13. Colored Processes

A complete statistical theory is readily available for white noise processes. But construction of the periodogram or spectral density estimate of real processes (e.g., Fig. 13) quickly shows that a white noise process is both uninteresting, and not generally applicable in nature.

For colored noise processes, many of the useful simplifications for white noise break down (e.g., the zero covariance/correlation between Fourier coefficients of different frequencies). Fortunately, many of these relations are still asymptotically (as $N \to \infty$) valid, and for finite N, usually remain a good zero-order approximation. Thus for example, it is commonplace to continue to use the confidence limits derived for the white noise spectrum even when the actual spectrum is a colored one. This degree of approximation is generally acceptable, as long as the statistical inference is buttressed with some physical reasoning, and/or the need for statistical rigor is not too great. Fortunately in most of the geophysical sciences, one needs rough rules of thumb, and not rigid demands for certainty that something is significant at 95% but not 99% confidence.

As an example, we display in Fig. 14 the power density spectral estimate for the current meter component whose periodogram is shown in Fig. 13. The 95% confidence interval shown is only approximate, as it is derived rigorously for the white noise case. Notice however, that the very high-frequency jitter is encompassed by the interval shown, suggesting that the calculated confidence interval is close to the true value. The inertial peak is evidently significant, relative to the background continuum spectrum at 95% confidence. The tidal peak is only marginally significant (and there are formal tests for peaks in white noise; see e.g., Chapter 8 of Priestley). Nonetheless, the existence of a theory which suggests that there should be a peak at exactly the observed period of 12.42 hours would convince most observers of the reality of the peak irrespective of the formal confidence limit. Note however, that should the peak appear at some unexpected place (e.g., 10 hours) where there is no reason to anticipate anything special, one would seek a rigorous demonstration of its reality before becoming too excited about it.

Another application of this type of analysis can be seen in Fig. 15, where the power density spectra of the tide gauge records at two Pacific islands are shown. For many years it had been known that there was a conspicuous, statistically significant, but unexplained, spectral peak near 4 days period at Canton Island, which is very near the equator.

Plotting of Power Density Spectra

Some words about the plotting of spectra (and periodograms) is helpful. For most geophysical purposes, the default form of plot is a log-log one where both estimated power density and the frequency scale are logarithmic. There are several reasons for this (1) many geophysical processes produce spectral densities which are power laws, s^{-q} over one or more ranges of frequencies. These appear as straight
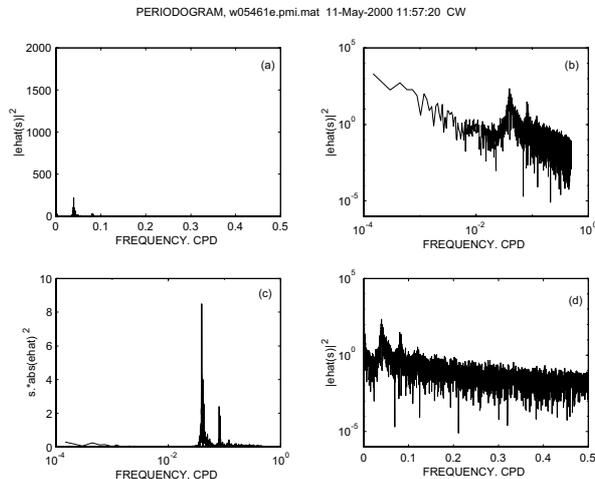
PERIODOGRAM, w05461e.pmi.mat 11-May-2000 11:57:20 CW

FIGURE 13. Periodogram of the zonal component of velocity at a depth of 498m, 27.9°N,54.9°W in the North Atlantic, plotted in four different ways. (a) Both scales are linear. The highest value is invisible, being lost against the $y-$axis. The rest of the values are dominated by a peak at the local inertial frequency $(f/2\pi)$. In (b) a log-log scale is used. Now one sees the entirety of the values, and in particular the near power-law like behavior away from the inertial peak. But there are many more frequency points at the high frequency end than there are at the low frequency end, and in terms of relative power, the result is misleading.

(c) A so-called area-preserving form in which a linear-log plot is made of $s\,|\alpha_n|^2$ which compensates the abcissa for the crushing of the estimates at high frequencies. The relative power is proportional to the area under the curve (not so easy to judge here by eye), but suggesting that most of the power is in the inertial peak and internal waves $(s > f/2\pi)$. (d) shows a logarithmic ordinate against linear frequency, which emphasizes the very large number of high frequency components relative to the small number at frequencies below $f/2\pi$.

