How to Find the Power Spectrum

Bob Parker, IGPP

SEMC March 2012

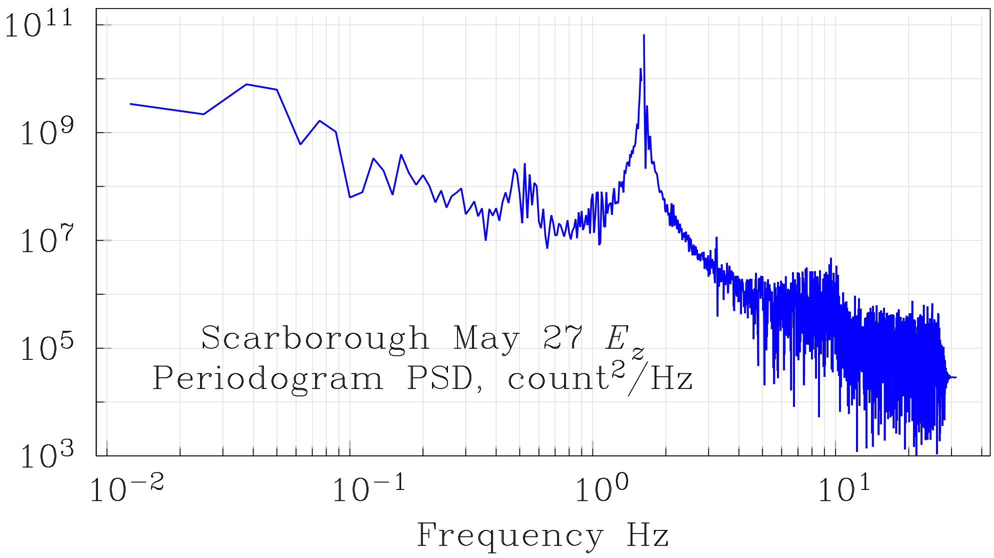
The power spectrum, or power spectral density (PSD), is the variance per unit bandwidth of a stationary stochastic process.

It is the analog of the Fourier transform for random functions.

The PSD often gives remarkable insights into a data record.

The naive estimator, the **periodogram**, takes the squared magnitude of the FFT of the record.

The result is ghastly: this estimator is **biased** and has **large variance**, which does not decrease with increasing record length (an inconsistent estimate).

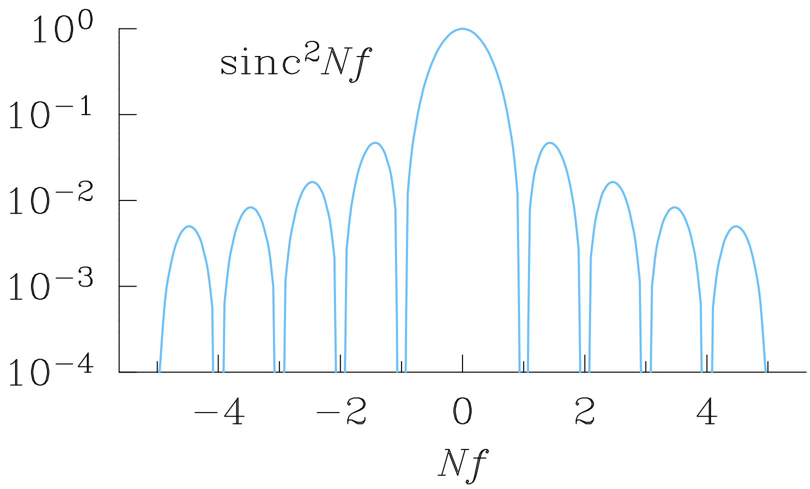


Bias in the periodogram takes the form of **spectral leakage**.

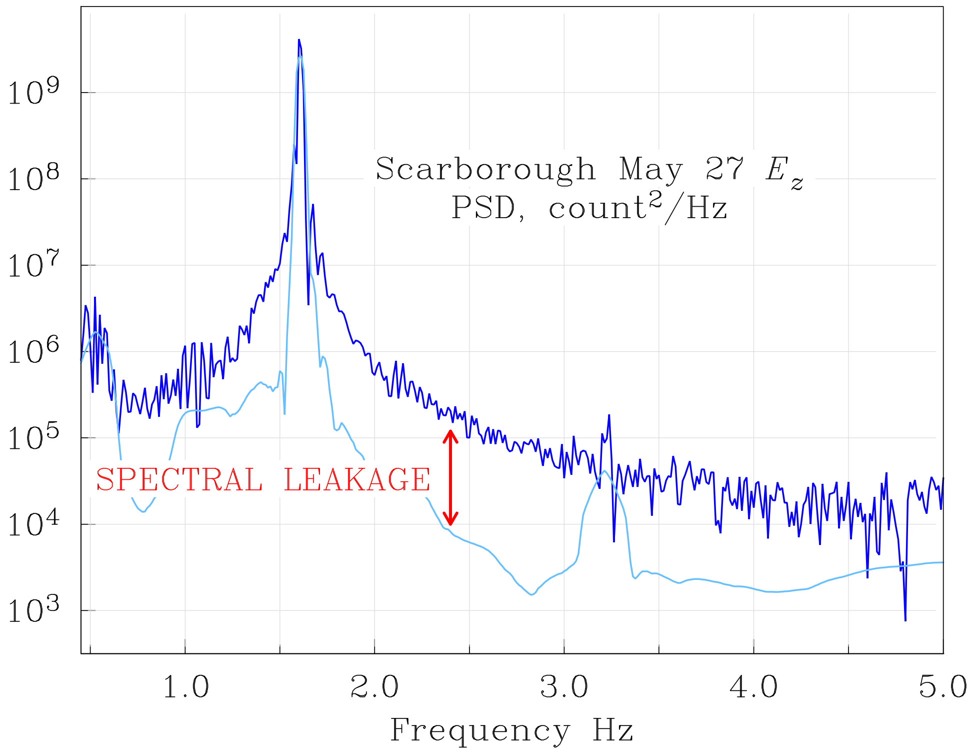
The true spectrum is smeared by convolution with a sinc-squared kernel:

