Fourier crash course for MPO 524
Brian Mapes Mar 2016

Fourier decomposition (or analysis, or transformation) is one instance of a "spectral" analysis (https://en.wikipedia.org/wiki/Spectrum). Data on some sequential domain (like time or space) are decomposed into a set of coefficients multiplying orthogonal functions. In Fourier analysis, the orthogonal functions are circular functions (sines and cosines, or more elegantly packaged into the complex exponential function). Sometimes Fourier analysis is used as a tool for deconstructing and reconstructing data, like for **filtering**, or to build synthetic data. Sometimes, scientific arguments or claims hinge on spectra.

1. Spectral (Fourier) analysis: why?

a. Time ←→ frequency

It is traditional to teach spectral (Fourier) analysis with respect to *time series* like f(t), but actually time is an awkward domain for Fourier decomposition. Time is not a periodic dimension (so far as we know). An infinite time series will take forever to collect, even if we knew the whole past. So the Fourier spectrum in time is always a fiction: an ideal, unknowable. Fourier analysis in time is therefore a bundle of compromises that should really be thought of as **estimation** of the unknowable true frequency spectrum. Why would we want to estimate a spectrum? Usually it is to seek **spectral peaks**, corresponding to periodicities in the time series.

Scientifically, a spectral peak might tell us about what processes are at work generating the things we observe. Circular functions are solutions of orbital equations, so in the ocean or atmosphere we may be trying to detect **forced oscillations,** like from periodic astronomical forcings (diurnal, annual, 22000 years, etc.) in a noisy world. But circular functions are also solutions to linear **wave** equations (with a "restoring" force proportional to -displacement), like "gravity" (buoyancy) waves with period > $2\pi/N \sim 10$ minutes in the tropical troposphere. Fluids also have **free oscillations** at special periodicities, like "inertial" motions with period $2\pi/f$ = 12 hours at the poles. Fluid dynamical **instabilities** sometimes have preferred frequencies, so detecting one frequency vs. another may serve as evidence for instability theories. Sometimes there are periodic **artifacts in data**, like say 60 Hz electronic noise entering some high frequency data sets. Besides wanting to just detect these things, we sometimes want to take them out (artifacts), or isolate them (take out everything else to study some pure oscillation). Spectral analysis is a method for such decomposition or **filtering** of data.

More practically, if a periodicity exists in time, that corresponds to **predictability**. In the extreme case of a single sinusoidal variation (where the spectrum is a single infinitely narrow peak), knowing two values (phase and amplitude) gives you all the values forever. Weaker but positive predictability is implied by spectral peaks of finite strength and width. (It is a little tricky to define a "peak" as we shall see). Efficient-market economists say there are no significant or real spectral peaks in the stock market, because if there were, someone would exploit the predictability by buying when the price cycle is low (driving up price) and selling at the price peak (which would damp the peak).

Like all objective types of **estimation**, spectrum estimation can be thought of as **2 parts: 1) obtaining a result** (known to be imperfect), and **2) obtaining error bars** around that result, which bound the true answer with some high probability (confidence level) -- commonly 95 or 99% in science, but perhaps much higher if engineering safety is involved. So the main purpose of this Crash Course is to understand how error bars in spectral space correspond to (and arise from) errors or imperfections in the time-domain data series f(t).

b. Longitude ←→ planetary wavenumber

In the atmosphere, we have a much more natural domain for spectral analysis: longitude. Variations along a latitude line actually ARE periodic, unlike time variations, so a discrete Fourier spectrum in longitude, with *integer* planetary wavenumbers 1,2,3... [units: cycles/(circumference of earth)], is actually a complete, uncompromised representation. In fact some atmosphere models use a spectral representation, because dynamical terms involving spatial derivatives (like advection or the PGF) can be computed exactly in analytic form [d/dx(sin x) = cos x *exactly*], rather than inaccurately with finite differences like $\Delta u/\Delta x$ on a grid. Of course, that just leads to other challenges.

As in the time domain, there are theories that predict periodicities in longitude, like the prefererred wavelength of baroclinic instability. So Fourier analysis of data bears on scientific theories. Also there are forced wavelengths in longitude (continent-ocean spacings). And again, like in time, there may be predictability to be gained: sometimes isolating a long wavelength feature (usually with a corresponding long time scale) can allow extrapolations (in other word, predictions) of large-scale aspects of the flow pattern.

In nature, time series are (essentially) continuous and infinite, while we always have data that are finite in length (with record length T), available only at discrete times (with spacing dt), and quantized in value (stored in bytes). These compromises all have different signatures in spectral space, as shown in Section 3. First, let's examine a reference "truth" time series, so we can see how degradations affect the spectral view.

2. A "truth" series (7 days repeated forever) and its true spectrum

Figure 1 shows L(t), a time series of downwelling longwave radiation measured by a TAO buoy on the equator at 95W, reported every 2 minutes. You can perhaps sense a diurnal cycle (maybe due to temperature or water vapor?), plus some higher frequency spikes of high downwelling radiation as clouds blow by overhead. I removed the mean (409 W m^{-2}) for clarity. The variance (mean of squared values) is 110.9 (W m^{-2} $)^2$, and the standard deviation is the square root of that, 10.5 W m^{-2}.

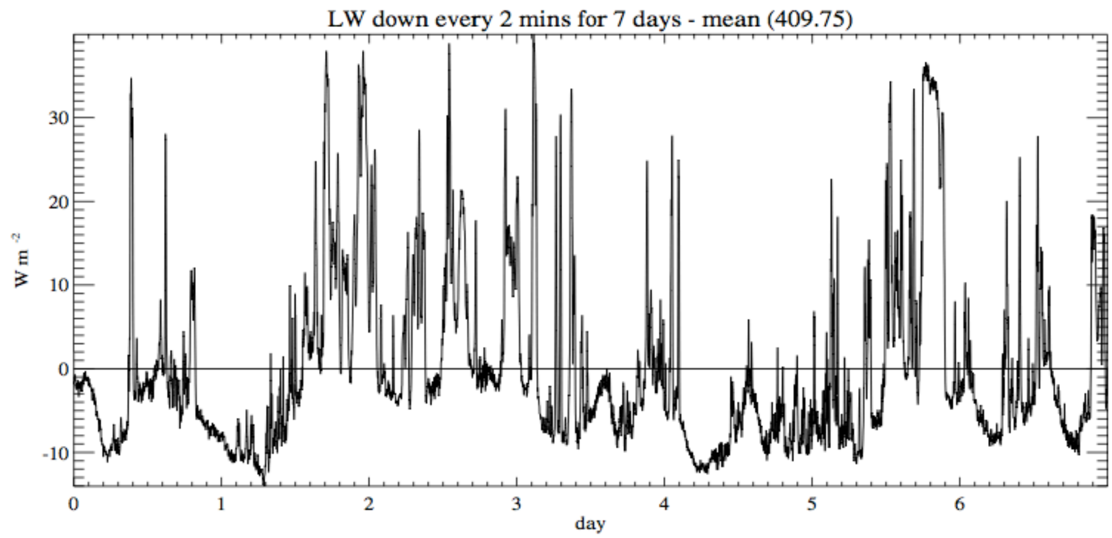


Figure 1. A time series L(t) of downwelling LW radiation on the equator, with its mean removed. The variance is 110.9 (W m^{-2})^2, so the standard deviation is 10.5 W m^{-2}.