lines on a log-log plot and so are easily recognized. A significant amount of theory (the most famous is Kolmogorov's wavenumber $k^{-5/3}$ rule, but many other theories for other power laws exist as well) leading to these laws has been constructed. (2) Long records are often the most precious and one is most interested in the behavior of the spectrum at the lowest frequencies. The logarithmic frequency scale compresses the high frequencies, rendering the low frequency portion of the spectrum more conspicuously. (3) The "red" character of many natural spectra produces such a large dynamic range in the spectrum that much of the result is invisible without a logarithmic scale. (4) The confidence interval is nearly constant on a log-power density scale; note that the confidence interval (and sample variance) are dependent upon $\tilde{\Phi}^\nu(s)$,
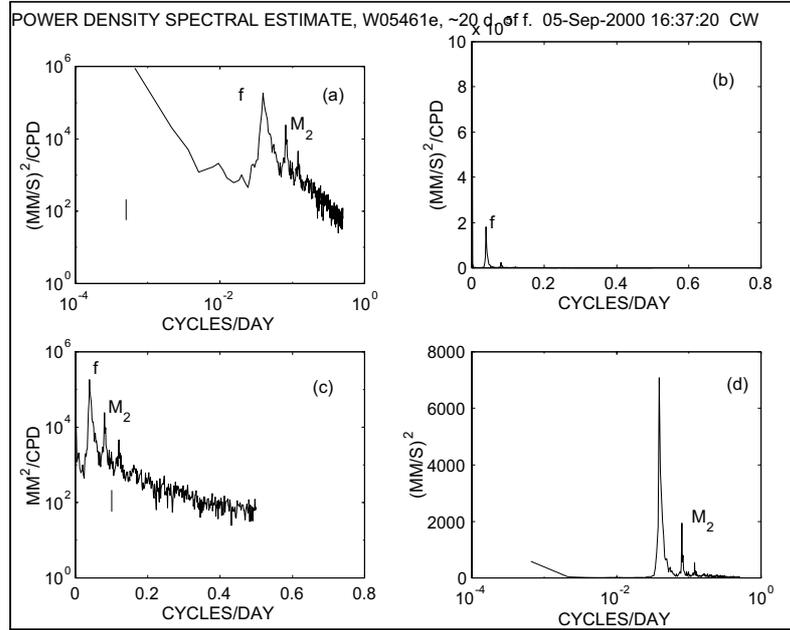
FIGURE 14. Four different plots of a power spectral density estimate for the east component of the current meter record whose periodogram is displayed in Fig. 13. There are approximately 20 degrees of freedom in the estimates. An approximate 95% confidence interval is shown for the two plots with a logarithmic power density scale; note that the high frequency estimates tend to oscillate within about this interval. (b) is a linear-linear plot, and (d) is a so-called area preserving plot, which is linear-log. The Coriolis frequency, denoted f, and the principal lunar tidal peaks (M_2) are marked. Other tidal overtones are apparently present.

with larger estimates having larger variance and confidence intervals. The fixed confidence interval on the logarithmic scale is a great convenience.

Several other plotting schemes are used. The logarithmic frequency scale emphasizes the low frequencies at the expense of the high frequencies. But the Parseval relationship says that all frequencies are on an equal footing, and one's eye is not good at compensating for the crowding of the high frequencies. To give a better pictorial representation of the energy distribution, many investigator's prefer to plot $s\tilde{\Phi}^\nu(s)$ on a linear scale, against the logarithm of s. This is sometimes known as an "area-preserving plot" because it compensates the squeezing of the frequency scale by the multiplication by s (simultaneously reducing the dynamic range by suppression of the low frequencies in red spectra). Consider how this behaves (Fig. 16) for white noise. The log-log plot is flat within the confidence limit, but it may not be immediately obvious that most of the energy is at the highest frequencies. The area-preserving plot looks "blue" (although we know this is a white noise spectrum). The area under the curve is proportional
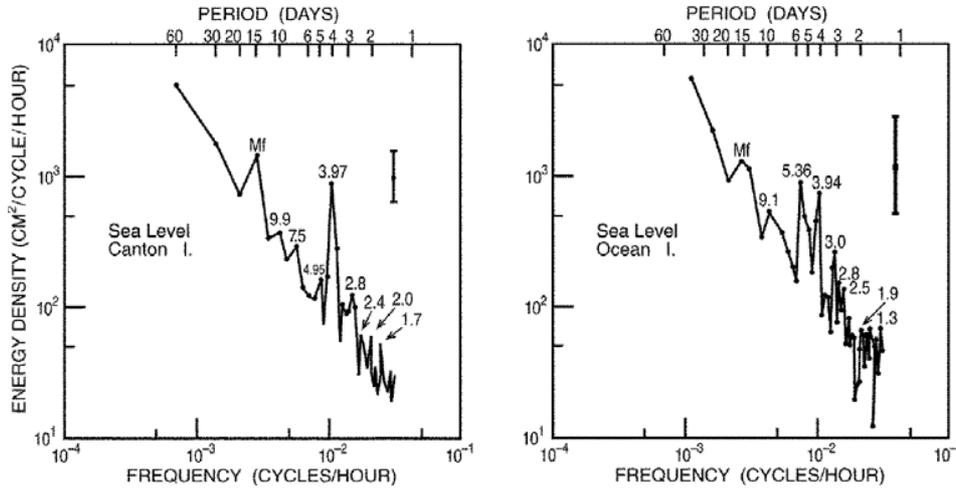
FIGURE 15. Power spectral density of sealevel variations at two near-equatorial islands in the tropical Pacific Ocean (Wunsch and Gill, 1976). An approximate 95% confidence limit is shown. Some of the peaks, not all of which are significant at 95% confidence, are labelled with periods in days (Mf is however, the fortnightly tide). Notice that in the vicinity of 4 days the Canton Island record shows a strong peak (although the actual sealevel variability is only about 1 cm rms), while that at Ocean I. shows not only the 4 day peak, but also one near 5 days. These results were explained by Wunsch and Gill (1976) as the surface manifestation of equatorially trapped baroclinic waves. To the extent that a theory predicts 4 and 5 day periods at these locations, the lack of rigor in the statistical arguments is less of a concern. (The confidence limit is rigorous only for a white noise spectrum, and these are quite "red". Note too that the high frequency part of the spectrum extending out to 0.5 cycles/hour has been omitted from the plots.)