$$\mathcal{E}\,[\tilde{S}] = S * F$$

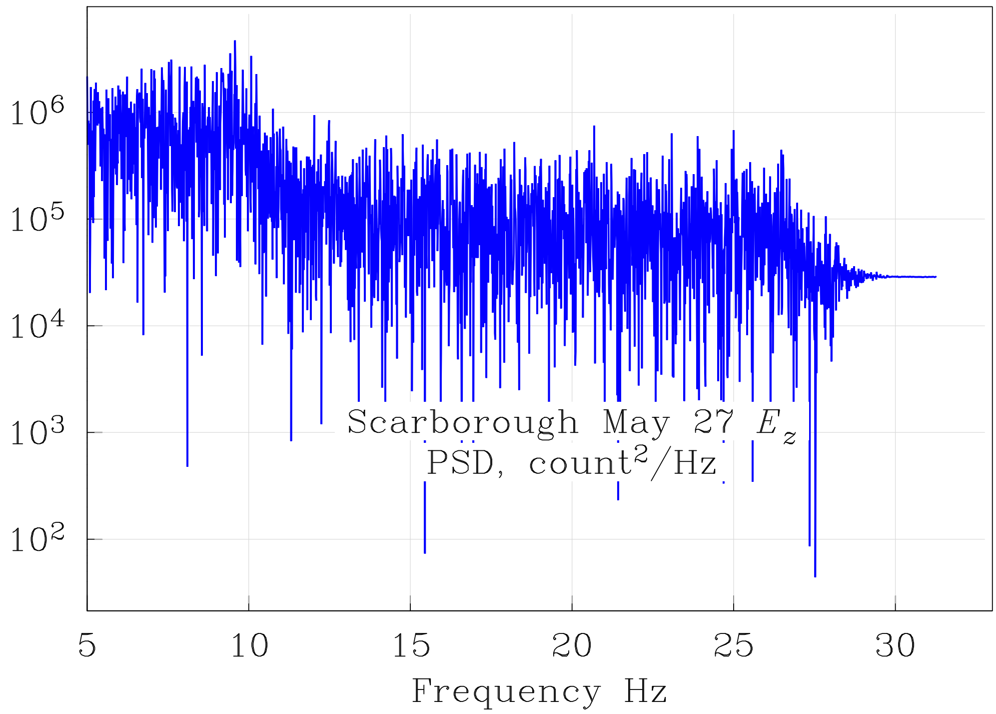
$$F(f) = N\,\mathrm{sinc}^2(Nf)$$



Energy from peaks is spread out into valleys.



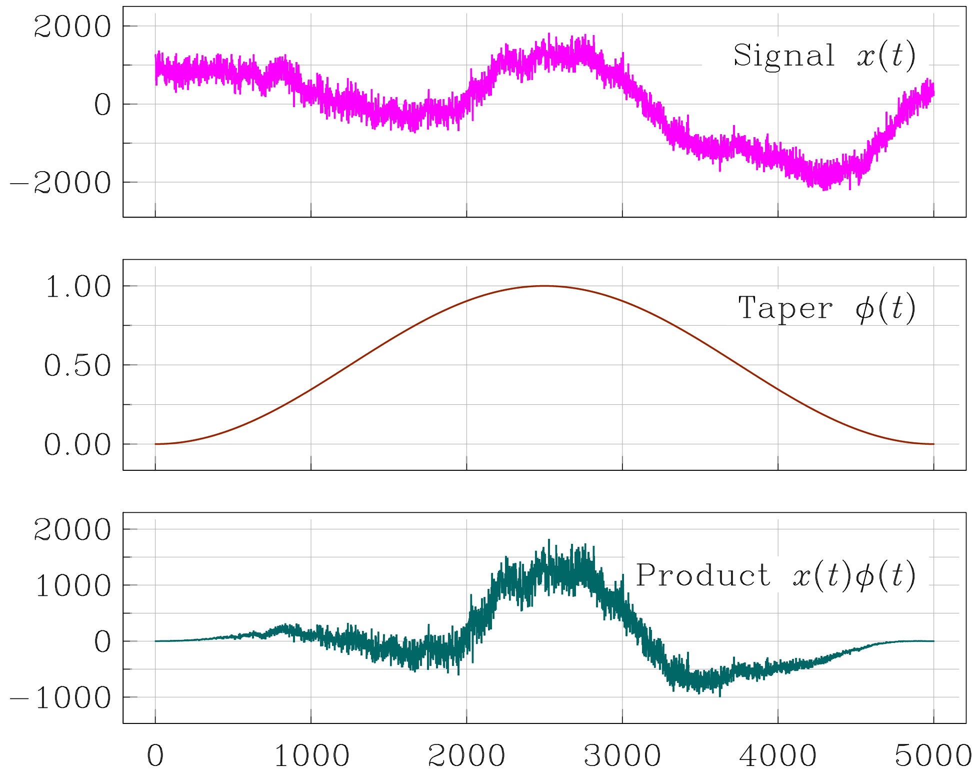
The high variance of the periodogram results in estimates with a 100% standard error: $\hat{S}(f) \pm \hat{S}(f)$.



Bias can be enormously **reduced by tapering**: multiplying the time series by a function $\phi(t)$ with a concentrated Fourier transform $\hat{\phi}(f)$.

Then one finds the periodogram of the tapered record.

