Fourier crash course for MPO/ATM 624
Brian Mapes April 2020

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}.

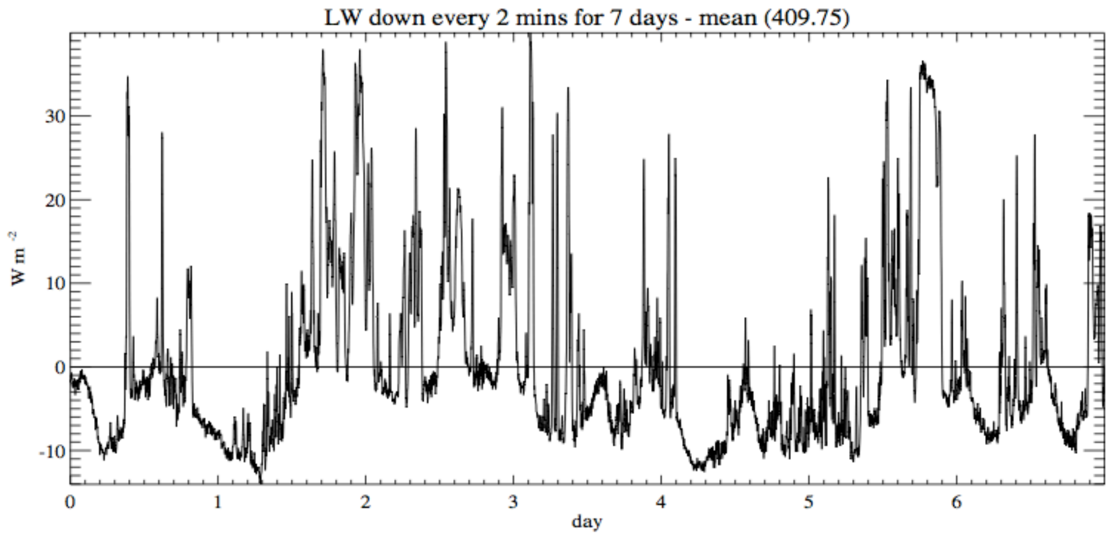


Figure 1. A time series L(t) of downwelling LW radiation on the equator, with its mean removed. The variance is 110.9 ($W\ m^{-2}$)^2, so the standard deviation is 10.5 $W\ m^{-2}$.

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 implicitly assuming that the data sequence is repeated periodically to infinity*. Hmm. Within that dubious 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) = atan2(imaginary/real). The square of amplitude is called Power P(f), apparently because it is like an energy (a squared thing). Historical naming.

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 (2π/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, below I divide these integers by 7.0 to express frequency units as **cycles per day** (**cpd**). Units of (cycles/time-interval-you-careabout) will save you and your readers a lot of confusion with Hz, radians, etc. which some people insist on using (sigh…).

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 (1 cycle)/(2 dt) 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):

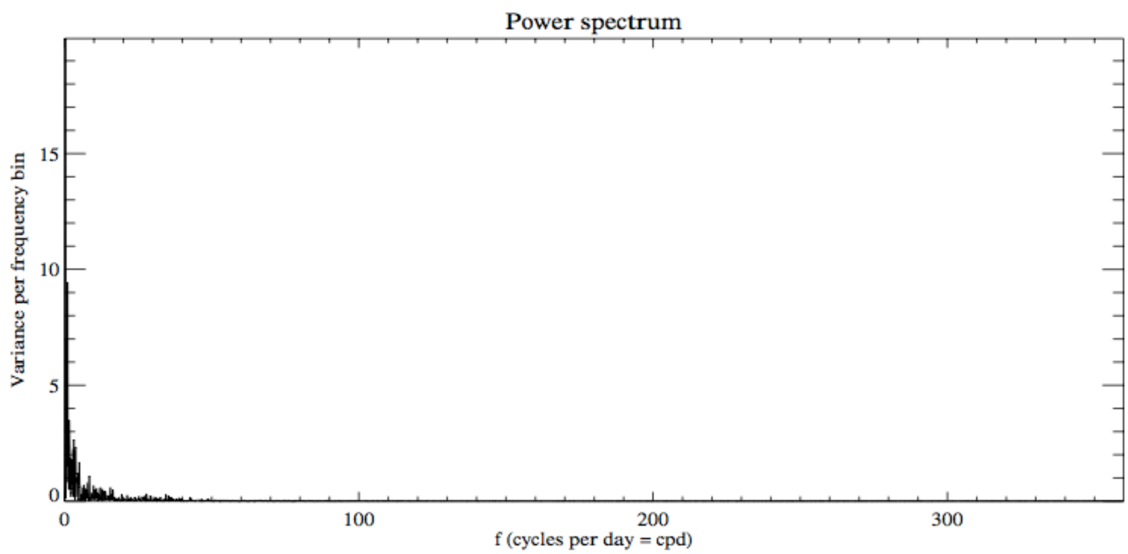


Figure 2. The power spectral density of Fig. 1. Only positive frequencies are shown. Sometimes a raw spectral power plot is called a "periodogram", a charming old timey word. Graphically terrible.

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

Parseval's Theorem (https://en.wikipedia.org/wiki/Parsevals_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), here 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 'variance = area under the curve' aspect is lost. A trick to retain that aspect is to plot fP(f) on a x-log axis, since $\int fPd\ln f = \int Pdf$.

I like to plot the **cumulative variance** ΣP(f), which asymptotes to the total variance (a nice clear illustration of Parseval's theorem and graphical check on scaling). This can be plotted against log (f) which helps emphasize the lowest frequencies, or log(period) so that the x axis labels are period instead of frequency. Cumulation 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 our cumulative spectrum with the x axis carefully offset by half a bandwidth for clarity. Reading off from about halfway up the y axis, we see that 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 distinct step up at 1.0 cycles per day indicates a peak in the spectrum at 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 ~ 0.2 cpd) 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?

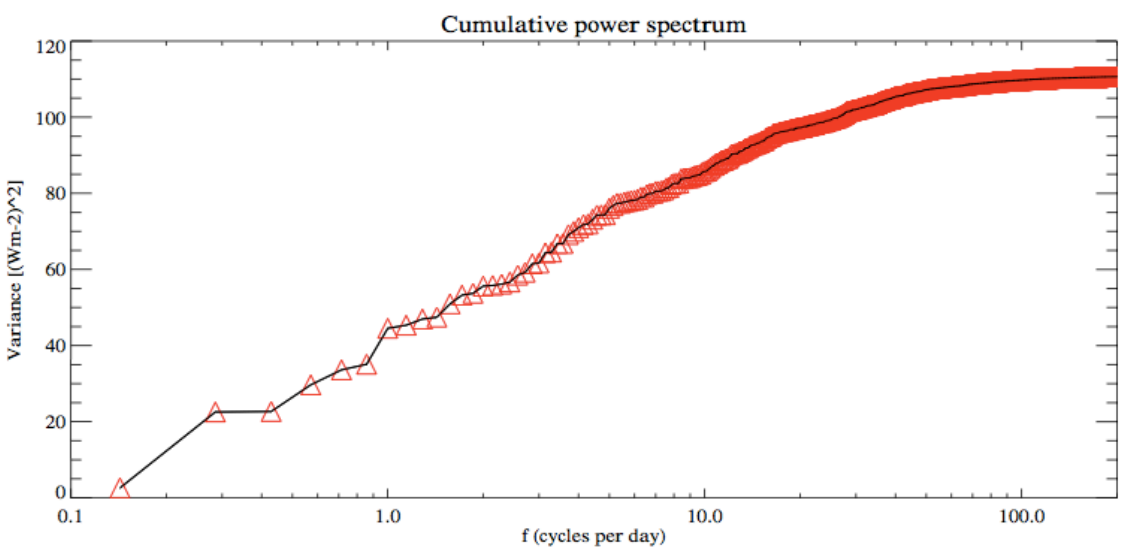


Figure 3. The cumulative power spectral density of Fig. 1. Only positive frequencies are shown, I doubled the spectrum to get this to reach the right total – a good easy check on methods.

Since fft(L) is reversible, ifft(fft(L)) = L, no information has been lost. In other words, we can reconstruct L(t) by adding up all the Fourier components (called **harmonics**). If only certain frequency bands are included, and others excluded (multiplied by zero) the reconstruction is called a **Fourier filtering** of the data.