to the fraction of the variance in any fixed logarithmic frequency interval, demonstrating that most of the record energy is in the high frequencies. One needs to be careful to recall that the underlying spectrum is constant. The confidence interval would differ in an obvious way at each frequency. A similar set of plots is shown in Fig. (14) for a real record.

Beginners often use linear-linear plots, as it seems more natural. This form of plot is perfectly acceptable over limited frequency ranges; it becomes extremely problematic when used over the complete frequency range of a real record, and its use has clearly distorted much of climate science. Consider Fig. 17 taken from a tuned ocean core record (Tiedemann, et al., 1994) and whose spectral density estimate (Fig. 19) is plotted on linear-linear scale.

One has the impression that the record is dominated by the energy in the Milankovitch peaks (as marked; the reader is cautioned that this core was "tuned" to produce these peaks and a discussion of their reality or otherwise is beyond our present scope). But in fact they contain only a small fraction
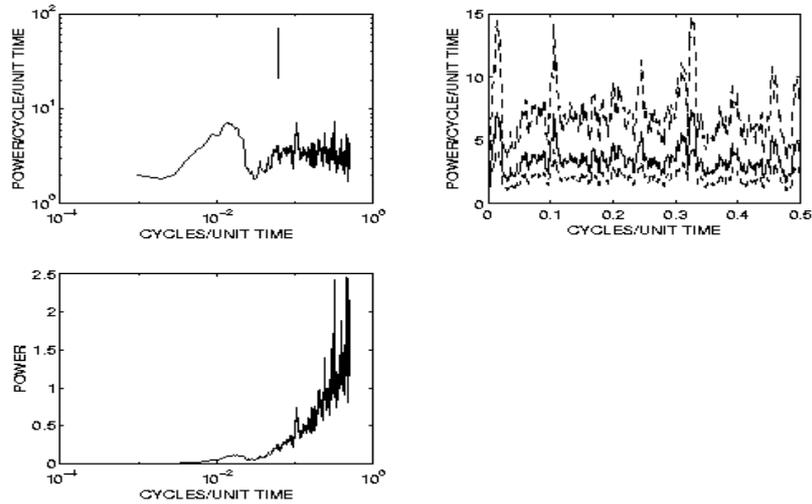
FIGURE 16. Power density spectral estimate for pseudorandom numbers (white noise) with about 12 degrees-of-freedom plotted in three different ways. Upper left figure is a log-log plot with a single confidence limit shown, equal to the average for all frequencies. Upper right is a linear plot showing the upper and lower confidence limits as a function of frequency, which is necessary on a linear scale. Lower left is the area-preserving form, showing that most of the energy on the logarithmic frequency scale lies at the high end, but giving the illusion of a peak at the very highest frequency.
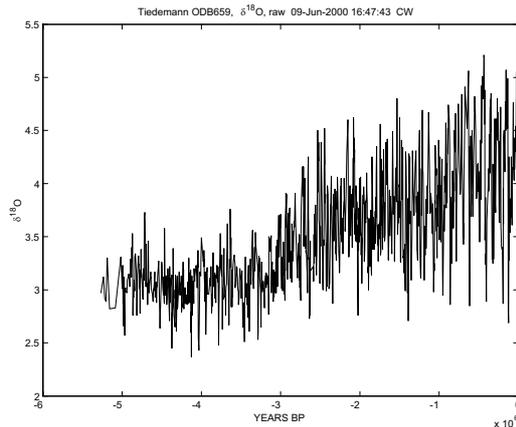

