Time-Series Analysis of Variable Star Data

Matthew Templeton
AAVSO, 25 Birch Street, Cambridge, MA 02138

Based on a workshop session at the 92nd Spring Meeting of the AAVSO, April 25, 2003; revised August 2004

Abstract Time-series analysis is a rich field of mathematical and statistical analysis, in which physical understanding of a time-varying system can be gained through the analysis of time-series measurements. There are several different techniques of time-series analysis that can be usefully applied to variable star data sets. Some of these techniques are particularly useful for data found in the AAVSO International Database. In this paper, I give a broad overview of time-series analysis techniques useful for variable star data, along with some practical suggestions for the application of different techniques to different types of variables. Included are elementary discussions of traditional Fourier methods, along with wavelet and autocorrelation analysis.

1. Introduction

Time-series analysis is the application of mathematical and statistical tests to time- varying data, both to quantify the variation itself and to learn something about the behavior of the system. Ultimately, the goal of time-series analysis is to gain some physical understanding of the system under observation: *what makes the system time-variable?; what makes the system similar to or different from other systems?; is the system predictable?; and can we place reliable limits on the behavior of the system?*

Clearly, simple forms of "time-series analysis" were known in ancient times, since many ancient civilizations made accurate predictions of various cyclical celestial phenomena. The birth of modern time-series analysis dates to the early 19th century, with Joseph Fourier's description of the Fourier series, and later, the Fourier transform. (Carl Friedrich Gauss first derived the Fast Fourier Transform around 1805, before Fourier published his work, but Gauss did not publish his results.) The greatest advances in time-series analysis coincided with the development of computing machines and the digital computer in the mid-20th century. The digital computer made possible the statistical analysis of large amounts of data in much less time than would be possible by human calculators. Along with this came the development of more efficient algorithms for time- series analysis (like the rediscovery of Gauss' Fast Fourier Transform by Danielson and Lanczos in the 1940s), and the development of new ideas such as wavelet analysis and chaos theory. Today, time-series analysis is regularly applied to a wide variety of problems in the real world, from radio and telecommunications engineering to financial forecasting.

In this paper, I will give a basic overview of several topics in time-series and statistical analysis relevant to variable star research. I will begin with a discussion of the Fourier transform and its many implementations and uses. Then I will discuss two other important types of time-series analysis: *wavelet analysis*, which is useful for studying systems which change over time, and *autocorrelation analysis*, which is useful for systems which may not exhibit coherent periodic behavior but nevertheless have characteristic periods. I conclude with a summary and a list of resources for those interested in conducting their own research in this field.

2. Fourier analysis and the Fourier transform

Fourier analysis is the technique of using an infinite number of sine and cosine functions with different periods, amplitudes, and phases to represent a given set of numerical data or analytic function. In so doing, you can estimate the period (or periods) of variability by determining which of these functions are statistically significant. The amplitudes (and phases) of these components are determined with a *Fourier transform*.

If we have a set of time-varying data, given by $x(t)$, then the *Fourier transform*, $F(\nu)$, is given by the integral

$$F(\nu) = \int_{-\infty}^{+\infty} x(t)\,\exp(-i2\pi\nu t)\,\mathrm{d}t,$$

where ν is the *frequency*, defined as $\nu = 1/P$, i is the imaginary square root of -1 used in complex numbers, π is a mathematical constant approximately equal to 3.14, and the sine and cosine functions are represented by the complex exponential function given by Euler's formula:

$$\exp(-i2\pi\nu t) = \cos(-2\pi\nu t) + i\,\sin(-2\pi\nu t).$$

(For a more complete discussion of the mathematics behind the Fourier transform, see Bracewell (2000).)

If a set of data, $x(t)$, contains a coherent signal at some frequency, ν', then the value of the Fourier transform, $F(\nu)$, should reach a local maximum at ν'. If the data contain several signals with different frequencies, then $F(\nu)$ should have local maxima at each, with the global maximum at the frequency having the largest amplitude.

The Fourier transform is an extremely powerful yet elegant technique that is used in many areas of mathematical analysis and the physical sciences. However, as one might expect, its power is finite, limited by the amount and quality of data that are transformed. The data place several limits on the usefulness of the transform, including the maximum and minimum periods testable, the accuracy of the period determination, and the minimum statistically significant amplitude that can be found.