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

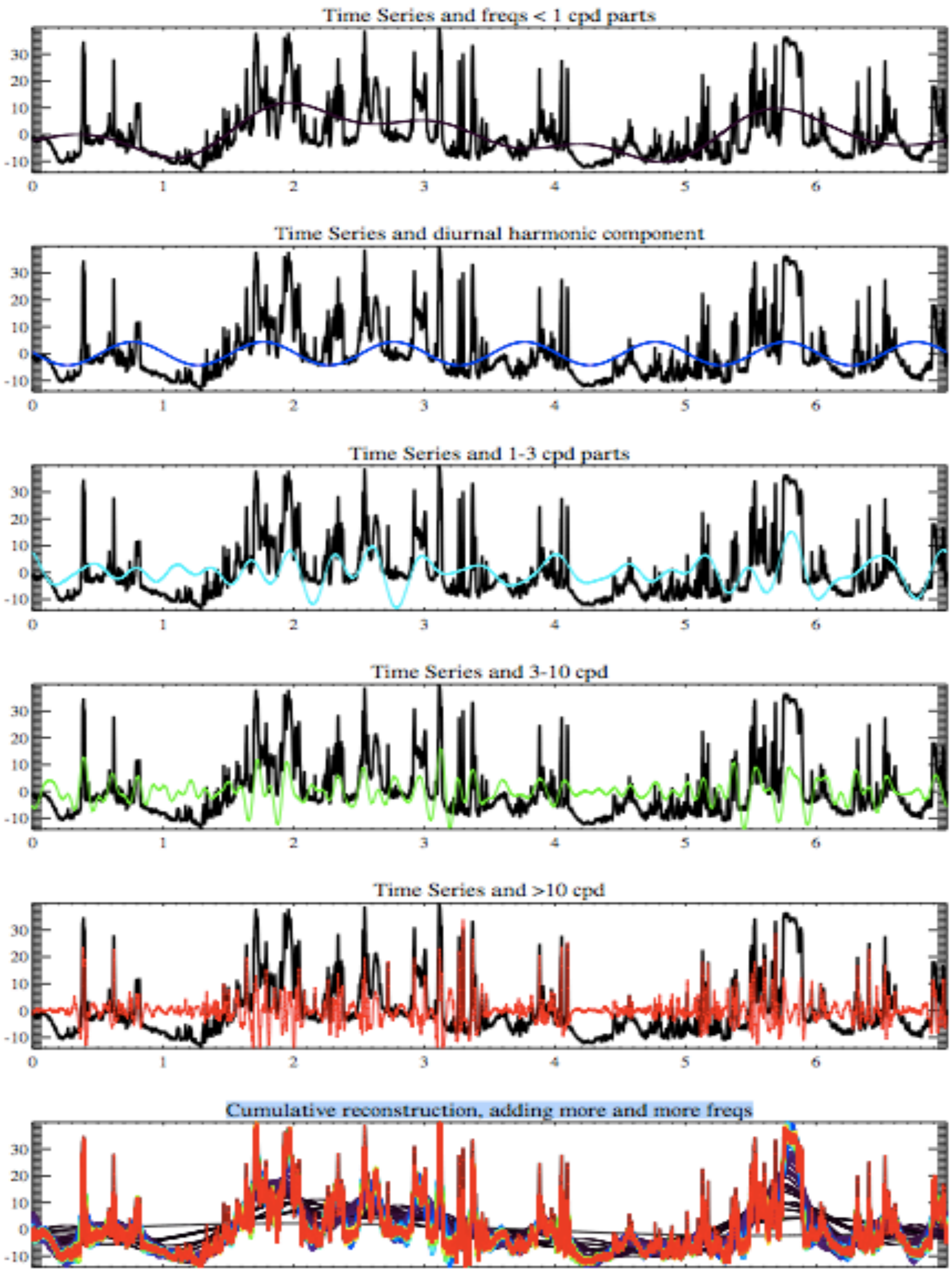


Figure 4. The total series L(t) in black, overlaid by various Fourier bands and components (colors), and then reconstructed in the last panel from the cumulative sum of all the components.

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, which should be discrete like a bar plot, but people do that anyway).

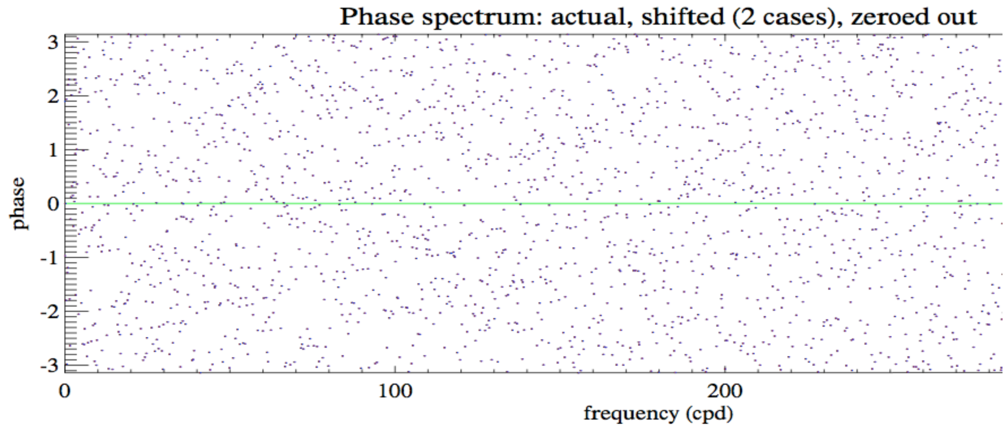


Figure 5. Phase spectrum of the time series, for positive frequencies.

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

c. Reconstruction with a modified phase spectrum

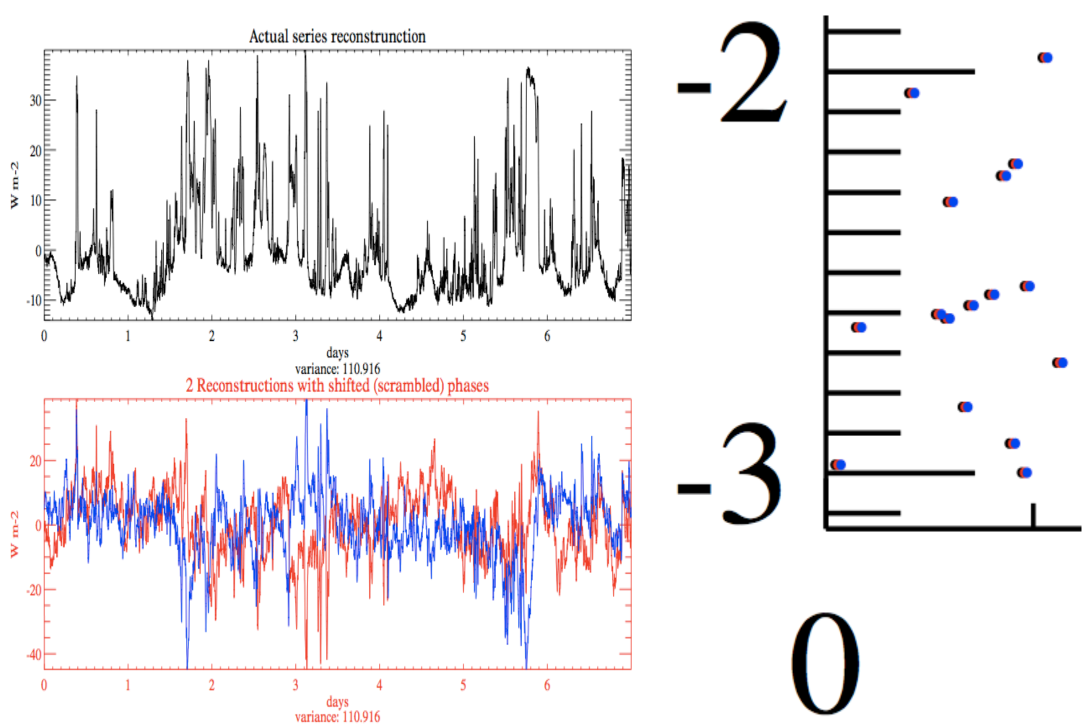


Figure 6. Left: Reconstructed time series with actual phase (black), and with the array of phase values shifted by 1 and then 2 frequency bins (red, then blue; right panel shows a super zoom of Fig. 5).

Fig. 6 shows what happens 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, and thus identical total variance. But the ***distribution (PDF)*** can change. Figure 6 shows L(t), with its skewed distribution (repeating Fig. 1). If *random* phase is assigned to the Fourier components, the PDF becomes Gaussian, by the Central Limit Theorem, since the reconstructed time series is now the sum of many i.i.d. variables (the sine and cosine components). 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 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, and so on. Slight change, drastic result! Already the skew is totally 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.

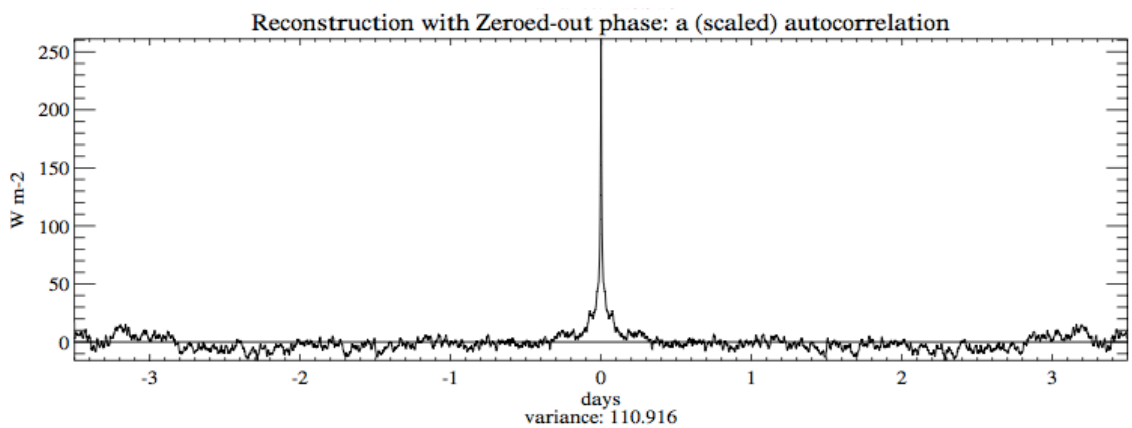


Figure 7. Reconstructed time series with all phases ϕ_i set to 0, so that $\cos(f_i t + \phi_i) = 1$ at t=0 for each Fourier component i. Since the time series is periodic, I can shift t=0 into the middle for clarity.