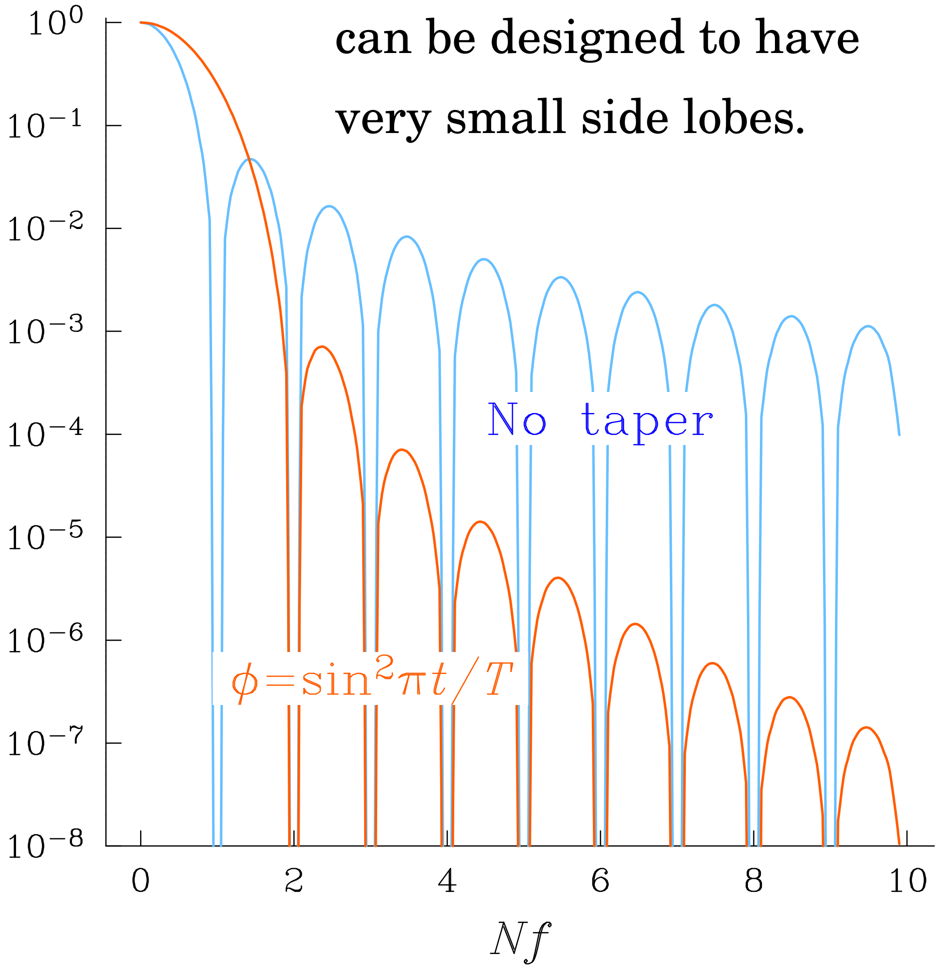
For example, like this:



There is still bias, but now

$\mathcal{E}\,[\tilde{S}] = S * |\hat{\phi}(f)|^2$

And the function $|\hat{\phi}(f)|^2$ can be designed to have very small side lobes.



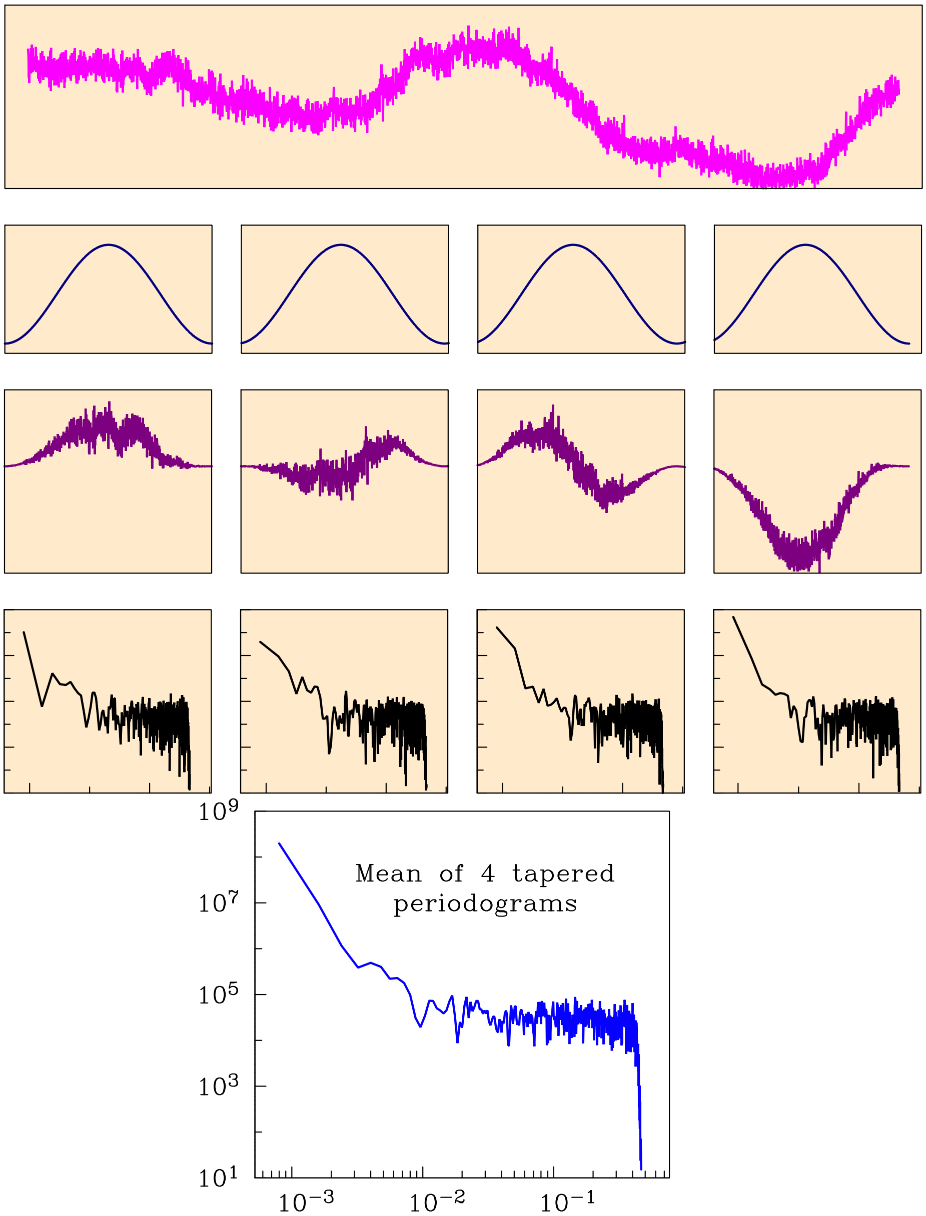
Tapering does nothing
to improve variance.

To reduce variance one must
average in some way.

For example:

Welch's method breaks the record into M equal-length segments, **tapers** each one, then **averages** the M periodograms together.

The method addresses both defects of the periodogram: bias and variance.