As an example, consider the following case: you have a data set spanning 5000 days, with an even sampling rate of 10 data points per day (or one data point every

2.4 hours). The maximum period detectable in this case is 5000 days, since the data should cover one complete cycle. However, this detection would be very unreliable since without additional data you have no idea whether the variation detected was truly periodic or simply a short-term fluctuation that merely looks like a 5000-day period. A more reliable limit is 5000/2 or 2500 days, since you could detect two complete cycles at that period within the data set. The span of the data set also determines the resolution of the Fourier transform, which is the precision to which the frequency (or period) may be determined. The sampling theorem defines the set of frequencies that may be measured by a given data set, and the separation between two adjacent frequencies defines the resolution of the transform. The resolution is defined by

$$d\nu = 1/N\Delta \quad \text{OR} \quad dP = P^2/N\Delta$$

where N is the total number of samples (50000), and Δ is the space between the samples (0.1 day). In terms of frequency, the resolution is simply the inverse of the span of the data; if the data span 5000 days, the frequency resolution is 1/5000 d^{-1}. An example of this is shown in Figure 1, where the peak of the Fourier transform of R CVn is shown for two different data sets, having different spans. The data set with the longer span clearly provides a much more precise determination of the period than the shorter data set. This makes clear the need for long sets of data, particularly when studying long period variables.

The minimum period detectable by our example data set is 0.2 day, corresponding to a maximum frequency of 5 cycles per day. This is because your *sampling frequency* of 10 points per day would (potentially) allow you to detect the object at maximum and minimum once each cycle. This is known as the *Nyquist frequency*. The Nyquist frequency is important not only because it defines the highest frequency (and shortest period) detectable with a given dataset, but also because it defines the maximum sampling rate you need in order to fully describe variations up to the maximum frequency.

In certain circumstances, it *is* possible to detect frequencies higher than the sampling rate. The transform will suffer from *aliasing*, in which several different peaks appear in the transform, along with the real one. The alias peaks are separated from the true frequency by integer multiples of the sampling frequency, such that the transform will look like a "picket fence" when plotted. In the case of regular sampling, the alias peaks will have equal statistical significance to the real peak, and it is therefore impossible to tell which peak in the power spectrum is the correct one. In the case of uneven sampling, you will still have aliasing, but the strengths of the alias peaks will generally be lower than that of the dominant one. An example of this can be seen in Figure 2, showing the Fourier transform of MACHO (Massive Compact Halo Objects) observations of a δ Scuti star (Alcock *et al.* 2000). Although the MACHO sampling rate was very low (one observation per day, on average), δ Scuti variations could be detected because of the uneven sampling and nearly complete phase coverage.

It must be stressed that the case of the δ Scuti stars shown here represent a very specific case where the variations are strictly periodic in the long term and have very regular light curves. Other cases where one may detect periods shorter than the sampling period are other regular pulsators like the δ Cepheids and RR Lyrae stars, and the strictly periodic eclipsing binary stars. In general, it is not possible to uniquely determine periods shorter than the sampling period, and any such analysis must be done with great care.

One major consideration in Fourier analysis is noise—both the intrinsic noise of photometric observations, and measurement errors of the data. Noise is always present in a given signal regardless of the quality of measuring devices. From a physical standpoint, the precision of photometry is limited by Poisson statistics—the more photons arrive at the detector, the more precisely one is able to measure the signal. The noise level (defined as the square root of the number of photons) decreases *but never vanishes* as the number of photons increases. From an observational standpoint, measurements of a given signal will always contain some error. For example, visual observations are rarely accurate to more than 0.1 magnitude for an individual observer, while systematic differences between observers can amount to 0.2 to 0.3 magnitude or more. CCD and photoelectric measurements