FIGURE 17. Time series from a deep-sea core (Tiedemann, et al., 1994). The time scale was tuned under the assumption that the Milankovitch frequencies should be prominent.
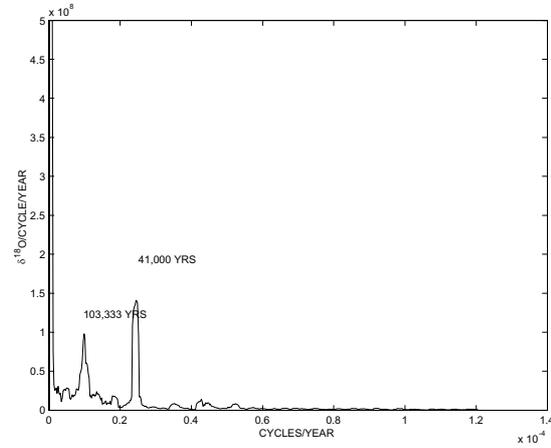
FIGURE 18. Linear-linear plot of the power density spectral estimate of the Tiedemann et al. record. This form may suggest that the Milankovitch periodicities dominate the record (although the tuning has assured they exist).
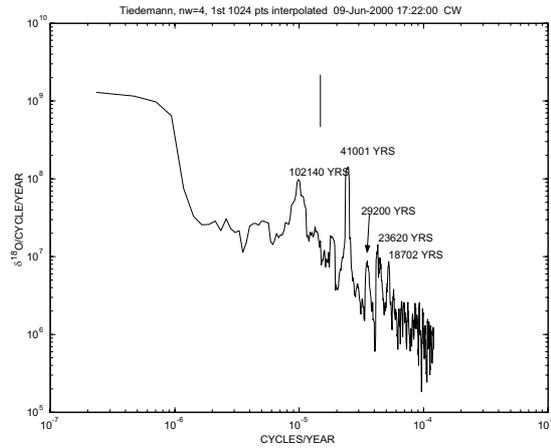

FIGURE 19. The estimated power density spectrum of the record from the core re-plotted on a log-log scale. Some of the significant peaks are indicated, but this form suggests that the peaks contain only a small fraction of the energy required to describe this record. An approximate 95% confidence limit is shown and which is a nearly uniform interval with frequency.

of the record variance, which is largely invisible solely because of the way the plot suppresses the lower values of the much larger number of continuum estimates. Again too, the confidence interval is different for every point on the plot.

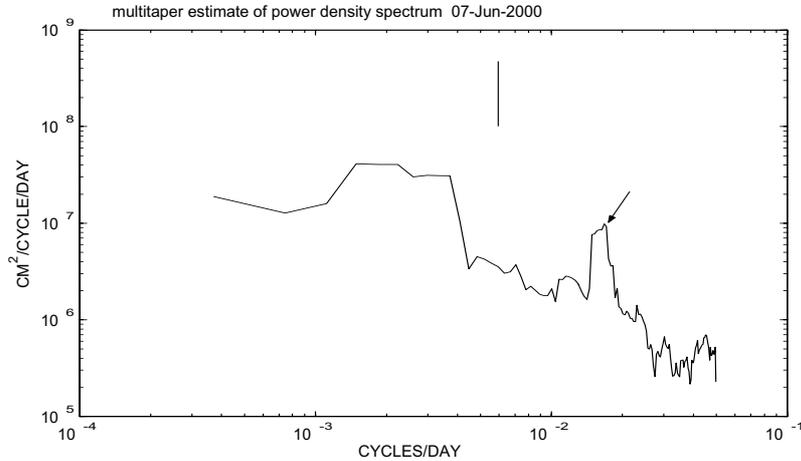multitaper estimate of power density spectrum 07-Jun-2000

FIGURE 20. Power density spectra from altimetric data in the western Mascarene Basin of the Indian Ocean. The result shows (B. Warren, personal communication, 2000) that an observed 60 day peak (arrow) in current meter records also appears conspicuously in the altimetric data from the same region. These same data show no such peak further east. As described earlier, there is a residual alias from the M_2 tidal component at almost exactly the observed period. Identifying the peak as beng a resonant mode, rather than a tidal line, relies on the narrow band (pure sinusoidal) nature of a tide, in contrast to the broadband character of what is actually observed. (Some part of the energy seen in the peak, may however be a tidal residual.)

As one more example, Fig. 20 shows a power density spectrum from altimetric data in the Mascarene Basin of the western Indian Ocean. This and equivalent spectra were used to confirm the existence of a spectral peak near 60 days in the Mascarene Basin, and which is absent elsewhere in the Indian Ocean.