A Fourier component (sinusoidal function) has a **frequency** or **wavenumber** (2π/**period** or 2π/**wavelength**). For each frequency, there is an **amplitude** and **phase**. **Power** is amplitude squared. The set is called a **spectrum.** So any series like Fig. 1 has an *amplitude spectrum* (whose square is the *power spectrum*), and also a *phase spectrum*. When we do Fourier analysis with sines and cosines that oscillate to infinity, *we are assuming that the data sequence is repeated periodically to infinity*. Hmm. Within that assumption, the series in Fig. 1 has a "true" spectrum.

Let's look at the variance (or power) spectrum. Section 3 will then show what happens in frequency space as we mess with the data in the time domain (undersample it, or average it into coarser time bins, or quantize its values).

a. Power spectrum

A simple call to the magic function **fft(L(t))** yields the **complex spectrum $\mathcal{L}$(f),** a *complex number* array, where f is frequency. A complex number can be unpacked into cosine (real) and sine (imaginary) components, or into amplitude abs($\mathcal{L}$(f)). and phase ϕ(f). The square of amplitude is called Power P(f).

Because the time interval is finite (and assumed periodic), the **frequencies are discrete**. A wave with frequency of 1.3 cycles per week would violate the assumption that the data sequence is repeated. The frequencies are equally spaced: 1,2,3,... cycles within the data period (T = 7 days). This equal spacing of the possible frequencies (1/T where T is the sequence length) is called the **bandwidth** of the spectrum. It is also equal to the lowest possible frequency, obviously. Because weeks aren't very fundamental, I divide these integers by 7.0 to express frequency units as **cycles per day** (**cpd**).

The highest frequency resolvable by these data is (1 cycle)/(4 minutes), since it takes at least two 2-minute data points to indicate a minimal cycle (a zig-zag). This highest possible frequency is called the **Nyquist** frequency.

The complex spectrum fft( ) is returned in a complex array of length N. How can that be? N numbers have been translated into 2N numbers, since each complex number has 2 numbers packed within it. The answer is that the frequency domain in $\mathscr{L}$(f) includes **negative frequencies** as well as positive. Wha?? For any *purely real L(t),* which data from the real world always are, the spectrum is *symmetric*, with the same value for positive and negative frequencies. So we can just *take the right half of the array and double it* to make a plot with positive frequency as the x axis.

Here is the Power spectrum (or spectral density) P(f):

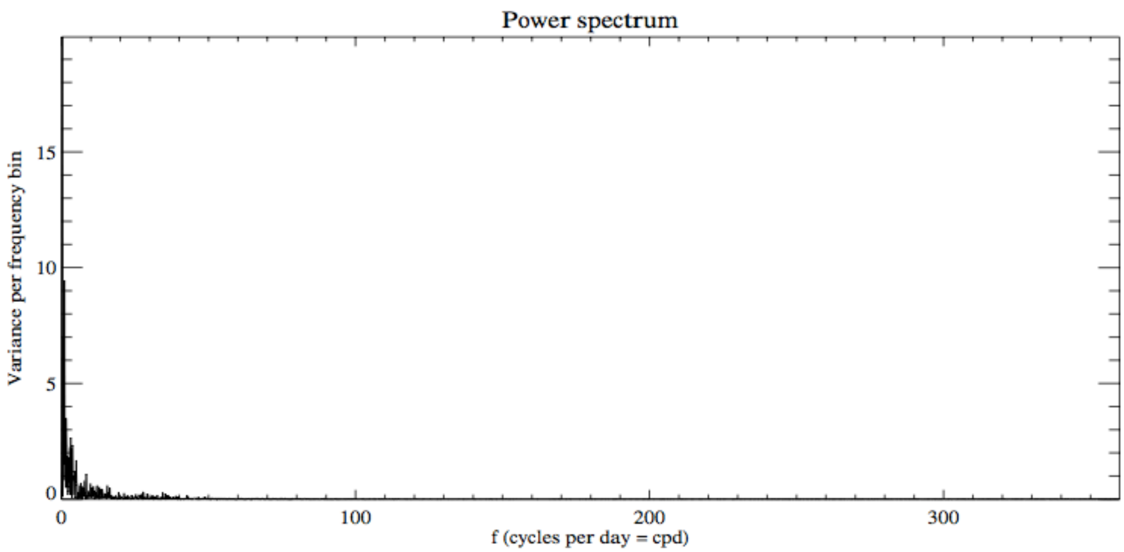


Figure 2. The power spectral density of Fig. 1. Only positive frequencies are shown. Sometimes a raw spectral power plot is called a "periodogram", a charming old timey word.

P(f) is a very spiky thing, with most of the power in the lowest few frequencies.

Parseval's Theorem tells us that the total power – the "area under the curve" if we view P(f) as a continuous spectral *density* by plotting it as a line curve as in Fig. 2, or the SUM of the DISCRETE values making up the spectrum which should really be a bar plot -- is equal to total variance of L(t), 110.9 W^2 m^{-4}. That is impossible to see in Fig. 2, which is a very unhelpful diagram.

Sometimes people use a y-log axis to squash down the tall vertical spikes and bring up the tiny values. Sometimes people use a x-log axis to emphasize the low frequencies that have most of the information. But then the 'area under the curve' aspect is lost. Instead, I like to plot the **cumulative variance** ΣP(f), which asymptotes to the total variance. This has simple variance units, and can be plotted against log (f) which helps emphasize the lowest frequencies. It also helps tamp down the spiky nature of spectra, which do tend to excite our brains irrationally by lighting up our *giraffe!* neurons.

Figure 3 shows the cumulative spectrum. About half the variance is in frequencies lower than 2 cycles per day in this case. Looking back at Fig. 1, does this agree with your intuition? A step up at 1 cycle per day indicates the diurnal cycle.

The red triangles emphasize the **discrete nature of the spectrum**, but a line connects them for eyeball convenience. The second-lowest frequency (2 cycles/ 7 days) contains a lot of power, but that is not a very well resolved period...you'd want a longer record to conclude anything. Can you see that cycle in Fig. 1?