This spike is related to the **auto-correlation function** https://en.wikipedia.org/wiki/Autocorrelation), which has the same power spectrum as the time series itself: (https://en.wikipedia.org/wiki/Wiener-Khinchin_theorem).

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 this will sample the peaks and troughs with total amplitude, the total variance of this undersampled series in the line with triangles, 117.564 $(Wm^{-2})^2$, is essentially 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 (variance) in the higher frequencies must somehow being improperly mapped or "aliased" into the resolved low frequencies of the spectrum.

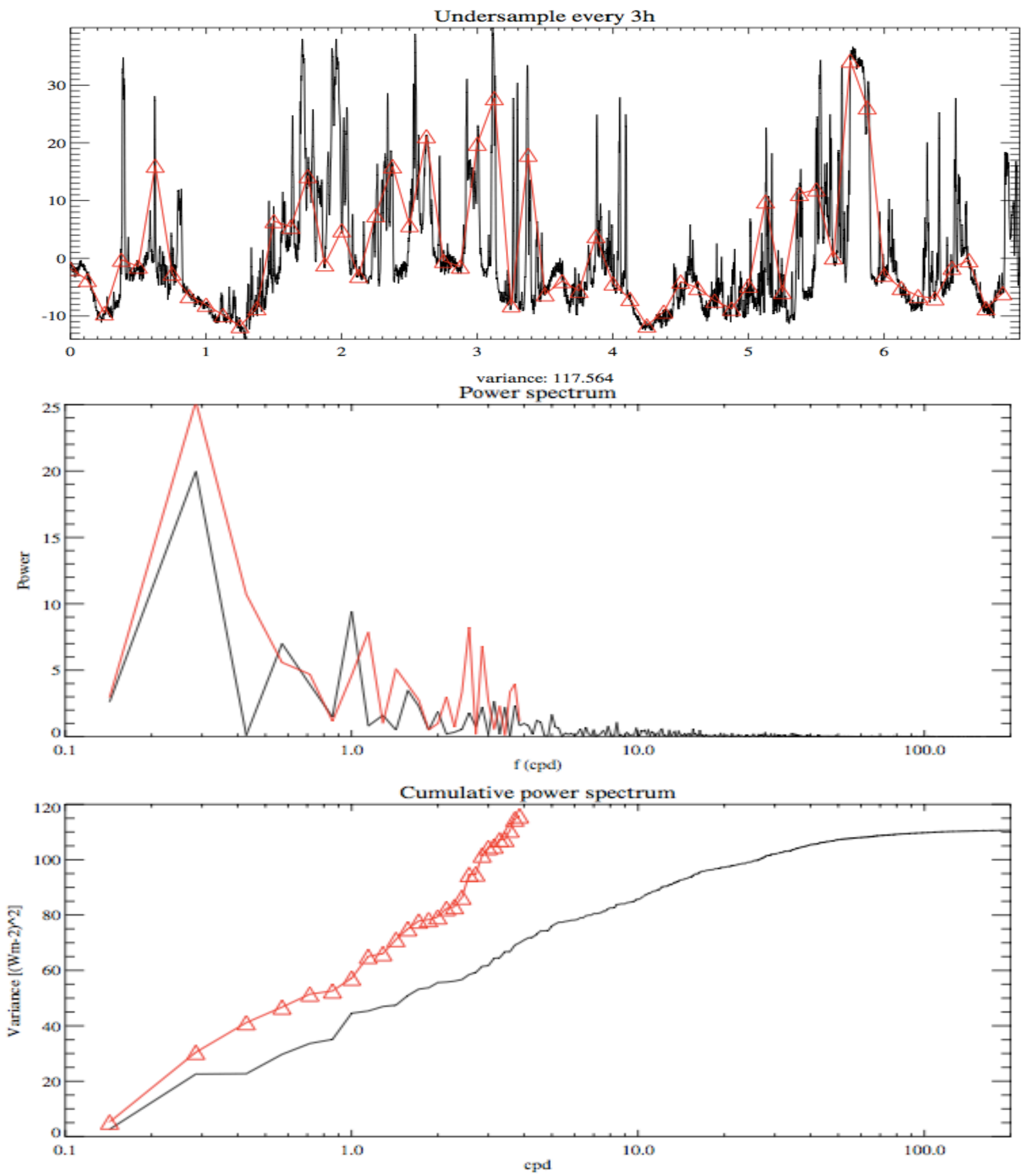
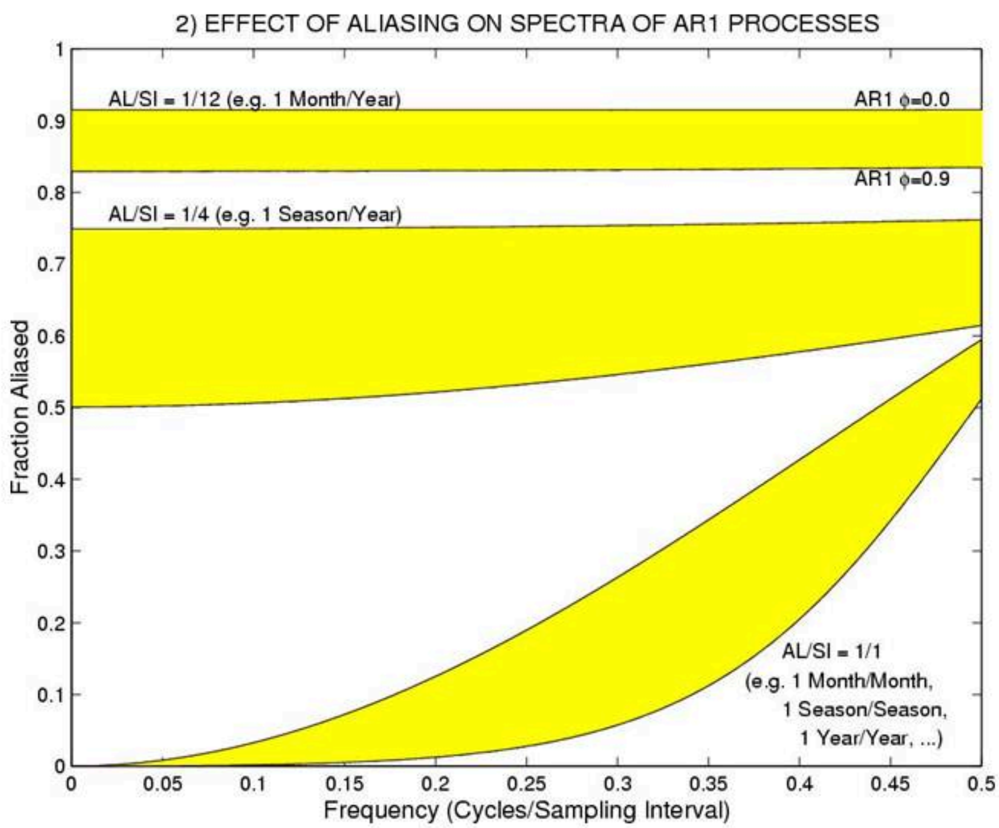


Figure 8. Undersampled data: a value is grabbed every 3 hours. The sample range and thus variance is just as large as the original data, but now there are no frequency "bins" higher than 1 cycle/ 6h. Somehow, extra variance must therefore be shifted or "aliased" into the low frequencies.

The undersampled series in Fig. 8 (red) has more power at most frequencies. The diurnal peak got moved by one frequency interval bin (from 7 cycles/7days = 1 cpd to 8cycles/7days). Evidently this is because the high values in the latter part of day 6 got missed by the sampling (the 2nd panel in Fig. 4 shows the pure diurnal harmonic).

Aliasing is a surprisingly powerful force for generating nonsense in spectra! Figure 9 gives results for colored AR1 noise from 'white' (where autocorrelation at lag=1 AR1 is ϕ =0.0) to very 'red' (ϕ =0.9). Data are sampled at an interval (SI), with a given Averaging Length (AL) that comprises each data "point." The yellow bands show what fraction of your diagnosed variance at low frequencies is aliased (spurious) rather than real.

Suppose we sample SST anomalies from only one month per year, to make a 'pseudo-climate' time series of September-only anomalies in the hurricane MDR for instance. The top yellow bar shows that *over 80% of the variance in such a time series is spurious, due to aliasing!* This is dangerous for seasonalized climate science, obviously.