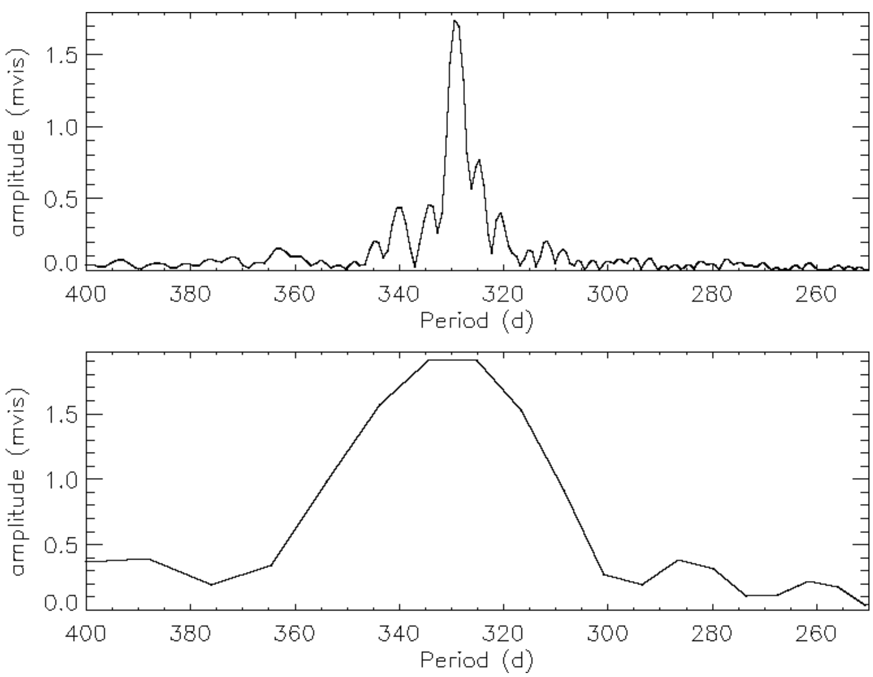

Figure 1. Fourier transforms for two data sets of R CVn. Top panel: AAVSO visual data spanning JD 2420232–2452906 (32674 days). Bottom panel: AAVSO visual data spanning JD 2449909–2452906 (2997 days). While both data sets show the spectral peak at nearly the same period, the pulsation period of 329 days is much better resolved in the longer data set than in the shorter one.

typically have smaller (but still significant!) errors caused by sky background and instrumental effects. Fourier analysis of a given set of data assumes that *everything* contained within a given data set is a signal. Thus *noise* will appear in the Fourier spectrum at some level defined by the strength of the true periodic modulations relative to the background noise. Measurement of this noise level is an important part of Fourier analysis, as it allows you to determine the reliability of your results.

Figure 2 also shows this quite well. Individual photometric data points from MACHO typically have errors between 0.05 and 0.2 magnitude, depending upon the brightness of the star and how crowded the fields are. These errors manifest

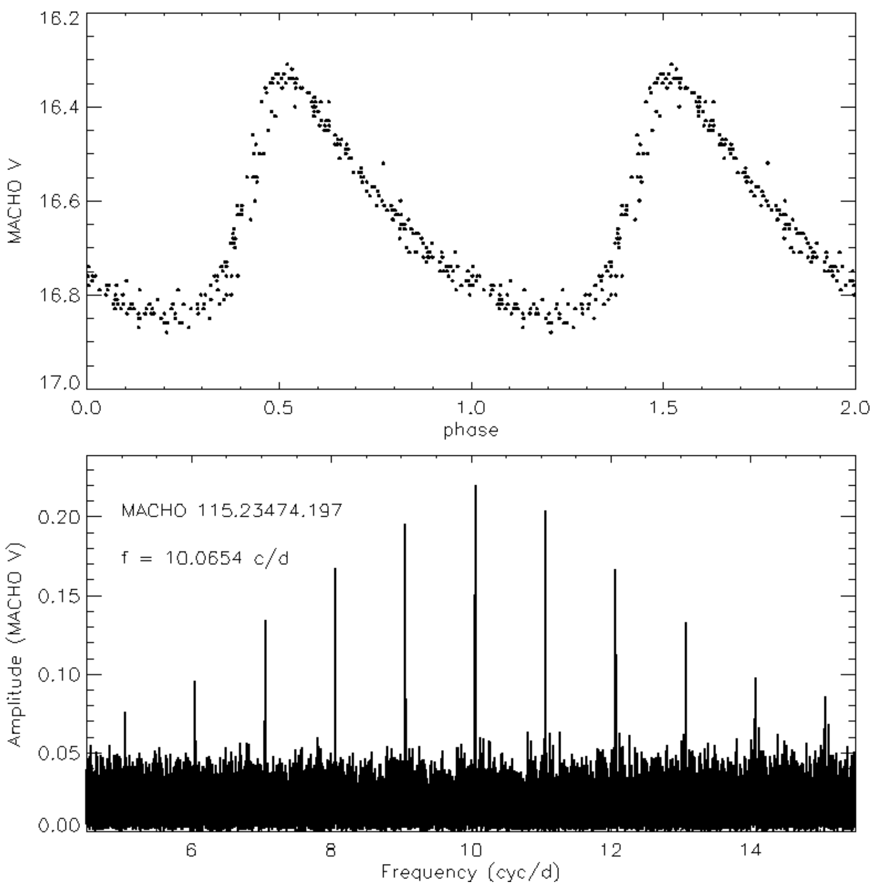

Figure 2. Folded data and Fourier transform of MACHO 115.23474.197, a high-amplitude delta Scuti star with a pulsation frequency of 10.06 cycles per day, much higher than the data sampling rate of about 1 per day. Because the sampling is so low, the transform suffers from aliasing shown by the secondary peaks offset from the main peak by integer multiples of 1 cycle per day.

themselves in the Fourier spectrum as the dense forest of peaks with a limit of about 0.05 magnitude. The mean value of these peaks provides some indication of the detection limit of any periodic signal. If the signal strength is large relative to the scatter in observations (as with Mira variables), then the noise level of the Fourier transform will be relatively low. If the signal strength is small (as with Cepheid or δ Scuti observations) then the noise level will be higher. In the case of the MACHO data shown in Figure 2, any signal with amplitude near to or less than about 0.05 magnitude would be indistinguishable from noise, and therefore undetectable.