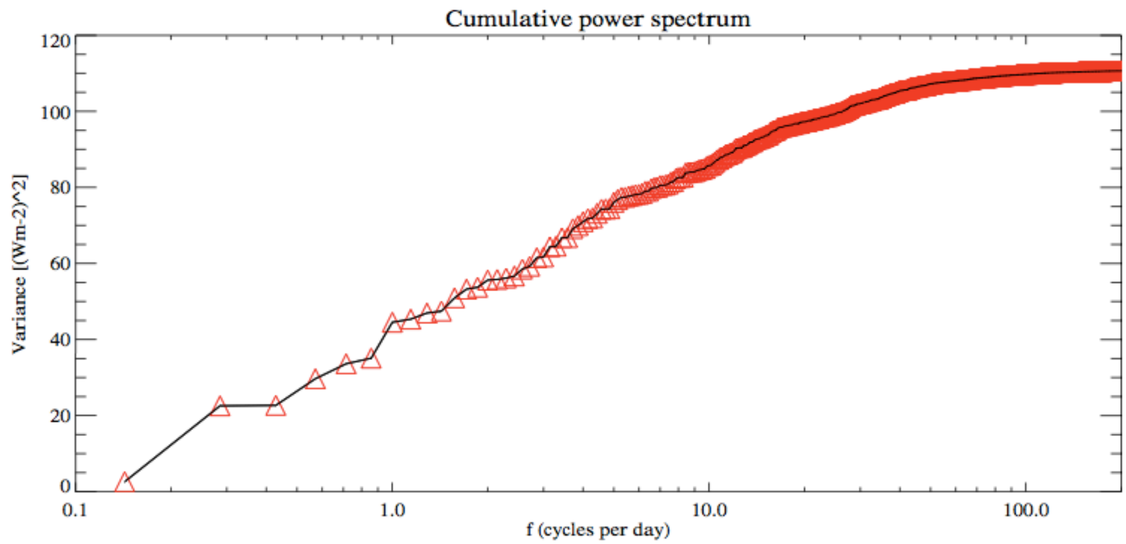


Figure 3. The cumulative power spectral density of Fig. 1. Only positive frequencies are shown, I doubled the spectrum to get this to reach the right total – a good easy check on methods.

Since the fft() is reversible by ifft(), no information has been lost, and we can reconstruct L(t) by adding up all the Fourier components (called **harmonics**).

If only certain frequency bands are included, the reconstruction is called **Fourier filtering**.

Figure 4 (next page) shows an example with several frequency bands, then a rainbow colored cumulative reconstruction using more and more frequencies until the full time series is recovered.

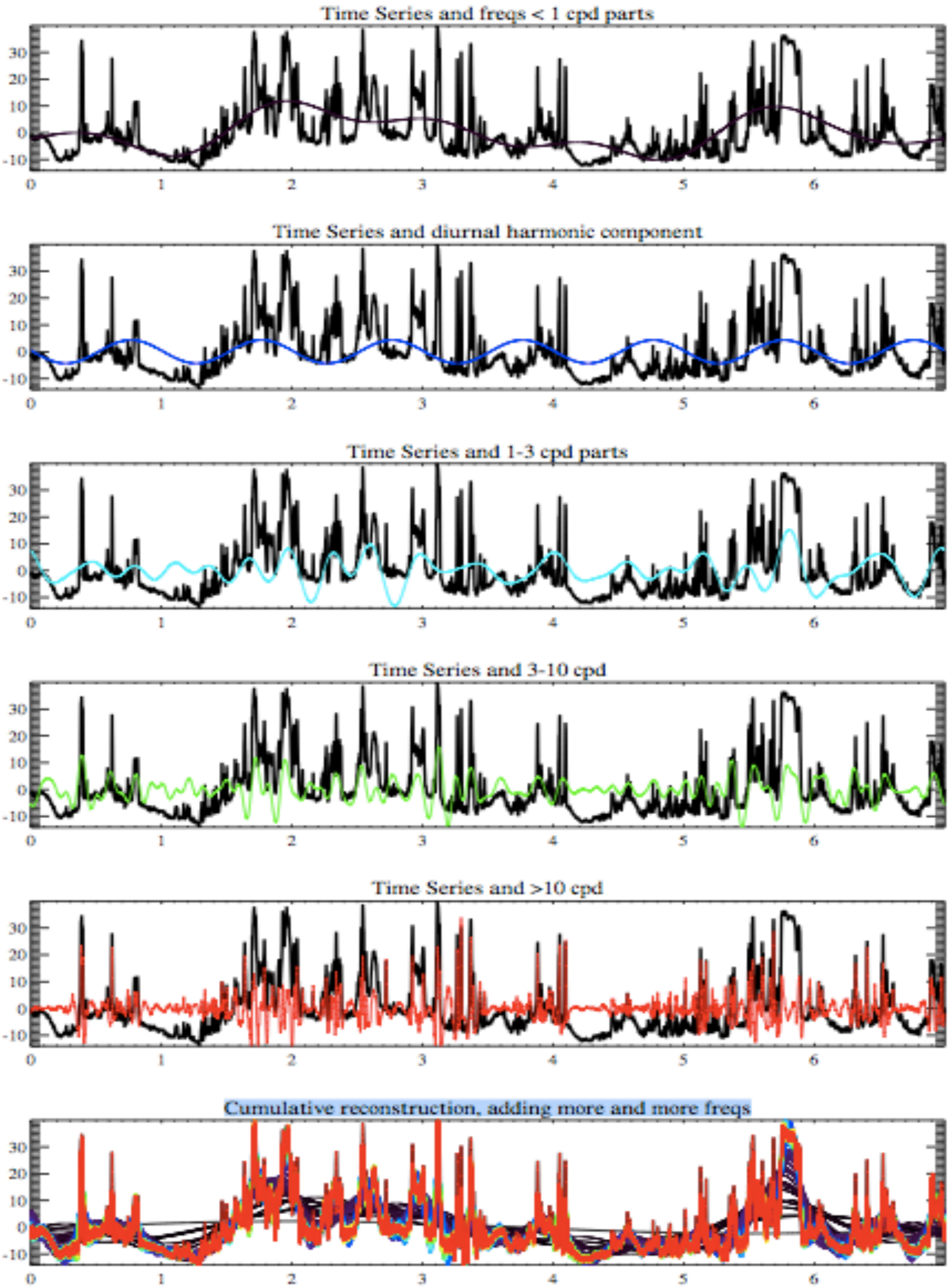


Figure 4. The total series L(t) in black, overlaid by various Fourier bands and components (colors), and then reconstructed in the last panel from the cumulative sum of all the components.

b. The phase spectrum

It is easy to forget that the power spectrum contains only half the information content in the time series. The phase spectrum is not human-readable at all, it is a random looking (*but not random*!) set of discrete values in [-π, π], one per frequency. A line plot connecting these dots would make no sense (and arguably makes little sense for power spectra too, but people do it anyway).