AL = averaging length, T ; SI = sampling interval

Figure 9. Fraction of spectal power that is spurious (aliased) as a function of frequency and AL/SI, for lag-1 autocorrelation values 0.0 to 0.9. 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.

b. Benefits of averaging instead of sampling

Using averages rather than samples is much better. It should be obvious that there is less total variance in averages, since high frequency fluctuations are damped by the averaging. The spectrum of 3h averages in the resolved frequencies is almost exactly right:

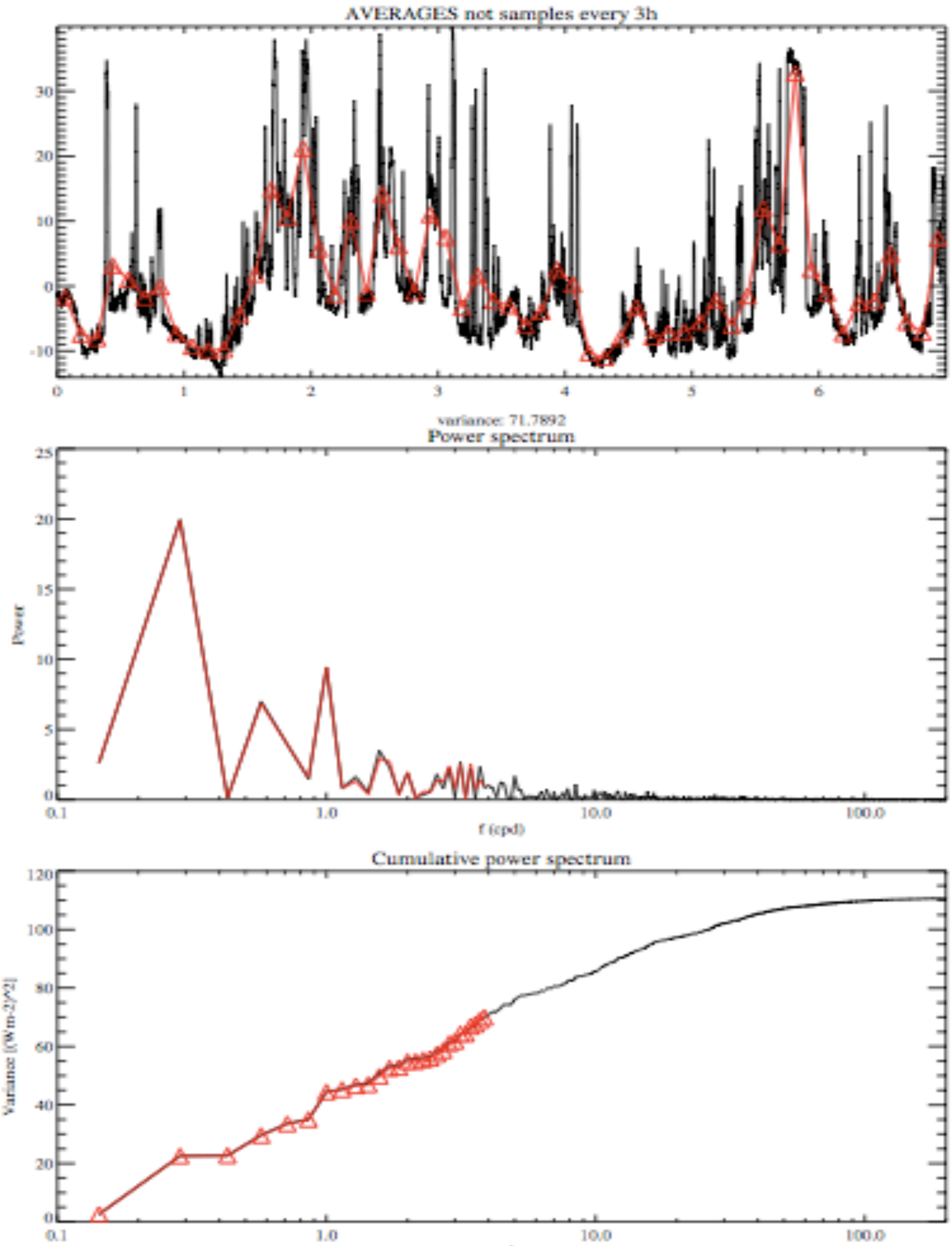


Figure 9. Time series and power spectrum from 3-hour averaging rather than sampling. Total variance is less, and is properly apportioned. No aliasing!

c. Quantizing the data

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

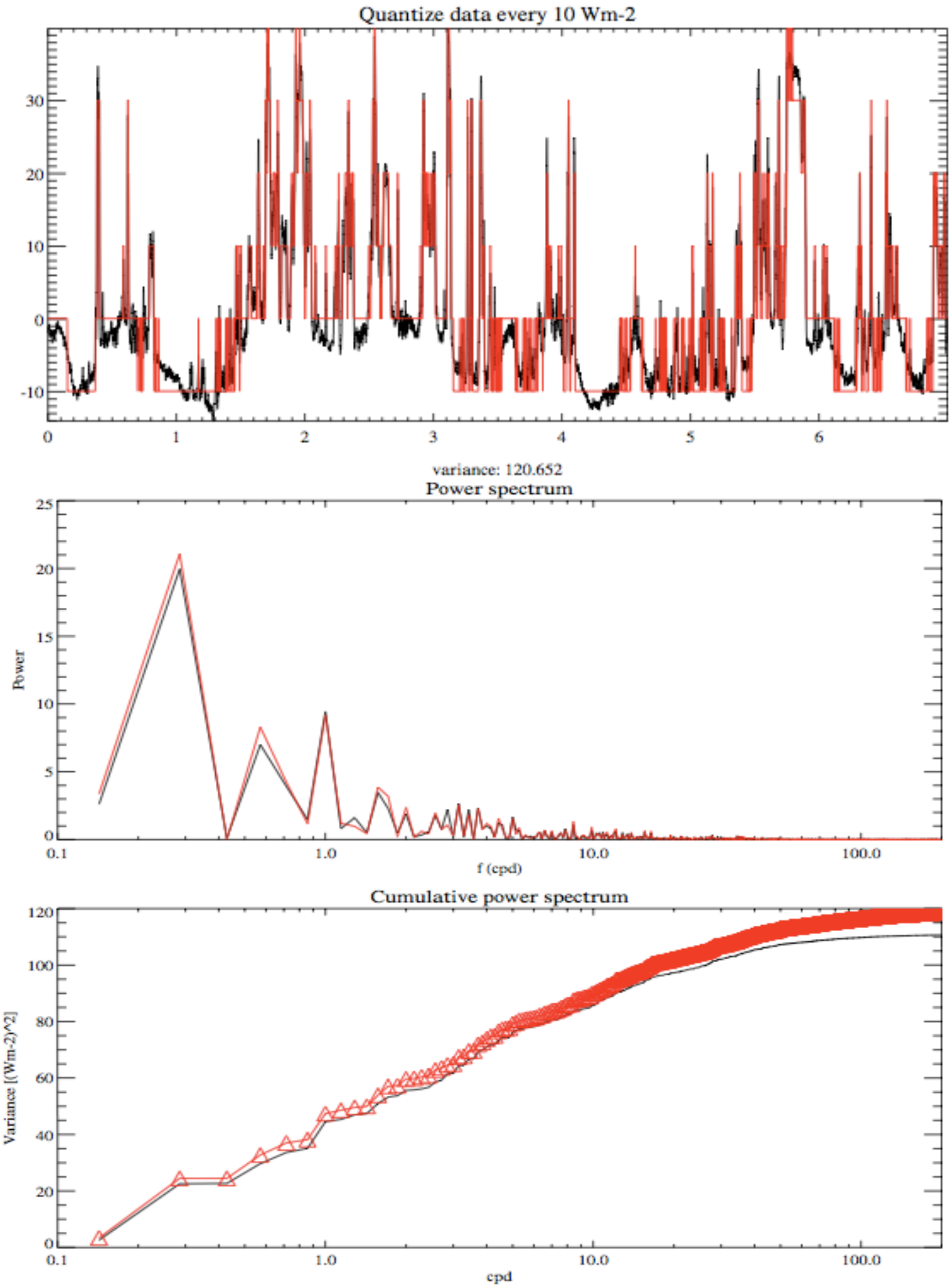


Figure 10. Spectrum from coarsely quantizing the data values.

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? Well, which segment do you happen to sample?

Here are the first 2 days (Fig. 11). The lowest frequency possible is now (1 cycle)/(2 days), so the spectrum (red) is now clipped at its *left* end. There is 117 W^2 m^{-4} of variance, but much of that is in a trend, so treating this segment as periodic which fft(L) implies produces big spurious lowest-frequency (0.5 cpd) power, part of the implied "sawtooth" pattern as the trend is repeated in each segment of the infinite periodic series. Spectral analysis usually starts with "detrending" the data for this reason. (Estimating a trend can be done various ways; I like just regressing L on t and removing that linear regressed part before taking the fft. Or there are routines; detrend(L) in Matlab.)