The determination of formal uncertainties on the spectral information from Fourier transforms can be non-trivial. Horne and Baliunas (1986) derived formal confidence limits for the Scargle (1982) *periodogram*, a variation on the Fourier transform. In their work, they proved that it was possible to detect extremely low-strength signals in the presence of noise, so long as the number of data points (and the span of the data) is large enough. In general, the "signals" generated by random noise in a given data set will asymptotically approach a limit defined by the noise level, while the true signal of interest will grow in statistical significance as more data are obtained. With AAVSO data, you may find it possible to obtain good results for low-amplitude stars even if the data are noisy, while in other cases, the available data do not allow the reliable detection of signals. Thus it is important to keep in mind the noise level and errors in photometric data when performing a Fourier analysis.

Most time-series analysis packages contain algorithms that calculate the noise level of a given set of data, as well as the statistical significance of any detected variability. For example, the AAVSO's TS package provides both period and amplitude error estimates for any peaks found by the program. These error estimates can help you to determine whether your results are real or spurious, though you should *always* double-check the computed values against what you expect.

A final caveat to those using Fourier analysis to detect periods is that discrete methods for analyzing unevenly sampled data will produce spectral artifacts of the *sampling* of the data, in addition to any signal contained within the data. For example, data taken over a series of nights (with daylight gaps between) will have *aliases* caused by the 1 cycle/day sampling windows. These alias frequencies are centered on any real signals in the data, offset from the central frequency by integer multiples of 1 cycle/day. The reason for this is that the data sampling produces a *window function* in the Fourier transform, which is convolved with the Fourier peak of the "real" signal. The result is the "picket fence" of frequencies in the Fourier spectrum, like that shown in Figure 2. In AAVSO data, other common examples of *aliasing* include the 1 cycle/year aliases seen in the Fourier spectra of some long period variables near the ecliptic, caused by the annual interference of the Sun. In general, one deals with aliasing in the spectrum by assuming the strongest peak observed is the correct frequency, but you should *always* use caution when interpreting the Fourier spectra of gappy data.

3. Fourier transform algorithms and applications

The Fourier transform has a huge number of practical applications not limited to astronomy and the physical sciences, or indeed even to *time*-series analysis (the Fourier transform is easily applicable to analysis of multidimensional *spatial* frequencies as well as temporal ones). There are also a large number of algorithms in existence to compute the Fourier transform, both because of its wide variety of applications, and because of varying needs for computational efficiency.

The simplest algorithm is a brute-force integration of equation (1), known as the *discrete Fourier transform*, in which the integral is carried out as a finite sum over data points $f(t)$. The discrete transform forms the basis for many Fourier methods, particularly those which deal with unevenly sampled data, like the *date-compensated discrete Fourier transform* (Ferraz-Mello 1981), the Lomb-Scargle periodogram (Scargle 1982), and Foster's CLEANest (Foster 1995). While powerful, discrete methods can be computationally expensive, since the computation time increases in proportion to the square of the number of data points, N. However, as mentioned in the introduction, Gauss invented a very fast implementation of the Fourier transform, which was later re-invented by Danielson and Lanczos in the 1940s, and is now known as the *fast Fourier transform* or FFT. This method permits the extremely fast computation of Fourier transforms, since the calculation time only increases as $N\,log_2(N)$, rather than the N^2 of discrete methods. The FFT is frequently used in large-scale data analysis, and in "real- time" Fourier analysis (like laboratory spectrum analyzers). Its major drawback is that it *requires* even data-sampling; you must either sample your data evenly (a rarity in long- term variable star observing), or else re-grid your data (which introduces errors). Given the computational power available with even basic home computers, the use of the fast Fourier transform is no longer a necessity in time-series analysis, even for relatively large data sets.

The Fourier transform has many uses in variable star research. Its most fundamental use is in finding periods in data. In the case of monoperiodic data, Fourier analysis should reveal the dominant period if the amplitude is a sufficient fraction of the noise level. However, it is rare that real stars have purely monoperiodic, sinusoidal light curves, and Fourier analysis can reveal additional information about the variability. For example, Fourier analysis is useful for analyzing stars with multiple periods, as many types of pulsating variables have. Or, if a star is monoperiodic, but has a light curve that is non- sinusoidal, then the Fourier transform can provide the amplitudes and phases of the *Fourier harmonics*—signals at integer multiples of the fundamental frequency that distort the fundamental sinusoid—which can in turn provide information about the physical properties of the star (see Simon and Lee 1981). The amplitude and phase information from the Fourier harmonics are commonly used in the analysis of pulsators like Cepheids and RR Lyrae stars, providing information on the pulsation mode type, metallicity, evolutionary state, and luminosity (see Morgan 2003 and references therein). Fourier analysis can provide other important information, such as the evolution of periods and amplitudes over time (Foster 1995), and the physical origin of variability in aperiodic variables like accretion-powered sources (e.g., van der Klis 1995).