Mean of 4 tapered periodograms
10^9
10^7
10^5
10^3
10^1
10^{-3}
10^{-2}
10^{-1}

Welch is still in wide use despite **deficiencies:**

- Severe loss of low frequency information.
- No theory for choosing the number of segments.
- Same resolution and variance reduction at all frequencies.

In 1982 David Thomson described **multitapers** and **Slepian** (prolate spheroidal) tapers.

In 1995 Riedel and Sidorenko proposed **sine multitapers**.

Thomson showed that if two tapers, ϕ_1 and ϕ_2, are **orthogonal,** that is,

$$\int_0^T \phi_1(t)\, \phi_2(t)\, dt = 0$$

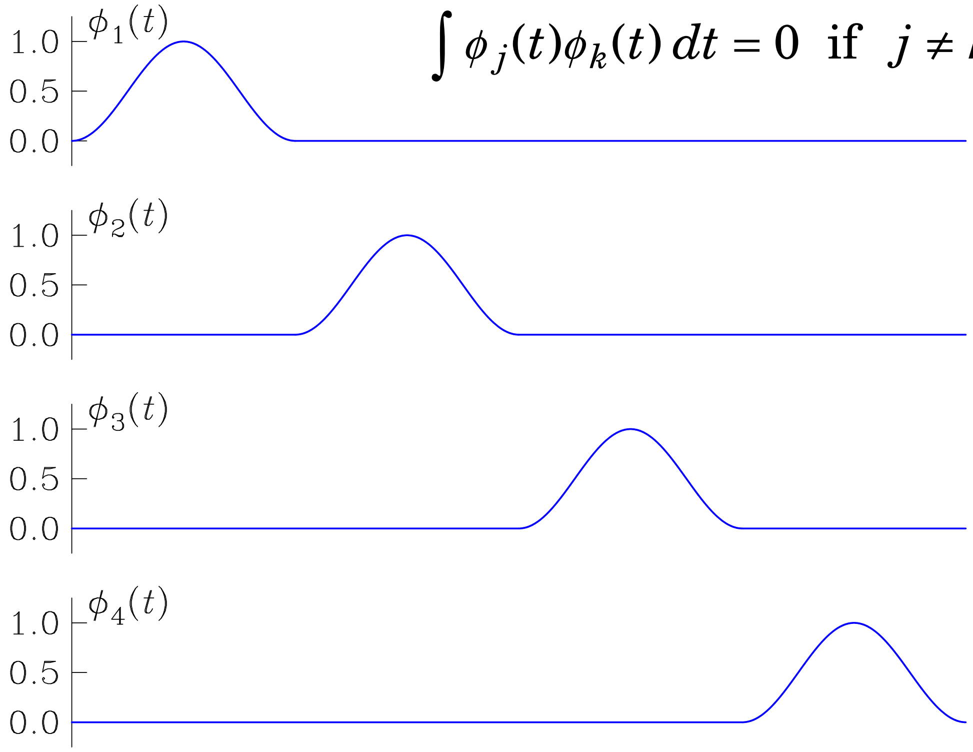
then the periodogram **estimates** from the two tapered records are **statistically independent**.

Averaging together estimates based on orthogonal tapers will yield **optimal variance reduction.**

It opens up the possibility of using **any orthogonal family** of functions as tapers.

In fact the Welch method (without overlap, as I described it here) employs an orthogonal family:

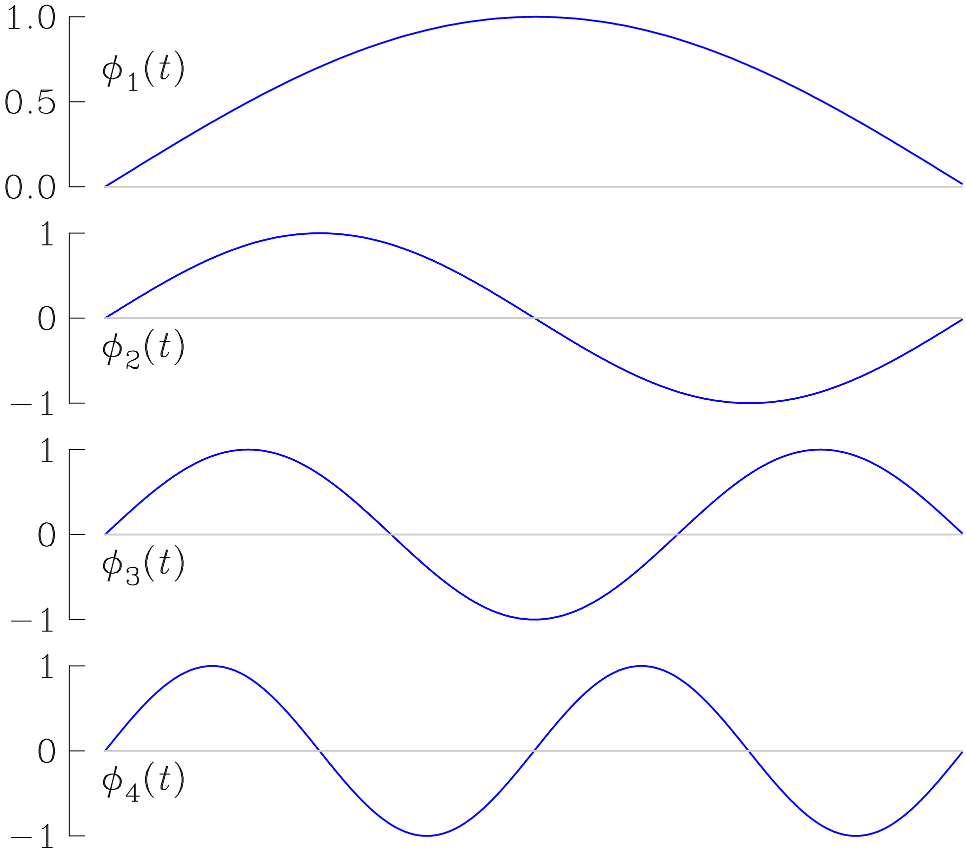
$$\int \phi_j(t)\phi_k(t)\,dt = 0 \quad \text{if} \quad j \neq k$$