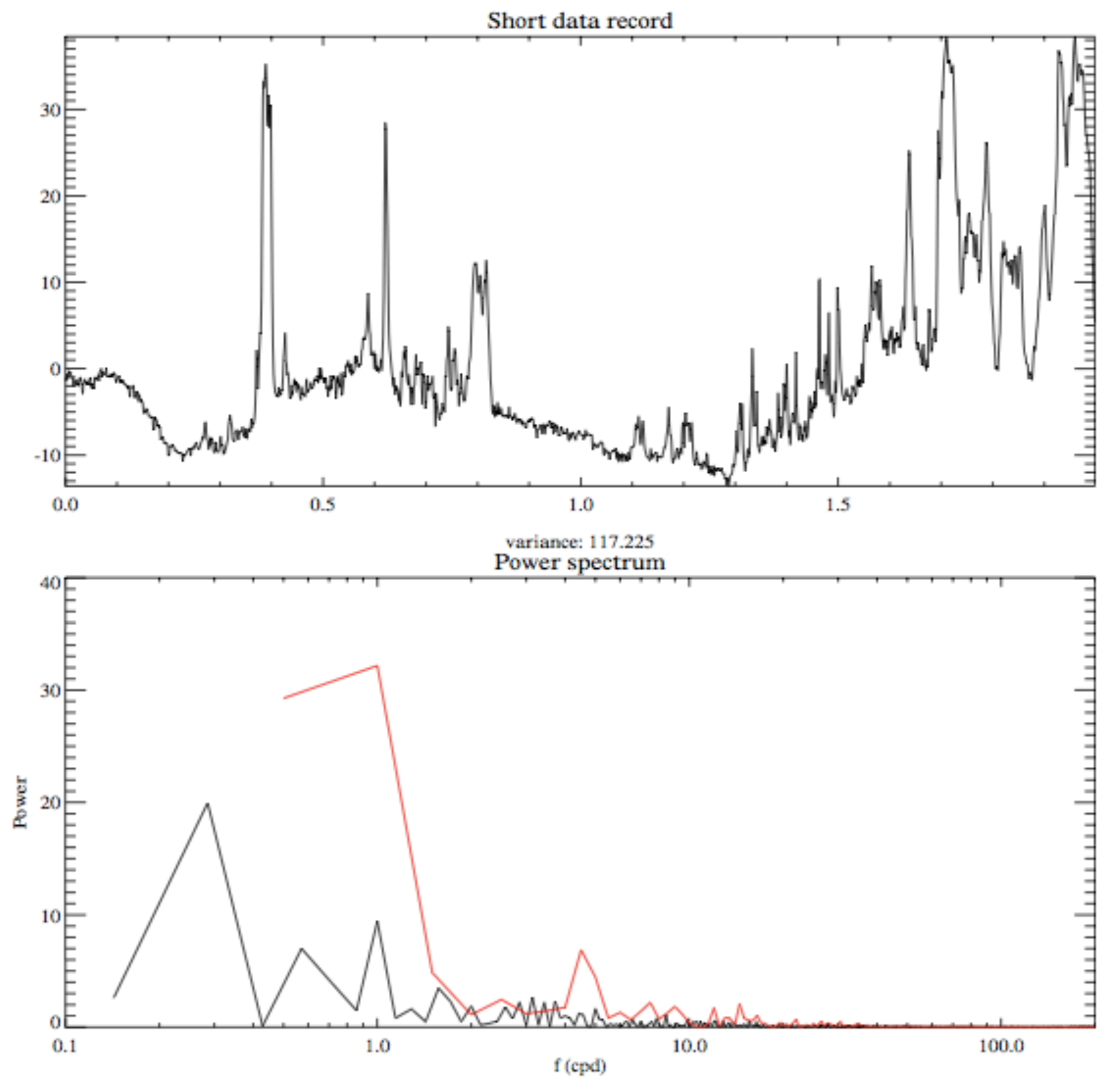


Figure 11. Power spectrum of first 2 days of data (red), compared to the true one (black). The lowest frequency has large spurious power because of the sawtooth pattern implied by the 0-2day trend.

If we instead sample days 2-3 (Fig. 12), we get a very different picture: There is less trend (but still some, making the lowest frequency partly spurious). There is no clear diurnal power in this case. Total variance is only 90 W^2 m^-4 in this case. What's the main spurious spectral peak you see? Can you match the red spectral peaks to a sketch of wiggles in the time series?

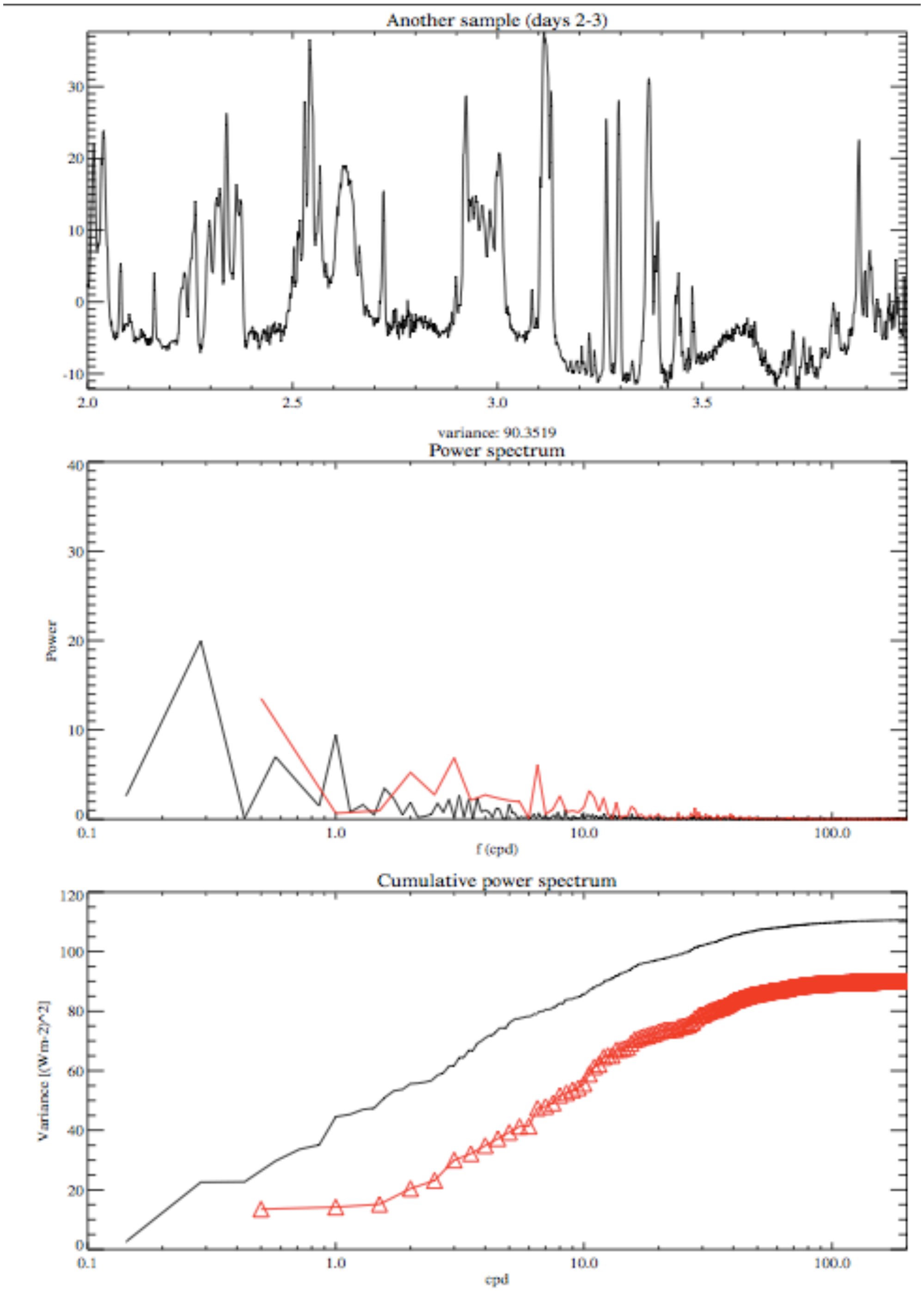


Figure 12. Spectrum of the second 2 day subset, treated as an infinitely periodic time series.

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

Convolution of 2 functions is a *definite integral of their product*:

$$\theta(t) = \int_{-\infty}^{+\infty} g(\tau)h(t-\tau)\,d\tau$$

$$\theta = g * h$$

a. Convolution as smoothing: the concept of a *kernel* function

A powerful use of convolution comes from having one function be your data and the other be a "kernel" or "weighting function" or "window" or "sampling mask". For example, thinking about a sampling mask made of equally spaced delta functions is the math that connects continuous Fourier analysis to its discrete cousin. Another use is a *smoothing kernel*, a function with finite values around τ=0 but decaying away to 0 for large | τ |. This might be a square "top-hat" or "boxcar" function, or a Gaussian or other bell curve, or a "1-2-1 filter", which is a narrow boxcar stacked on a wider boxcar (plotted below).

