosophical Transactions of the Royal Society of London A*, 330:601–616.

Torrence, C. and Compo, G. (1998). A Practical Guide to Wavelet Analysis. *Bulletin of the American Meteorological Society*, 79(1):61–78.