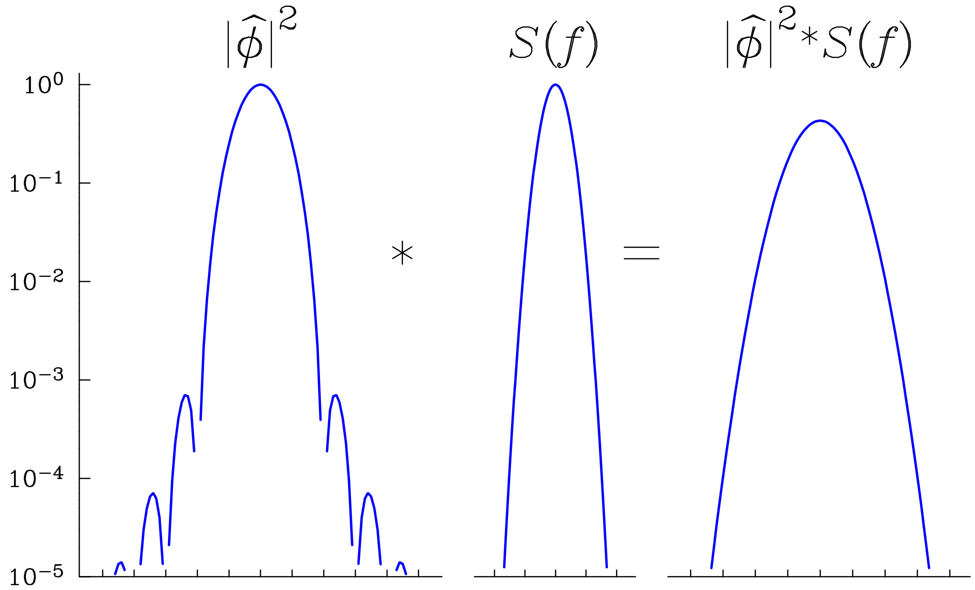
The **Welch** family of tapers makes very **poor use of the data;** overlapping segments yield a trivial improvement.

Thomson proposed the **prolate spheroidal** family, which has some remarkable properties, particularly effective for short series. I won't discuss that family here.

I want to discuss the **sine tapers.**



Tapering also introduces a **local bias** in addition to spectral leakage owing to convolution with $|\hat{\phi}(f)|^2$.

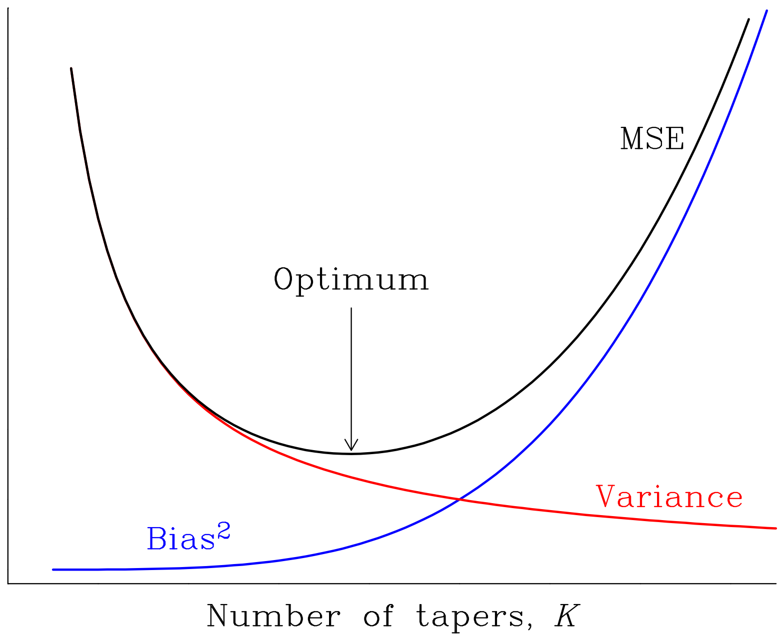


Riedel and Sidorenko showed that local bias is **minimized by the sine tapers.**

If the time series is long enough, local bias becomes more serious than spectral leakage.

They proposed the minimization of the **mean square error**:

$$\mathrm{MSE}[S] = \beta[S]^2 + \mathrm{var}\,[S]$$



The smallest MSE is obtained with

$$K_{\mathrm{opt}}(f) = \left(\frac{12T^2 S(f)}{|S''(f)|} \right)^{2/5}$$

K_{opt} is the number of tapered periodograms to be averaged at frequency f.

$$K_{\text{opt}}(f) = \left(\frac{12T^2 S(f)}{|S''(f)|} \right)^{2/5}$$

However, there are several problems:

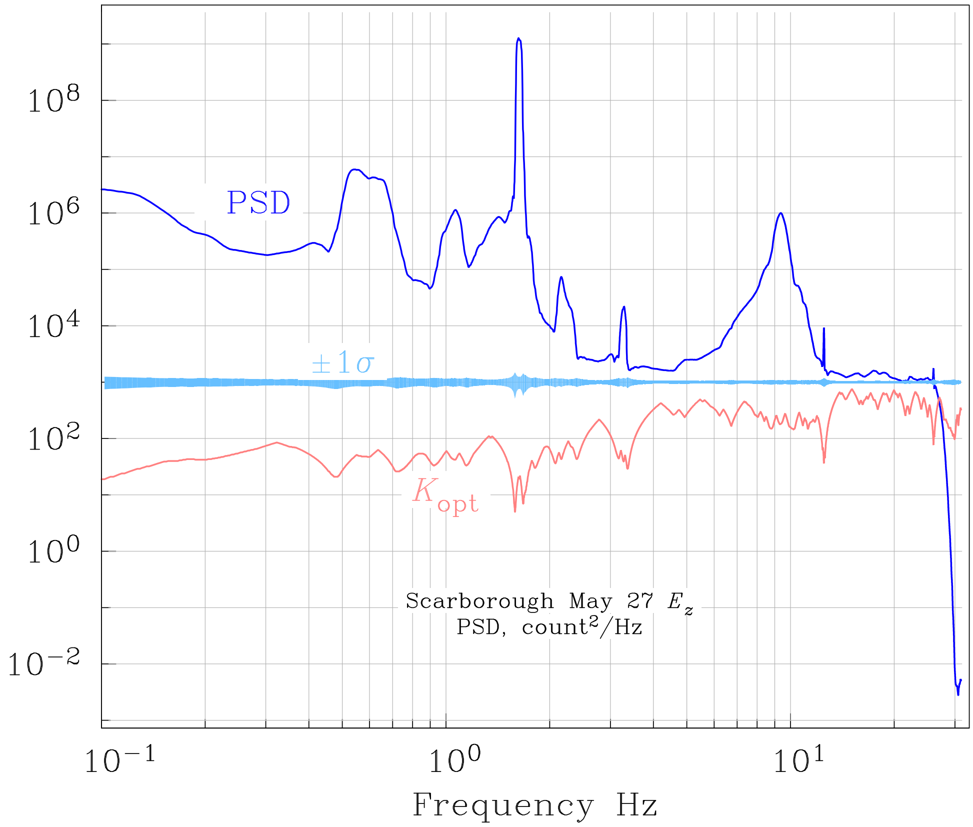
- One apparently needs to know S in order to estimate S.
- One needs to know the 2nd derivative, $S''(f)$.
- What happens when $S''(f) = 0$?
- One must apparently compute K_{opt} periodograms, and when K_{opt} is large (say > 100), that is a lot of work.