The convolution of a time series h(t) with a smoothing kernel g(τ) can be thought of as replacing each data point on the graph of h(t) with a little copy of the kernel function g(τ), centered at that time t and scaled by the magnitude x(t). Now add up all these kernels, and you get a smoothed time series. In basic calculus, you thought of an infinitesimal area x(t)dt under a graph as a narrow square "tower". If you are convolving h(t) with a bell shaped g(τ) kernel, you instead replace this little tower at each value of t by a bell curve with the same area under it. Sum up all the little bell curves, and what do you get? A smoothed (smeared) version of h(t).

Figure 13 shows a graphical example of the above idea.

(*Forgive the small mis-offsets of colored bars relative to vertices of the dotted time series; a graphing software annoyance*).

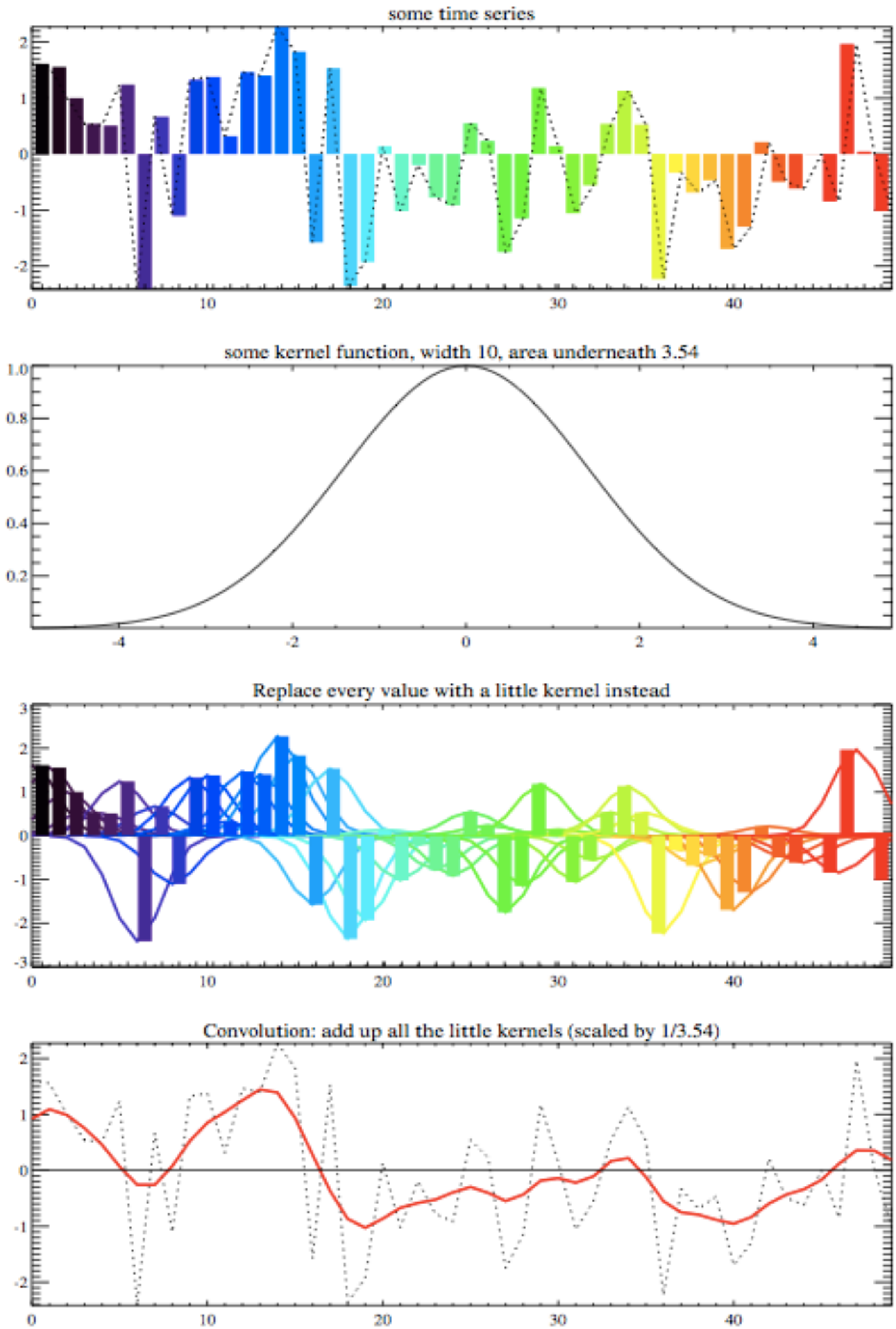


Figure 13. Illustration of the convolution of a random time series in [0,50] (dotted curve and bars) with a Gaussian kernel of half-width 2 (solid curve) to get a kernel-smoothed time series (red).

b. The Convolution Theorem links spectral and physical spaces:



I learned this profound fact from the lectures and notes of Dr. Kevin Leaman (1949-2015), teacher of our predecessor course for many years.

The convolution theorem:

The result is that convolution in the time domain corresponds to multiplication in the frequency domain. The converse is also true.

Let's use this profundity to revisit the problem of "**spectrum estimation**" – trying to understand the error bars in the *frequency domain* that result from various imperfections and finiteness problems with the data in the physical or *time domain*. We saw that taking a *finite time segment* of length T makes the power spectrum *discrete*, with only *integer* frequencies 1,2,3... cycles/T. **Padding** that with zeroes at both ends, out to infinity, would make the spectrum more continuous. What are the effects of padding?

i. Padding the time series:

A zero-padded series is the product of an infinitely repeating set of copies of our time series L(t) on [0,T] (Fig. 1 above), multiplied by a boxcar(t) masking function which is 0 everywhere except on [0,T], like the top panel of Fig. 14. Transforming this product into spectral space, the *multiplication becomes a convolution*. In other words, the spectrum is convolved with (or smoothed by) a kernel that is the Fourier transform of the boxcar function.

TIME: Zero-padded series	=	periodic L(t)	***times***	boxcar(0-7 days)
FREQ: Power(" ")	=	$\vert\mathcal{L}(f)\vert^2$	***convolved with***	FT(boxcar)

What does that kernel look like? bottom panel:

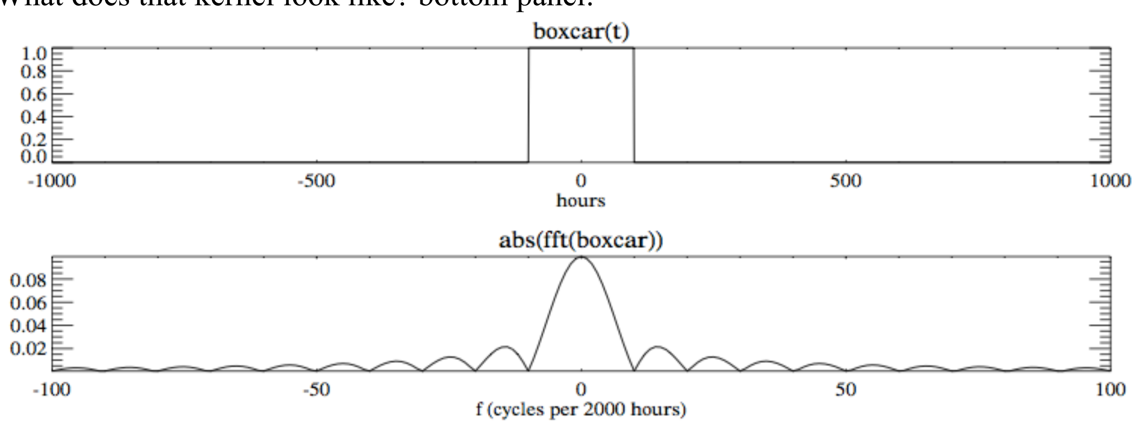


Figure 14. The boxcar function and its Fourier amplitude spectrum. The nonzero mean of boxcar(t) corresponds to nonzero power at frequency=0 in the transform (spectrum). The side lobes have alternating sign, hidden by the abs() operation.