4. Wavelet analysis

Wavelet analysis and the wavelet transform are relatively recent developments in time-series analysis. The development of the wavelet transform came from the need to analyze signals that were transient and/or non-sinusoidal in nature. The wavelet transform of a set of time-series data, *x(t)*, is given by

$$W(\omega,\tau;\, x(t)) = \omega^{1/2} \int x(t)\, f^{*}(\omega,(t-\tau))\, dt \quad \text{(eq. 3; Foster 1996)}$$

where ω is a test frequency, τ is a "lag" time or a position within the light curve, and the function, *f*, is called the "mother wavelet"—a function which determines how the signal should vary with time, frequency, and position within the light curve. (The "*" indicates the *complex conjugate* of the function, *f*, is used.) The wavelet transform is *extremely flexible* because the mother wavelet can be nearly any function at all. This means, for example, one can include both a specific waveform (e.g., sinusoid) to search for a periodicity, and a time-varying weighting function (like a sliding window) to study the time-dependence of the signal. In this way, one could study both the frequency spectrum of a given signal, as well as the evolution of that spectrum as a function of time.

This analysis method has great utility in several areas of astronomy and astrophysics, since many objects have varying periods, or have no fixed period at all and instead show transient periodicities or quasiperiodicities. For example, the long-period Mira stars have long been known to exhibit slightly varying periods from cycle to cycle, while a few of these stars (like R Aquilae) are known to have strongly varying periods indicative of evolutionary changes within the star. Other stars, like the semiregular variables and the RV Tauri stars, do not have a constant period but instead vary with one or more characteristic periods which become incoherent when viewed over the full light curve. Still other stars exhibit temporary periods or quasiperiods, like accreting dwarf novae stars having superhumps, or X-ray binaries with high-frequency quasiperiodicities. In all cases, wavelet analysis enables you to look for transient or time-varying behavior within a given data set.

However, like more traditional Fourier analysis techniques, the wavelet transform also has limitations. The major limitation, as with Fourier analysis, is that the data set must be long enough and well-sampled enough to adequately measure the periods of interest. If the data only span 1000 days, it would be meaningless to test periods longer than 500 days. Additionally, when the wavelet contains a window function, the data should span a length of time such that the window is meaningful. For example, if the data span 1000 days, the period of interest is 200 days, and the wavelet window covers five cycles, the wavelet analysis will not give meaningful information about the time evolution of the signal—nearly all of the data will lie within the window for any chosen value of τ.

The AAVSO has a very powerful tool available for computing wavelet transforms called WWZ, for *weighted wavelet z-transform*. The algorithm was developed by Foster (1996) specifically with AAVSO data in mind, and may be downloaded in BASIC and FORTRAN77 versions from our website. Foster's algorithm performs the wavelet transform given in equation 3, using a wavelet function which includes both a periodic, sinusoidal test function of the form $exp(\iota\omega\tau(t-\tau))$ and a Gaussian window function of the form $exp(-c\omega^2(t-\tau)^2)$, where both the frequency, ω, and the user-defined constant, c, determine the width of the window. The algorithm fits a sinusoidal wavelet to the data, but as it does so, it weights the data points by applying the sliding window function to the data; points near the center of the window have the heaviest weights in the fit, while those near the edges have smaller weights. The window slides along the data set, giving us a representation of the spectral content of the signal at times corresponding to the center of that window.

When analyzing AAVSO data with WWZ, there are a few things to keep in mind. For one, the data set should have a reasonably long time-span, preferably much longer than the expected period of the star. If you were interested in studying period evolution in a variable, it would be best to have a span of data many times longer (perhaps by a factor of 50 or more) than the mean period of the variable. This will allow the algorithm to slide the window over a large span of data and determine the best-fitting period over completely independent regions of the light curve.

Another thing to note is that WWZ allows you to select the width of the window (via the constant, c), which gives you some flexibility in the timescales you wish to investigate for period changes. However, there are tradeoffs when making the window narrower or wider. Recall from the discussion about Fourier analysis that the span of the data affects the period range and resolution. Since the data window acts to change the span of the data, making the window narrower effectively reduces the span of the data and consequently makes the period resolution worse. But by doing so, you can study period changes over very short timescales. Likewise, if you widen the window to improve the period resolution, you worsen the time resolution of the transform, making it difficult to detect short-term variations.

As an example, the wavelet transforms of the semiregular variable Z Aurigae are shown in Figure 3. The star is believed to undergo "mode switches" where the period of the star suddenly changes from one period to another—in this case from a 110-day to a 137-day period. Z Aurigae has made the switch between these two periods a few times over the past century of observations, and these mode switches can be easily detected with WWZ.