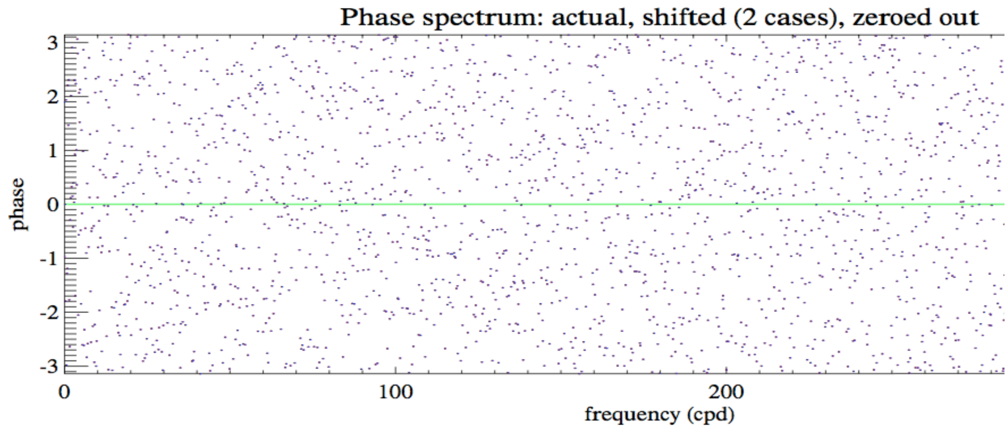


Figure 5. Phase spectrum of the time series, for positive frequencies.

The only way to see the meaning in the phase spectrum is to use it in reconstruction.

c. Reconstruction with a modified phase spectrum

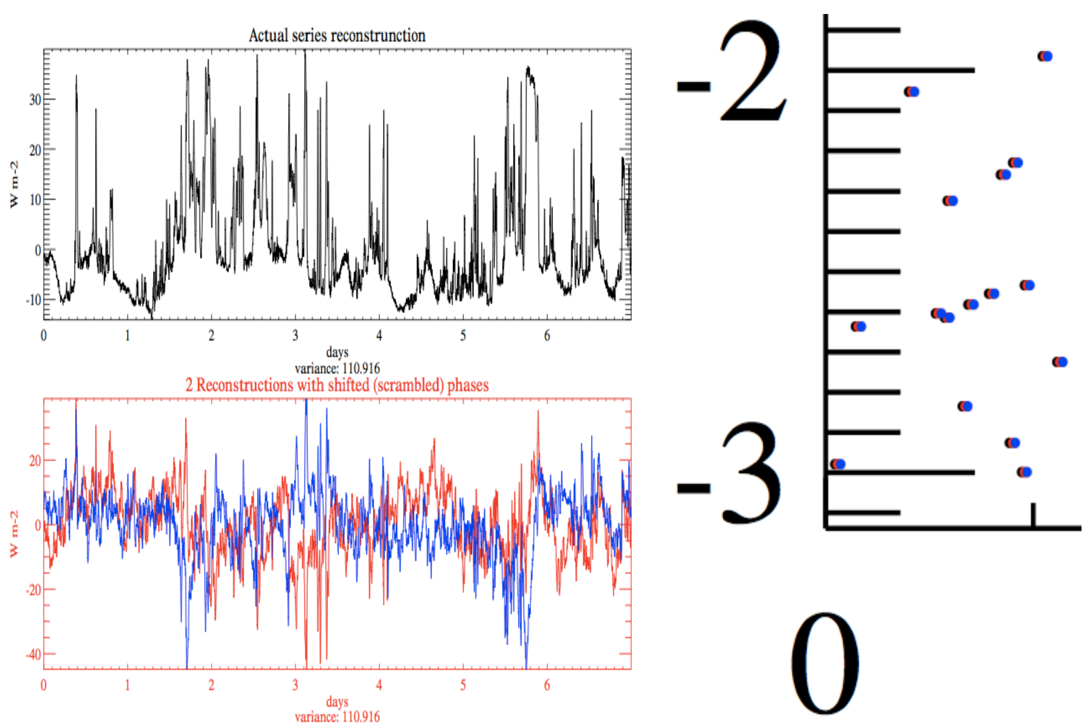


Figure 6. Left: Reconstructed time series with actual phase (black), and with the array of phase values shifted periodically by 1 or 2 frequency bins (red, blue; right panel shows a super zoom of Fig. 5).

What if we reconstruct a time series by keeping P(f), but messing up the phase information? By definition we will get a time series with the identical power spectrum, identical total variance. But the ***distribution*** may change. Figure 6 shows L(t), with its skewed distribution (Fig. 1). If random phase is assigned to the Fourier components, the pdf gets Gaussian: the reconstructed time series is now the sum of many i.i.d. variables (the sines and cosines). In other words, all that skew in the original L(t) is encoded somehow in the *exact details* of the phase spectrum, and is easily lost in phase scrambling.

The red and blue curves in Fig. 6 come from just periodically shifting the phase array by 1 and 2 positions: For red, I mis-assigned the phase of frequency 2520 to frequency 1, 1 to 2, 2 to 3, for example. Slight change, drastic result! Already the skew is lost, for instance. *Phase information is delicate and subtle.*

How about when we set all the phases to 0 (the green line in Fig. 5)? Now all the cosines interfere constructively at t=0, but they wiggles tend to cancel elsewhere. Again, the same total variance is there in the series (since phase information is independent of amplitude or Power), but now the time series is one huge spike near t=0.

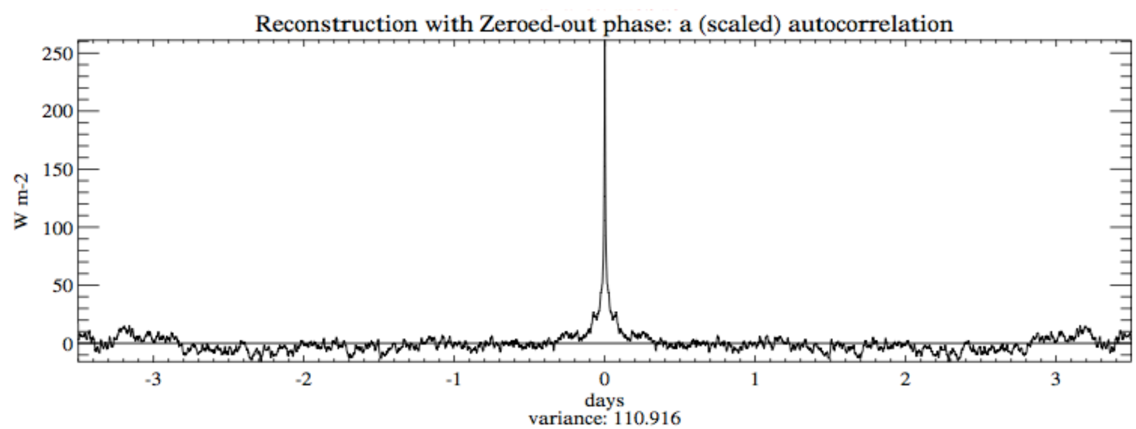


Figure 7. Reconstructed time series with all phases ϕ_i set to 0, so that $\cos(f_i t + \phi_i) = 1$ for each Fourier component at t=0. Since the time series is periodic, I shifted t=0 into the middle for clarity.