So if we pad our 7-day L(t) series from Fig. 1 with zeros, the spectrum of the resulting longer or infinite series is the spectrum (Fig. 2) *convolved with* (smeared or smoothed by) the *kernel* function whose absolute value is in the lower panel of Fig. 14 (note: the lobes oscillate in sign). In other words, at every frequency in the spectrum of Fig. 2, you replace each local spectral power value with a little copy of the kernel centered there, and add them up, as done in Fig. 13. A sharp peak in the spectrum of the repeated periodic L(t) would become a smeared peak, with "echoes" from the side-lobes of Fig. 14 further spreading its power across the frequency domain. This smearing is called "spectral leakage". The "no free lunch" principle states one can't gain more spectral resolution by padding with zeros, unsurprisingly!

i. Windowing or tapering the time series:

To reduce the side-lobes and long-range leakage, it is better to smoothly **window or taper** the ends of our padded data sequence(https://en.wikipedia.org/wiki/Window_function) . That is, we would multiply our infinite periodic data record by some kind of a rounded window rather than the square boxcar (mask) function above. Let's imagine the results, based on the fft of different window functions:

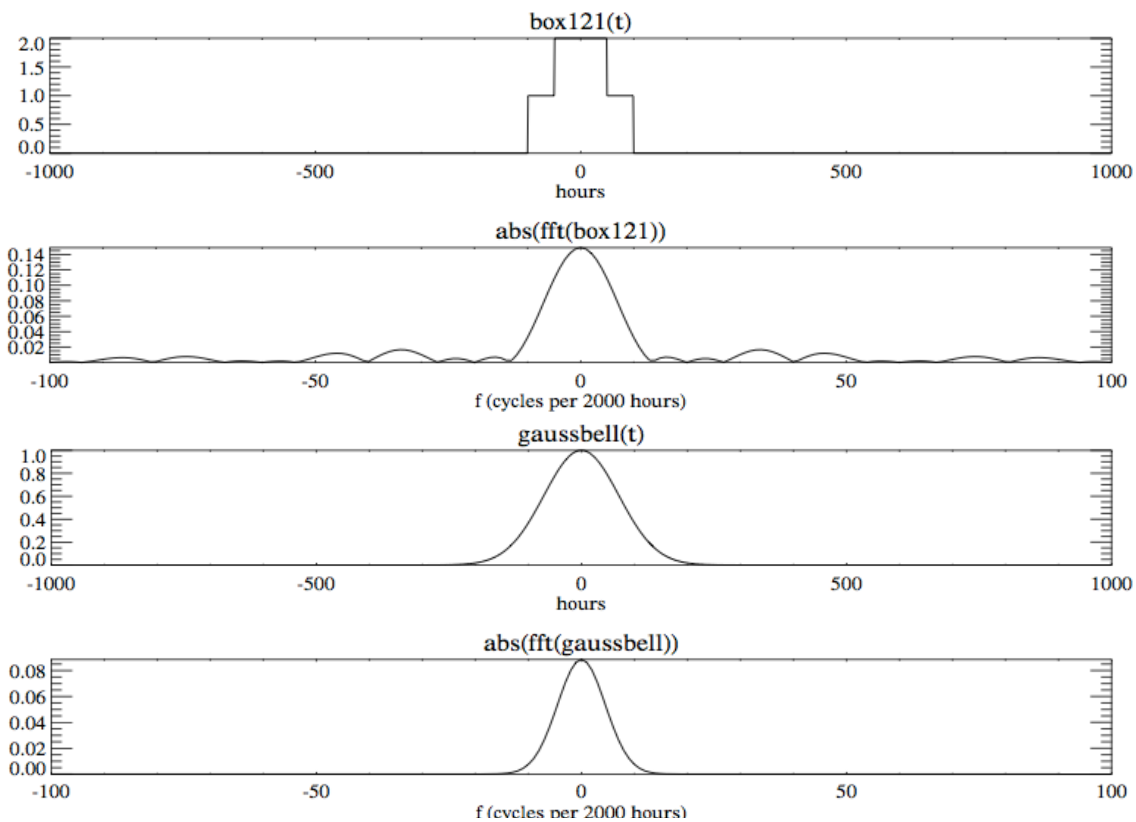


Figure 15. As in Fig. 14, but for "window" functions that taper, instead of a square boxcar.

WOW! **The Fourier transform of a Gaussian is a Gaussian** (spooky? or just beautiful?). *No side lobes.*

Now remember, this theorem works both ways:

converse is also true.

So when somebody says they used a "1-2-1 smoother" on their data, it means they *convolved* their time series with the 121 function of Fig. 15a. Often people will do this several times in succession to make things more smooth. Well, the effect on the power spectrum is equivalent to multiplying that spectrum by the fft(box121) function in Fig. 15b (perhaps many times). That is called a "low pass" filter: low frequencies are almost unaffected while frequencies comparable or higher than 1 cycle/(boxcar width) are damped. *Smoothing in time damps high frequencies in spectral space*.

If you really want a pure low-pass filter, it should have a boxcar-like step-function structure in *frequency* (low frequencies completely passed, high frequencies completely blocked). To do that with a smoothing kernel, you'd need a complicated wiggly kernel like fft(boxcar) on Fig. 14. But that kernel requires a wide support or stencil or 'footprint' in the time domain, with all those sidelobes, so it's impossible to apply near the ends of a time series. These are some of the inescapable trade-offs of having limited information content in a time series. There is an uncertainty principle to this: <u>there is an inescapable tradeoff between resolution in the time domain and resolution in the frequency domain</u>.

c. A glimpse of wavelet analysis

Wavelet analysis is an attempt to optimize that inescapable time-frequency tradeoff. Instead of decomposing a time series into sines and cosines that oscillate to infinity, a basis set of *wavelet functions* (https://en.wikipedia.org/wiki/Wavelet) is used. Wavelet functions $W(f,\tau)$ are like localized wave *packets*. They oscillate at some frequency (characterized by some pseudo-frequency f) around $\tau=0$, but with an amplitude that decays at large $|\tau|$. Wavelet basis functions probe the time series for *local wiggly features*, not just for global periodicities that exist somewhere within the whole time series like Fourier harmonics.

Recall that Fourier coefficients are the "projection" (the mean of the product) of your data with sine and cosine functions. The wavelet spectrum is the *projection* (resemblance, measured by the integral of the product) of your data with the wavelet basis function. A **wavelet power spectrum** is a function of both time and frequency, $\mathcal{L}_w$**(f,t).**

If your wavelet basis function has a lot of wiggles, its projection onto an oscillatory time series will clearly distinguish the *frequency* of oscillations. But in order to have a lot of wiggles, it must have a wide footprint or stencil in the time domain. That means the projection will light up at any central time when that long wiggly wavelet stencil overlaps with a wiggle in the time series. As a result, the wavelet projection will be similar for all the times that are within its time stencil width, leading to poor time resolution.

On the other hand, a more compact-in-time wavelet function like the DOG (derivative of Gaussian) or Mexican Hat (you can google image these) just has one sidelobe, so it can't distinguish frequencies very sharply, but its localization in time makes it good at localizing time series peaks or "events". *Frequency vs. time resolution is a trade-off.*

So you choose your wavelet function based on your purposes, and what you are trying to extract or emphasize about your time series. For a time series with "event-like" features, a compact wavelet is good; for a time series with "wave-like" features whose frequency you want to resolve exactly, it makes sense to use a more wavelike wavelet function, with Fourier analysis as the limit (the Fourier power spectrum is not a function of time at all).

d. Revisiting the problem of aliasing using the convolution theorem

Let's revisit the aliasing problem of Fig. 8, the 3-hourly subsampling of 2-minute L(t) from section 3a above. Figure 15 shows the "shah" (comb or spikes) function that, when multiplied by the full time series, gives the 3-hourly samples. Its amplitude spectrum (bottom) is also a series of spikes, now in frequency, spaced 8 cycles/day apart. Notice that this is twice the Nyquist frequency of 4 cycles per day.

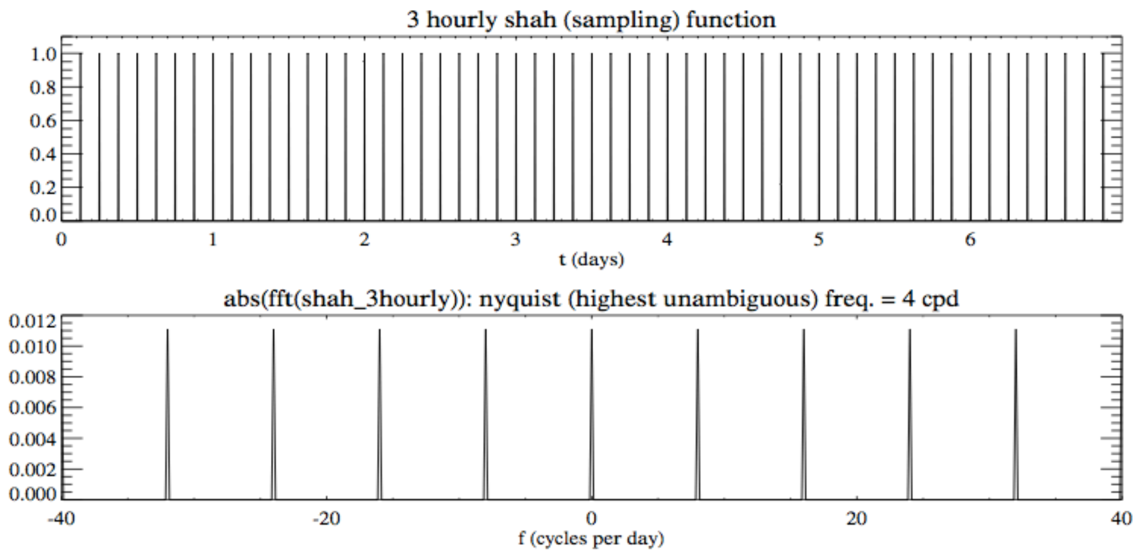


Figure 15. The comb or "shah" function: a set of equally spaced spikes. Its product with our L(t) data gives the subsampled data (red triangles in Fig. 8a). Its Fourier spectrum is shown at bottom. Notice that the nonzero mean of the function value gives a nonzero transform at frequency f=0.