In the top panel, I've chosen a wide window, which gives very fine period resolution, but has smoothed out much of the temporal variation. In fact, so much smoothing is used, that the transform missed two mode switches at JD 2429000 and 2439000. The amplitude variation (not shown) is also greatly reduced, since it, too, is calculated as a weighted average of the amplitude over the entire window. In the bottom panel, I chose a much narrower window, which brings out the time-varying nature of the spectrum. In particular, the transform has detected the two very short

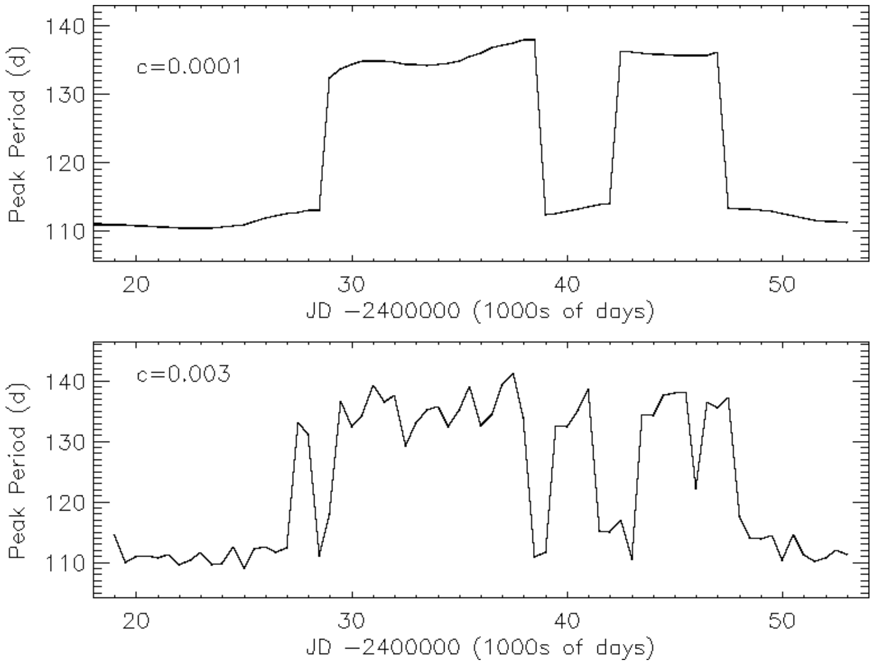

Figure 3. wavelet transforms of Z Aurigae. Top panel: wavelet transform with a wide window, showing better period resolution at the expense of temporal resolution. Bottom panel: wavelet transform with a narrow window, showing better temporal resolution at the expense of period resolution.

mode switches noted above. However, it does so at the cost of period resolution. There is clearly increased (and largely unphysical) scatter about the mean periods of 137 and 110 days, and the periods cannot be defined more accurately than to within about ± 10 days.

So when you use wwz to analyze your data, be aware that both the data and your selection of analysis parameters will have some effect on the results. As with any analysis method, wavelet analysis techniques must be applied wisely, with an eye toward both the limitations of your data, and the information you wish to obtain. No data set is of sufficient quality to provide infinite temporal *and* spectral resolution, and even large spans of data, like those available from the AAVSO, have their limits.

5. Statistical methods

There is another branch of time-series analysis that attempts to quantify temporal variations using statistical means, rather than assuming sinusoidal variations. One commonly used method in time-series analysis is *autocorrelation*. Autocorrelation compares pairs of points $(x(t), x(t+\tau))$ to see whether points separated by a test interval, τ, are similar; if the data have a characteristic timescale similar to this interval, then the autocorrelation function should reach a maximum at that value of τ. There are other methods as well, including the *analysis of variance* (ANOVA) *test*, and *phase dispersion minimization* (PDM). Below, I discuss autocorrelation in some detail, and briefly mention some other methods useful in astronomical time-series.

5.1 Autocorrelation

Unlike Fourier-based methods, instead of asking whether the light curve can be represented by periodic functions, autocorrelation (Box and Jenkins 1976) essentially asks *"what does the light curve look like at times separated by some interval, τ?"* What autocorrelation does is take each data point, measured at time t, and then compare the value of that data point to another at time $t + \tau$. If you perform this test for each pair of data points separated by an interval, τ, you can then average the differences together and see whether most points separated by τ are similar or different. If they're very similar, there will be a peak in the autocorrelation function at τ. If they're very different, there will be a trough. We would expect points separated by τ to be very similar if the data contained some variability with period τ, so the autocorrelation function will have peaks corresponding to periods of variability in the data.

This can be a very powerful method for stars with irregular light curves. This includes stars that are *almost* periodic, like some of the semiregular and RV Tauri stars, and those that may have transient periods or may have a "characteristic period" but have irregular light curves. However, it isn't just meant for irregular variables. You can also use it for strictly variable stars of all types. The only difference then is whether other methods like Fourier analysis might give you more information than the autocorrelation function. The only case where autocorrelation doesn't work so well is in stars that have multiple, simultaneous periods present in the data because the different periods interfere with one another in the autocorrelation spectrum. Fourier methods are far more straightforward in this case.