This spike is proportional to the **auto-correlation function** (whose t=0 value would be 1 without units). This is a rather profound and perhaps surprising result, relating the correlation of shifted time series to this Fourier reconstruction. (Or perhaps not surprising, since it is just mathematically true and thus maybe “obvious” to the properly educated and clear-thinking mind!)

3. Effects of degrading our data about the “true” time series above

a. Undersampling: the BIG problem of “aliasing”

Suppose we only grab a value every 3 hours. Because we sample the peaks and troughs, the total variance of this undersampled series (red, 117.564) is almost the same as the original data. But the discrete frequencies resolved in this dataset are only from 1 cycle/week to 1 cycle/6h. So we know already that the power in the higher frequencies are somehow being improperly mapped “aliased” into the resolved part of the spectrum.

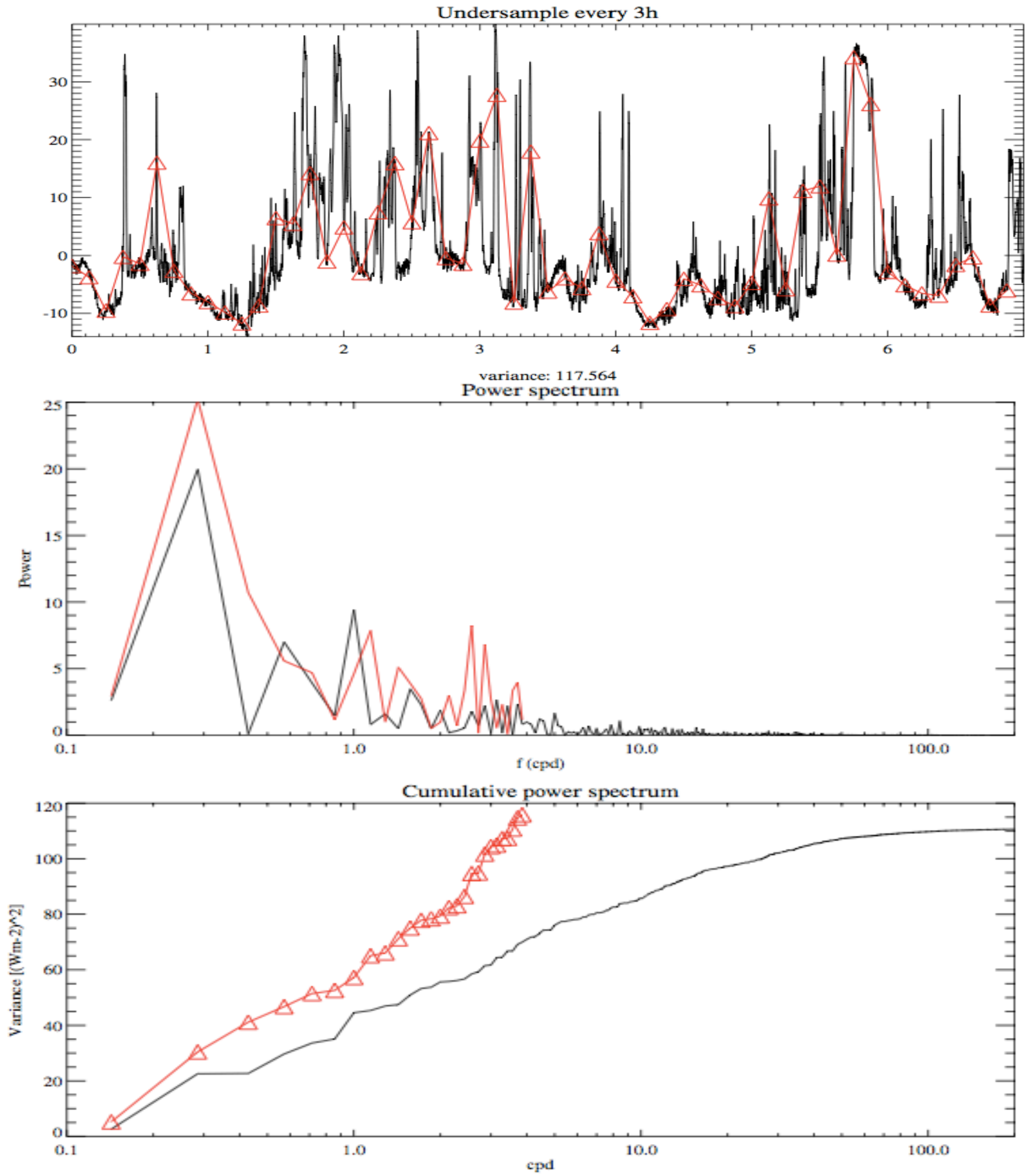
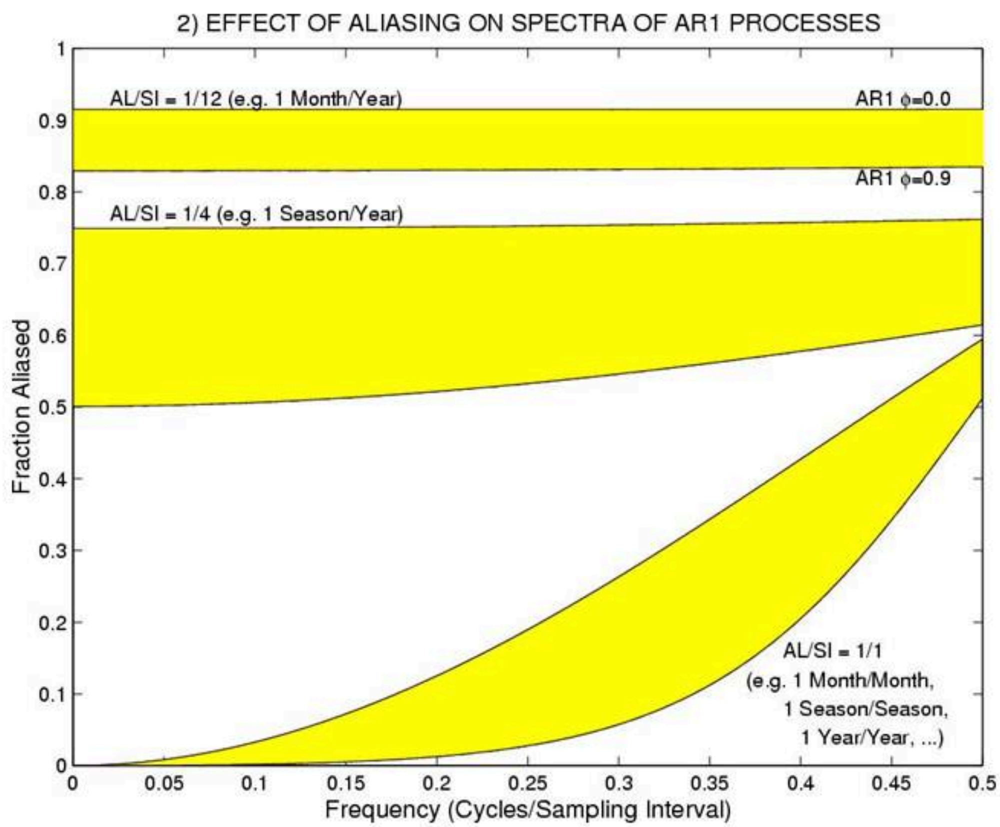