According to the convolution theorem, the effect of multiplying L(t) by the 3-hourly shah(t) function in time is to *convolve* $\mathcal{L}$(f) with shah(f) in the frequency domain. In this case, our spectrum $\mathcal{L}$(f) is the function that is more compact around the origin (since geophysical spectra usually are "red"), while shah(f) is the function that extends to infinity (Fig. 15b). So it is probably clearer to think of our spectrum as the 'kernel' and shah(f) the thing being 'smoothed' by that kernel.

Figure 16 shows the spectrum of the full data, plotted as amplitude (not power), and symmetric (not just the positive frequencies). It's the same information as Fig. 2 (repeated at right here as a reminder):

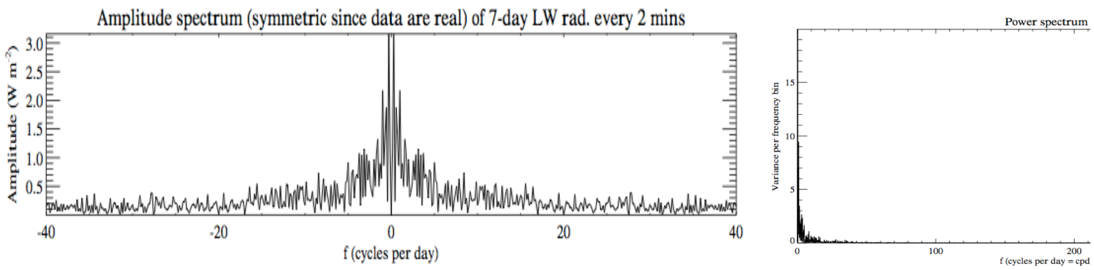


Figure 15. The amplitude spectrum of L(t), shown in symmetric form around f=0 for clarity.

Now we can build the convolution as a kernel sum, just like in Fig. 13:

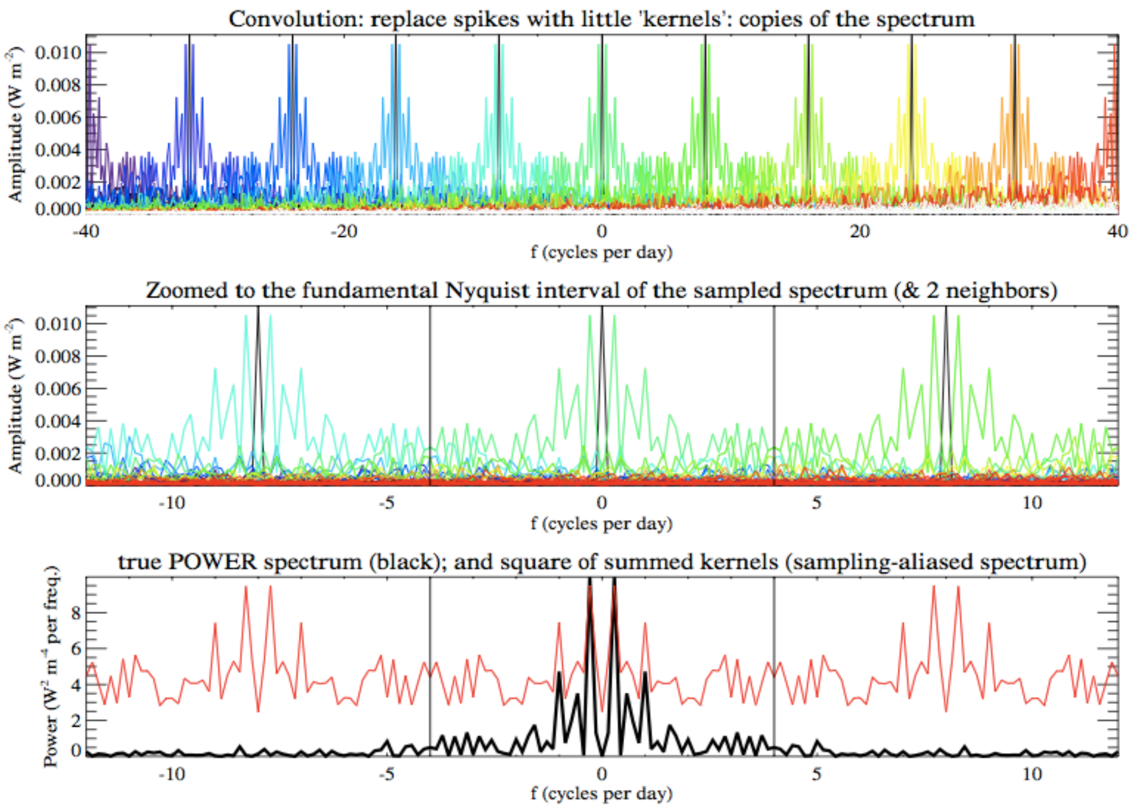


Figure 16. Summing of kernels as in Fig. 13. The true spectrum $\mathcal{L}$ (f) is treated as a "kernel" for the sampling or shah function whose values are replaced by the kernel. In the bottom panel, the true power spectrum is in black, and the aliased power spectrum due to 3-hour undersampling is in red. The red-black gap is a noise level or error bar for all possible alignments of the 3h sampling.

When we sum up all the kernels in Fig. 16, the long tails of high-frequency power from all the faraway copies of the spectrum add up to a large amount of spurious (aliased) power. The red-black gap at bottom here is a vertical "error bar" for spectra. As we saw in Fig. 8, 3h subsampling gives modest error on our lowest frequency peak (2 cycles/7 days). For the diurnal peak (1cpd), 3h sampling produces about a ~100% error bar (spurious, aliased power about equal to true power). For f > 1cpd, the spurious power is larger than the true power. That's consistent with the results of our one particular sampling *realization* from Fig. 8 (Fig. 17): the low frequency is maybe mostly right, diurnal is problematic, and the rest is mostly spurious.

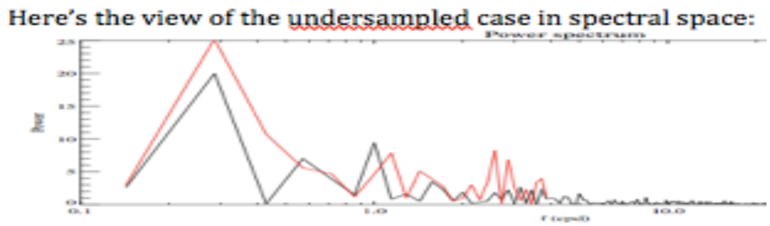


Figure 17. Repeating Fig. 8: the sampling-aliased spectrum for 3-hourly sampling aligned with t=0.

5. Summary of "ERROR BARS" in frequency space arising from data shortcomings in physical space (or time)

Taking the Fourier transform of a **finite segment** [0,T] of data L(t) to get $\mathcal{L}$(f) tacitly assumes that the patterns in L(t) repeat periodically every interval T, forever in all t. The finiteness of T gives a **lower bound** on the lowest frequency (1 cycle per record length T) in $\mathcal{L}$(f). It also **discretizes the frequencies** (1,2,3,... cycles per T). This discrete frequency spacing ("**bandwidth**") of 1 cycle/T is, in estimation terms, a form of horizontal "error bar" in frequency space.

Spectral "leakage" is the continuous version of this discretization of bandwidth. For example, say we pad the T length data sequence with 5T worth of zeros on each end, giving a total padded record length of 11T. Analyzing this longer sequence (which, again, assumes this longer 11T sequence is periodically repeated to infinity) will give 11x finer spectral resolution (1 cycle/11T instead of 1 cycle/T). But does this mean we will be able to actually discriminate frequencies that are closer together than 1 cycle/T? No, because spectral "leakage" or smearing will still act as an error bar in the frequency domain. This should not seem surprising, if you think about it: information-wise, you can't get something for nothing (where padding with zeros is literally nothing!).

When we have a **discrete** sequence of data values at times separated by dt, it gives an **upper bound** to the frequencies we can resolve: (1 cycle per 2dt), the **Nyquist frequency**. Folding (aliasing) of power from beyond the Nyquist frequency is shown in Fig. 16: the tails of power beyond the central Nyquist interval (-4 to 4 cpd in Fig. 16) add up to produce aliased power, providing a **vertical error bar on our estimates of the power spectrum**. *Much the power in an undersampled spectrum is spurious (aliased)*. Revisit Fig. 9 for a sobering reminder of how much.

"Statistical significance" of spectral peaks is assessed using the F test for the ratio of its variance to the variance at that frequency in AR1 noise with the same total variance and the same lag-1 autocorrelation ϕ as measured from your time series. Since the AR1 noise is known perfectly, it has infinite degrees of freedom, while your spectrum has 2 DOFs per frequency (from the sine component and the cosine component). The F test statistic for a p-value of 0.1 is F(2,infinity) = 9.49. In other words, *an expected peak must be 10x above the background to be 90% significant as a finding*! This is daunting, and many early periodogram studies got fooled! If a peak is 5 bandwidths wide, then it has to exceed F(10,infinity) = 2.1 to be *a priori* significant at p=0.1. Be careful! https://www.statology.org/wp-content/uploads/2018/09/f1.png