Figure 4 shows an example of autocorrelation analysis applied to the RV Tauri star R Scuti, which has a period of about 146 days. The top panel shows the AAVSO light curve of R Scuti measured over 2000 days. While a "period" of about 146 days might be apparent to the eye, the light curve is far from regular. There is substantial amplitude variability, and sometimes the variability nearly disappears (as around JD 2448700). By comparing the light curve at times separated by some time-difference, τ, you determine whether these points are generally correlated or not. As shown in the bottom panel, there is a peak in the autocorrelation function at a period of about 146 days. Integer multiples of this period (292 days, 438 days, and so on) are also correlated, suggesting that the 146-day period remains coherent for at least a few cycles, even though the amplitude is highly variable. However, the declining amplitudes of the correlation function's peaks show that the coherence decays over time.

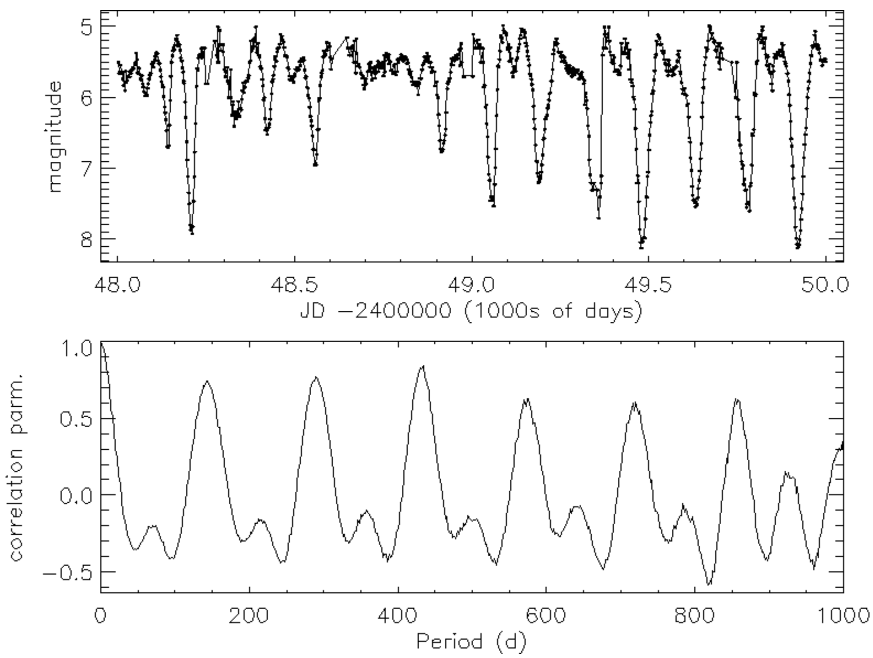

Figure 4. Two thousand days of AAVSO data (top panel) and autocorrelation function (bottom panel) for R Scuti, an RV Tauri star with a period of 146 days. While the data look very irregular, there is a strong characteristic period which autocorrelation clearly reveals.

5.2 Other statistical methods

Autocorrelation is by no means the only statistical method for performing time-series analysis. One period-finding algorithm is the *analysis of variance* or ANOVA test. Here, the variance of the signal is the important thing. In the ANOVA procedure, you bin and fold the data with a series of test periods, and then measure the variance of the data points within each bin. If you find a minimum value of the variance at some folding period, then it is likely that the star has a real period at or near that folding period. This is the time-series algorithm that is contained within the VSTAR program of the AAVSO's *Hands-On Astrophysics* kit, so you may be familiar with it already. It is quite powerful, and is equally adept at detecting both sinusoidal variations and non-sinusoidal or pulsed variations (Schwarzenberg-Czerny 1989). Another statistical method related to ANOVA is *phase dispersion minimization*, or PDM (Stellingwerf 1978). This method is also commonly used in variable star research, though Schwarzenberg-Czerny (1989) suggests that ANOVA is generally superior to PDM, particularly when the number of data points is small. A statistical method *not* related to ANOVA is *maximum entropy analysis* (Fahlman and Ulrych 1982). This method is sometimes used when reconstructing long-term data with gaps is the goal. In general, statistical methods are a very rich and varied field of time-series analysis, and I recommend you investigate this area to determine whether such methods may be useful to you.

6. Summary

The goal of this paper was to introduce the casual reader to some basic principles of time-series analysis, as well as to outline some of the considerations that go into selecting appropriate methods for analyzing a given set of data. The few methods outlined in detail here by no means give a comprehensive description of this rich field of research, but will hopefully serve as a starting point for your investigations. As with any field of research, you should investigate other methods of time-series analysis and their applications to see whether there might be more suitable alternatives to what I've presented here. While several different analysis methods may provide "correct" results for a given project, there are likely other analysis methods unsuitable for your data, and these may give misleading results. A little research in advance may save you time and trouble later on.