Figure 8. Undersampled data: a value is grabbed every 3 hours. The sample range and thus variance is as large as the original data, but now there are no frequency "bins" higher than 1 cycle/ 6h. Somehow, extra variance is being shifted or "aliased" into the low frequencies.

The undersampled series in Fig. 8 (red) has more power at most frequencies (aliasing adds power, and it has to go somewhere). The diurnal peak gets moved by one frequency interval bin (from 7 cycles/7days = 1 cpd to 8cycles/7days). Perhaps this is because the high values in the latter part of already-peculiar day 6 got missed by the sampling -- see the 2nd panel in Fig. 4 for the pure diurnal harmonic.

Aliasing is a surprisingly powerful force for generating nonsense! Roland Madden sent me this picture in Jan 2013 (from Madden, Roland A., Richard H. Jones, 2001: A Quantitative Estimate of the Effect of Aliasing in Climatological Time Series. J. Climate, 14, 3987–3993.)

The plot below gives results for white noise (lag-1 autocorrelation ϕ =0.0) up toward red noise (ϕ =0.9). Data are sampled with an interval factor (AL/SI). The yellow bands show what fraction of your diagnosed variance is aliased (nonsense) rather than real. Suppose SST anomalies have a 0.0-0.9 range of lag-1 autocorrelations, and we sample only one month per year (like a 'climate' time series of September-only anomalies in the hurricane MDR). The top yellow bar shows that *80-90% of the variance in such a time series, at all frequencies, is spurious, due to aliasing!*



AL = averaging length, T ; SI = sampling interval

Figure 9. Madden, Roland A., Richard H. Jones, 2001: A Quantitative Estimate of the Effect of Aliasing in Climatological Time Series. J. Climate, 14, 3987–3993.

b. Benefits of averaging instead of sampling

Using averages rather than samples every 3 hours is much better. There is less total variance, since high frequency fluctuations are damped by the averaging. The spectrum in the resolved frequencies is almost exactly right (red spectra on top of black).

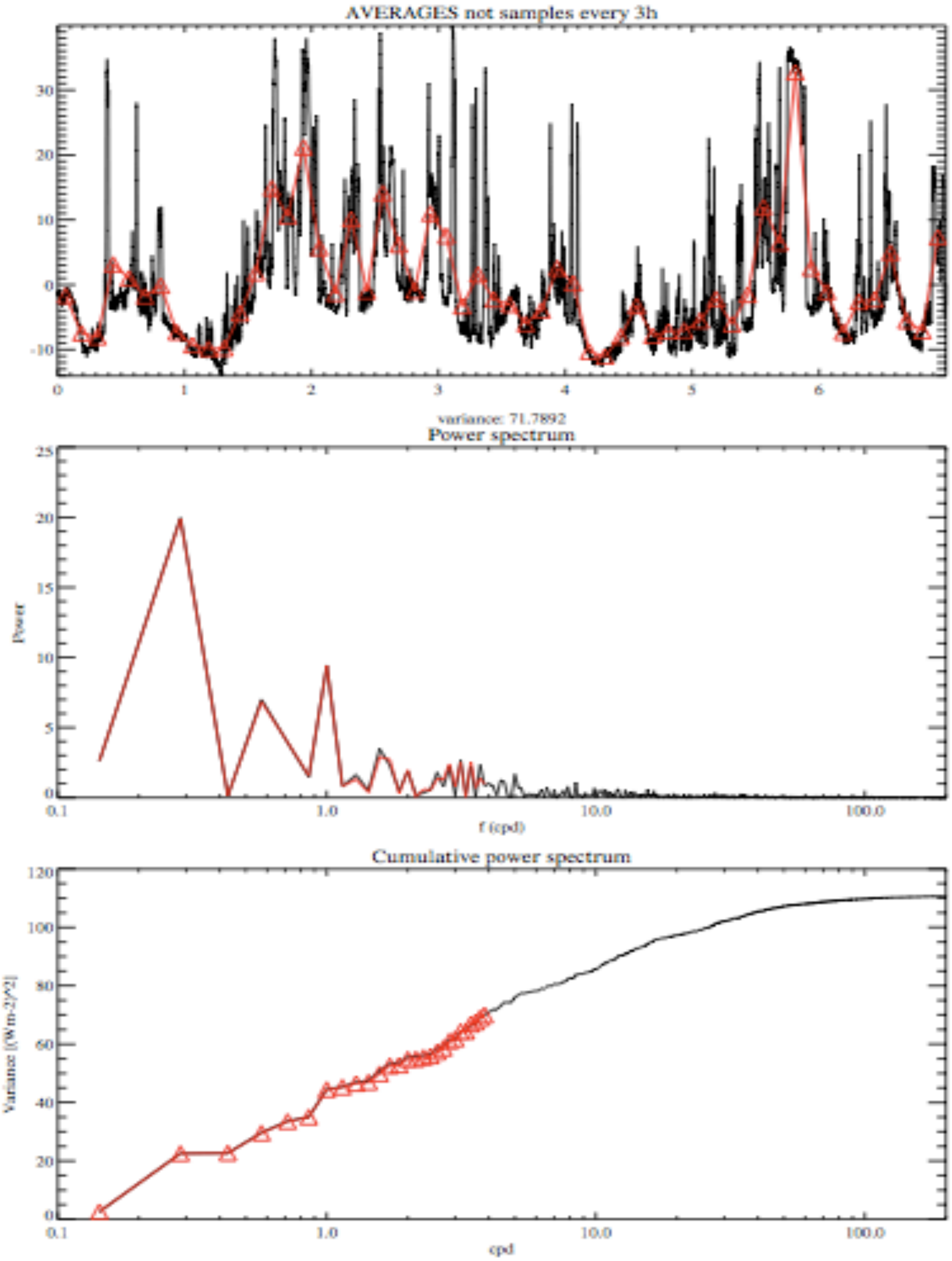


Figure 9. Spectrum from 3-hour averaging of data.

c. Quantizing

Now let's quantize the data (red): only 5 different values occur (-10,0,10,20,30,40).