- We solve the first by a **bootstrap iteration**.
- We make a curvature-adaptive, **least-squares estimate** for S''.
- Small S'' is fixed by a heuristic kludge.
- **All** the tapered periodograms can be computed from **a single FFT** of the series by juggling indices.

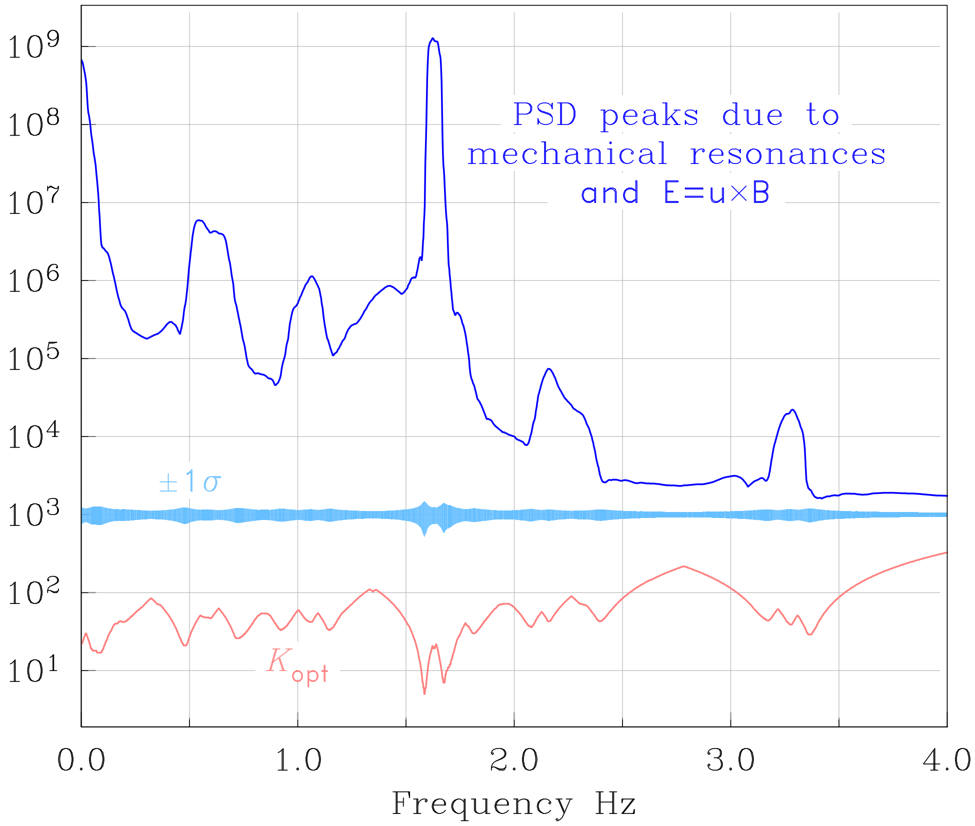
The result is a spectral estimate whose **resolution and accuracy vary with frequency:**

- Where the spectrum is smooth we get smaller variance from a lot of averaging.
- Where the spectrum varies rapidly, higher resolution, and larger variance.

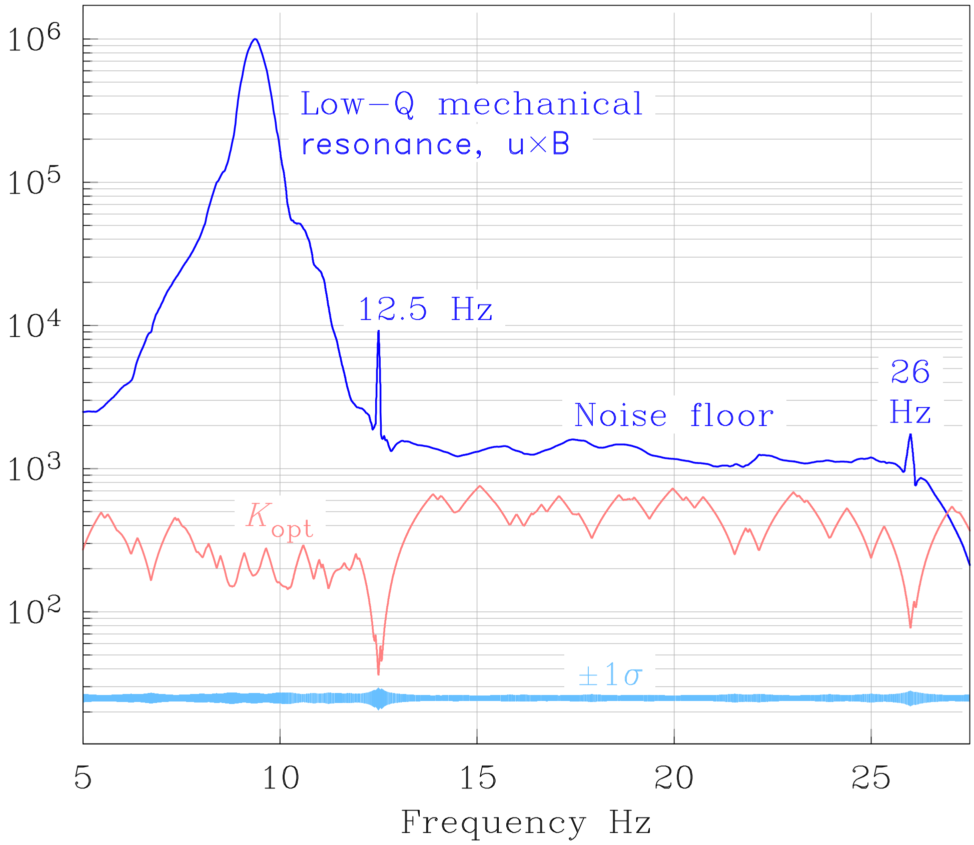
Here is the sine multitaper spectrum of the Scarborough E_z field.