Another point I wish to stress is that you should be aware of both the strengths and limitations of the data you wish to analyze *before* you begin your analysis. The amount and quality of a given set of data may not be sufficient to detect the type of stellar variability you are investigating. Measurement errors and noise place limits on the amplitudes detectable in a given data set, while the span and sampling rates of data limit the precision to which periods can be defined. This is particularly important to remember when dealing with visual data from the AAVSO; while visual data may not be suitable for detecting small-amplitude variability, they are often ideally suited for studying long-term changes in pulsation behavior. Again, a little foreknowledge of the goals of your analysis and the limits of your data can save you much time and trouble later on.

7. Additional resources

Beyond the information given in this paper, there are several additional resources for those interested in time-series analysis, including publicly available computer programs useful for astronomy. The best starting point for those interested in analyzing AAVSO data would be the publicly-available time-series analysis codes available from the AAVSO itself—the Fourier analysis code TS, and the wavelet analysis code WWZ. Both codes now exist in BASIC versions suitable for Microsoft Windows machines and in FORTRAN77 versions for any computer with an ANSI-compliant FORTRAN compiler. Both are available free-of-charge from our website:

http://www.aavso.org/cdata/software.stm

Another excellent, publicly available program is the PERIOD98 software package available from the δ Scuti Network of the University of Vienna, Austria. This program was designed for the analysis of multi-periodic stars (like δ Scuti stars), but will work perfectly on monoperiodic variables. The website of this software package is:

http://www.astro.univie.ac.at/~dsn/dsn/Period98/current/

Two other resources for time-series (and other statistical analysis methods) are the Penn State University Astronomy Department's "StatCodes" archive, available at

http://www.astro.psu.edu/statcodes/

and John Percy's "Astrolab" page at the University of Toronto:

http://www.astro.utoronto.ca/~percy/analysis.htm

which includes a "self-correlation" analysis program, a slight variation on the autocorrelation method described above.

Finally, to explore the relatively new field of non-linear time-series analysis and chaos theory, I strongly recommend investigating the TISEAN (Hegger, Kantz, and Schrieber 1999) analysis package and its accompanying documentation. The package may be found at

http://www.mpipks-dresden.mpg.de/~tisean/

The application of chaos and non-linear theory to variable star analysis is a relatively new and unexplored field, but data from the AAVSO have already been used to study the applicability of chaos theory to variable star research (see Jevtic *et al.* 2003, Kollath *et al.* 1998, and Buchler *et al.* 1996).

References

Alcock, C., *et al.* 2000, *Astrophys. J.*, **536**, 798.

Box, G. E. P., and Jenkins, G. 1976, *Time Series Analysis: Forecasting and Control*, Holden-Day Series in Time-Series Analysis, Holden-Day, San Francisco.

Bracewell, R. N. 2000, *The Fourier Transform and Its Applications*, 3rd edition, McGraw-Hill, Boston.

Buchler, J. R., Kollath, Z., Serre, T., and Mattei, J. A. 1996, *Astrophys. J.*, **462**, 489.

Fahlman, G. G., and Ulrych, T. J. 1982, *Mon. Not. Roy. Astron. Soc.*, **199**, 53.

Ferraz-Mello, S. 1981, *Astron. J.*, **86**, 619.

Foster, G. 1996, *Astron. J.*, **112**, 1709.

Foster, G. 1995, *Astron. J.*, **109**, 1889.

Hegger, R., Kantz, H., and Schreiber, T. 1999, *CHAOS*, **9**, 413.

Jevtic, N., Mattei, J. A., and Schweitzer, J. 2003, *J. Amer. Asoc. Var. Star Obs.*, **31**, 138.

Kollath, Z., Buchler, J. R., Serre, T., and Mattei, J. A. 1998, *Astron. Astrophys.*, **329**, 147.

Morgan, S. 2003, *Publ. Astron. Soc. Pacific*, **115**, 1250.

Scargle, J. D. 1982, *Astrophys. J.*, **263**, 835.

Schwarzenberg-Czerny, A. 1989, *Mon. Not. Roy. Astron. Soc.*, **241**, 153.

Simon, N. R., and Lee, A. S. 1981, *Astrophys. J.*, **248**, 291.

Stellingwerf, R. F. 1978, *Astrophys. J.*, **224**, 953.

van der Klis, M. 1995, "Rapid aperiodic variability in X-ray binaries," in *X-Ray Binaries*, W. H. G. Lewin, J. van Paradijs, and E. P. J. van den Heuvel, eds. Cambridge Univ. P., Cambridge.