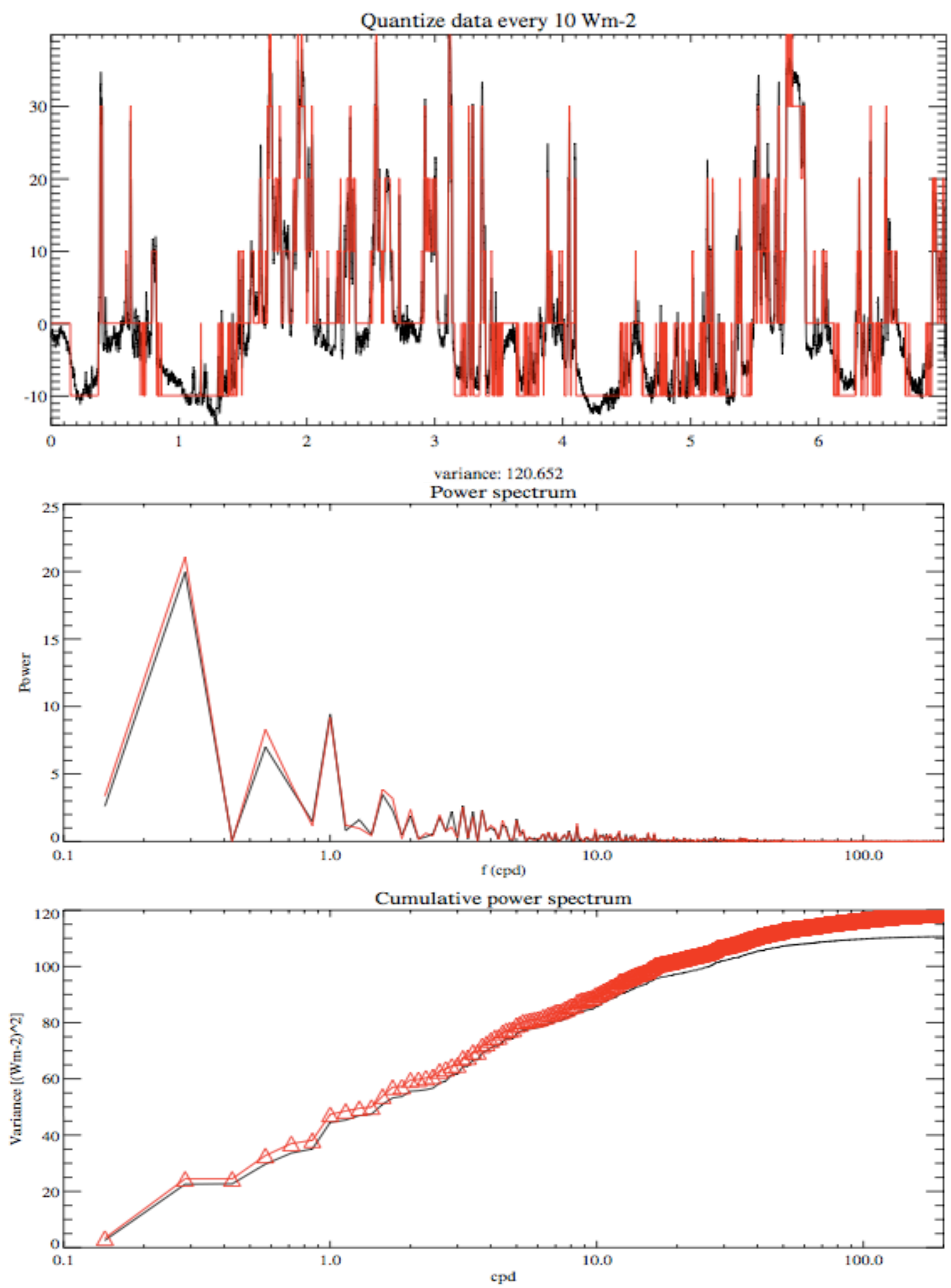


Figure 10. Spectrum from coarsely quantizing the data values.

The total variance is almost unchanged (increased a bit in this case). How about the spectrum? Nothing drastic: a bit of spurious power, much of it at f> 10 cpd, perhaps associated with the sharp edges of the square peaks and valleys?

d. How about a finite data record? Which segment do you happen to sample?

Here are the first 2 days. There is 117 W^2 m^-4 of variance, but much of that is in a big trend, so treating this segment as periodic (as xhat = fft(x) implies) gives a big spurious 0.5 cpd signal. This is why spectral analysis usually starts with “detrending” the data. (Estimating a trend can be done various ways, I like just regressing x on t and removing that linear regressed part before taking the fft.)

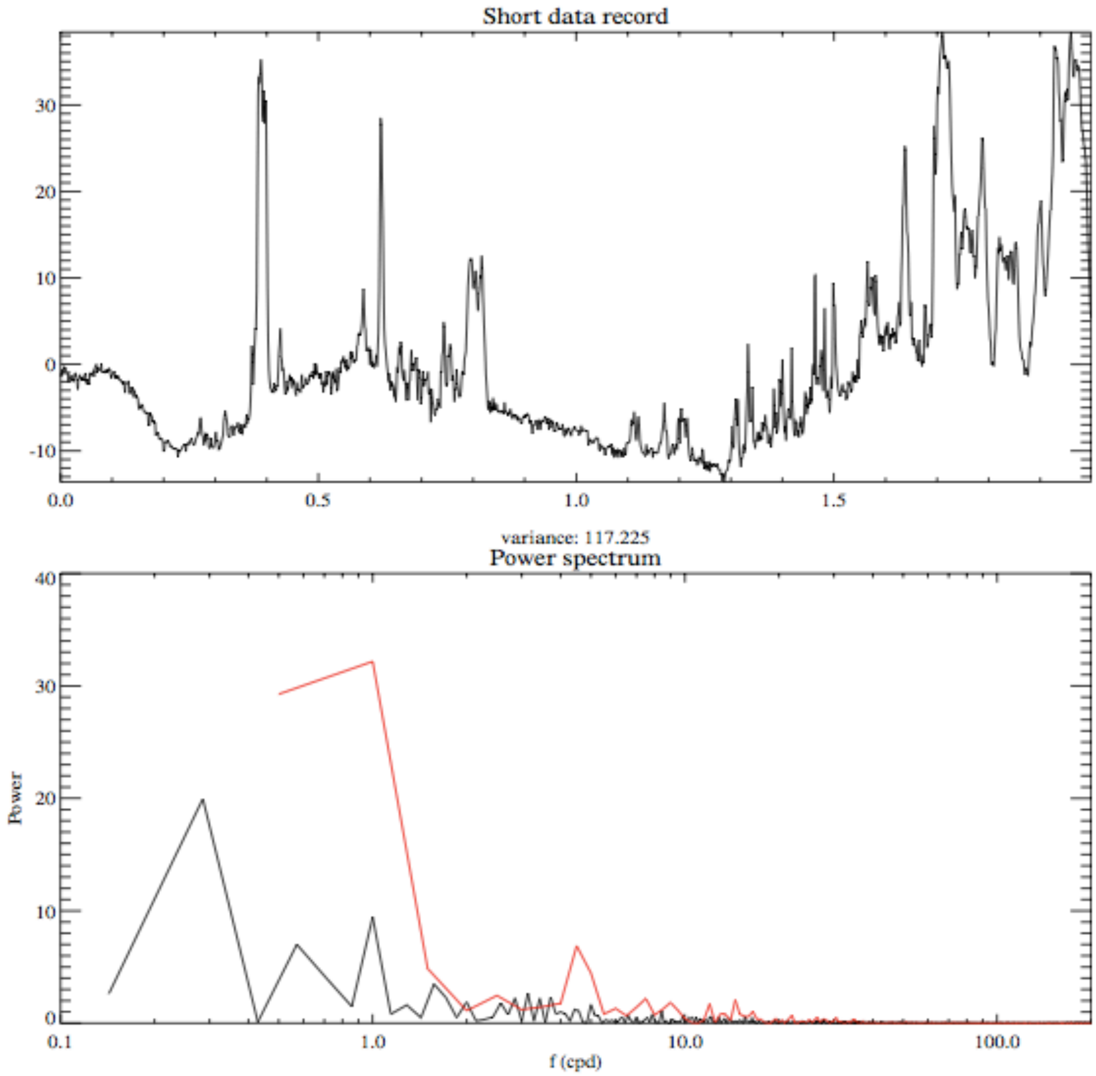
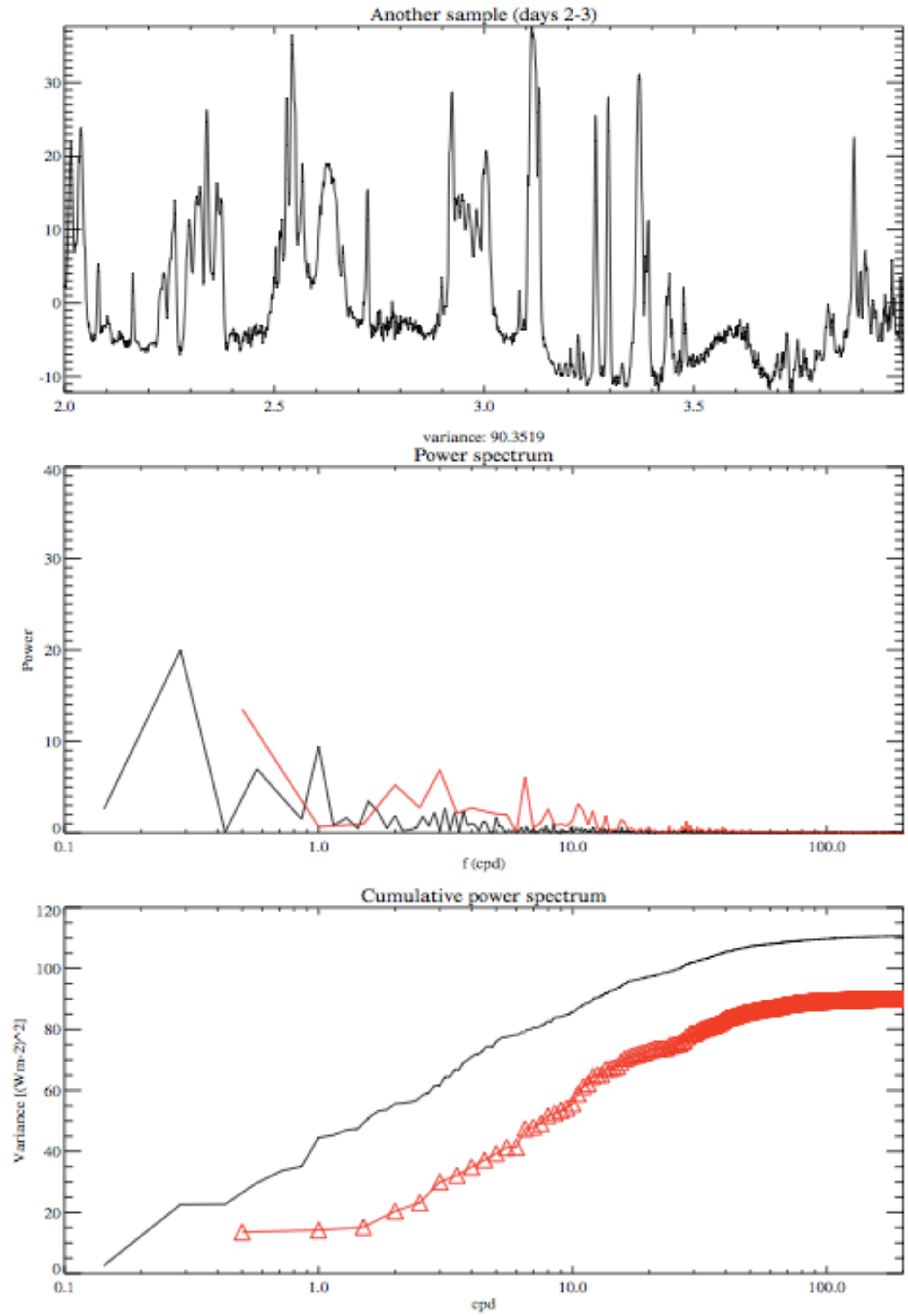


Figure11. Spectrum of first 2 days of data, treated as infinitely periodic time series.

Looking only at days 2-3 give a very different picture: less trend (but still some), and no clear diurnal power. Total variance is only 90 W^2 m^-4 in this case.



4. The convolution theorem and padding/windowing/smoothing:

Convolution of 2 functions x(t) and y(t) is an integral of their product.