Closer look, 0-4 Hz.



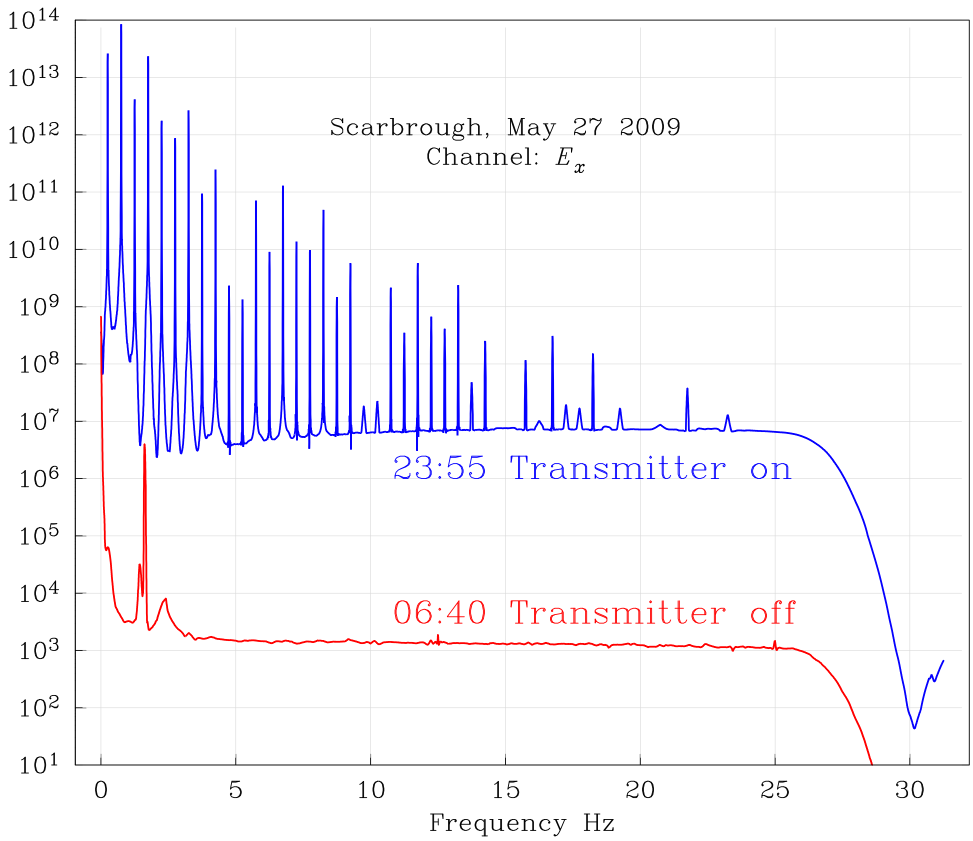
Closer look, 5-28 Hz.



Another illustration, this time with the Controlled Source transmitter turned on.

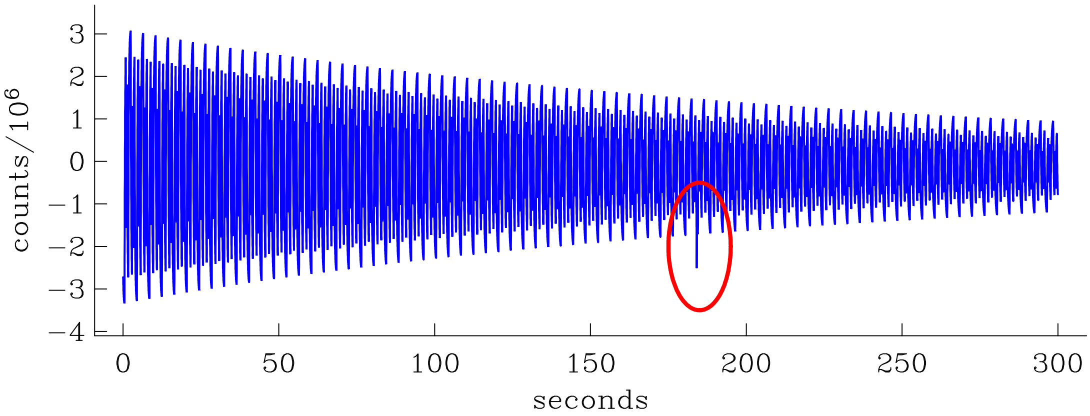
It demonstrates the incredible sensitivity of the PSD.

The apparent 'noise floor' is 4 orders of magnitude too high!



Not spectral leakage.

But there is a **spike** in the record $\cdots$



which spreads energy over the entire band.

I edited out the spike, but the problem was not fixed.

Perhaps there are other, less obvious glitches in the record.

Look at the **innovation** to detect departures from stationarity.

Model the data record as an **autoregressive process:**

$$X_n = \sum_{j=1}^{J} a_j X_{n-j} + \eta_n$$

where η_n is a white noise.

Estimate the coefficients a_j from the power spectrum via the **Yule-Walker equations.**

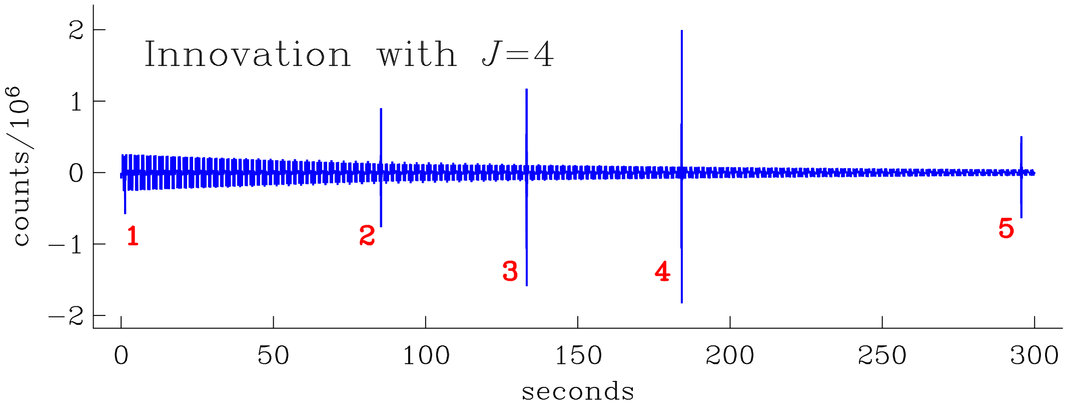
Then find the **innovation** by using the coefficients to filter the record:

$$\eta_n = X_n - \sum_{j=1}^{J} a_j X_{n-j}$$

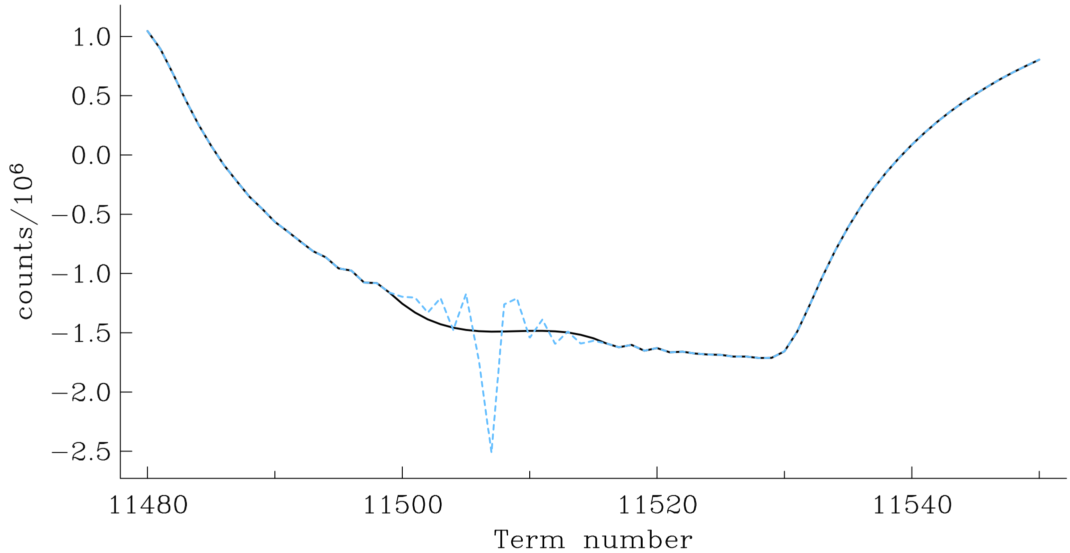
The new series will approximate white noise when X_n is a stationary process.

The procedure is often used to **pre-whiten** data in spectral estimation.

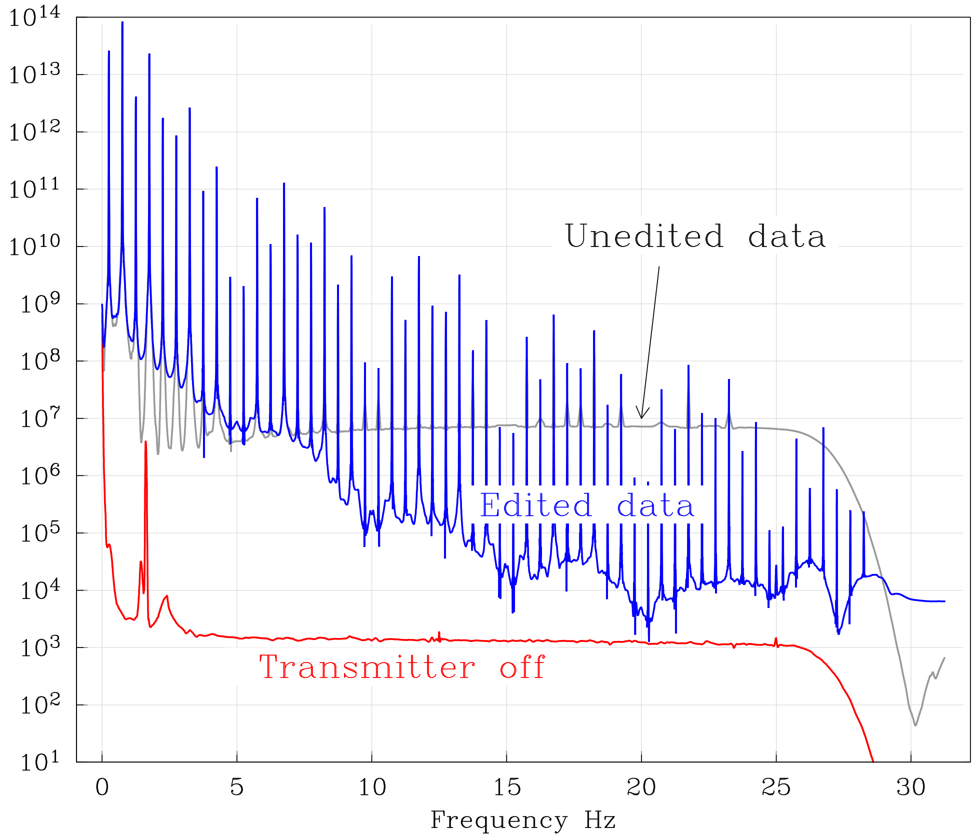
The innovation reveals 5 data spikes in the record.



I edited out the spikes with a spline; this is number 4.



The new spectrum is much better, but imperfect edits still cause distortion.



There are no spikes between terms 11,200 and 18,000. Here is the spectrum over that interval: