12.714 Computational Data Analysis

Alan Chave (alan@whoi.edu)
Thomas Herring (tah@mit.edu),
http://geoweb.mit.edu/~tah/12.714

Today's class

- Asymptotic distribution of lag window estimators
- Examples of lag window estimators
  - Bartlett
  - Daniell
  - Parzen
  - Papoulis
- Welch's overlapping segment averaging (woca)
- Example using Matlab pwelch routine
- Multi-taper methods

Asymptotic distribution of lag window estimators

- The lag window estimator

$$\boxed{\hat{S}^{(lw)}(f) = \int_{-f_{(N)}}^{f_{(N)}} W_m(f-\phi)\hat{S}^{(d)}(\phi)d\phi \approx \frac{1}{2N\Delta t}\sum_{j=-(N-1)}^{N} W_m(\tilde{f}_j)\hat{S}^{(d)}(f-\tilde{f}_{-j})}$$

- (summation is exact when $f=f_k=k/(2N\Delta t)$
- The direct estimates are chi-squared distributed with 2 degrees of freedom, and so S(lw) is the sum of the these χ^2 distributed random variables. What is its distribution?
- Approximation: $\boxed{\hat{S}^{(lw)}(f) \overset{d}{=} \alpha\chi_\nu^2}$
- Distributed with α times a ν degrees of freedom χ^2

S$^{(lw)}$ distribution

- To determine the values of α and ν, we have

$$\boxed{\begin{aligned} &E\{\hat{S}^{(lw)}(f)\} = E\{\alpha\chi_\nu^2\} = \alpha\nu \\ &\mathrm{var}\{\hat{S}^{(lw)}(f)\} = \mathrm{var}\{\alpha\chi_\nu^2\} = 2\alpha^2\nu. \\ &\therefore \nu = \frac{2\left(E\{\hat{S}^{(lw)}(f)\}\right)^2}{\mathrm{var}\{\hat{S}^{(lw)}(f)\}} \text{ and } \alpha = \frac{E\{\hat{S}^{(lw)}(f)\}}{\nu} \end{aligned}}$$

- For large numbers of samples, we can relate these expectations to known quantities and we have

$$\boxed{\nu = \frac{2N\Delta t}{C_h \int_{-f_{(N)}}^{f_{(N)}} W_m^2(\phi)d\phi} = \frac{2N}{C_h \sum_{\tau=-(N-1)}^{(N-1)} w_{\tau,m}^2} \quad \alpha = \frac{S(f)}{\nu}}$$

$S^{(lw)}$ distribution

- The quantity ν is called the *equivalent degrees of freedom*. It can also be related to the bandwidth of the smoothing window and allows us to calculate the variance of the spectral estimates.

$$\nu = \frac{2NB_W\Delta t}{C_h} \qquad \mathrm{var}\{\hat{S}^{(lw)}(f)\} = \frac{2S^2(f)}{\nu}$$

- As ν increases the variance decreases (but possibly at the expense of bias).
- If the bandwidth is increased too much, the mse can increase because of increased bias.

Confidence intervals

- If $Q_\nu(p)$ is the p*100% percentage point for χ^2_ν then we can compute confidence intervals. We have

$$P[Q_\nu(p) \le \chi^2_\nu \le Q_\nu(1-p)] = 1-2p$$

$$P[Q_\nu(p) \le \frac{\nu\hat{S}^{(lw)}(f)}{S(f)} \le Q_\nu(1-p)] =$$

$$P[\frac{\nu\hat{S}^{(lw)}(f)}{Q_\nu(1-p)} \le S(f) \le \frac{\nu\hat{S}^{(lw)}(f)}{Q_\nu(p)}] = 1-2p$$

The 100(1 - 2p)% confidence interval for S(f) is

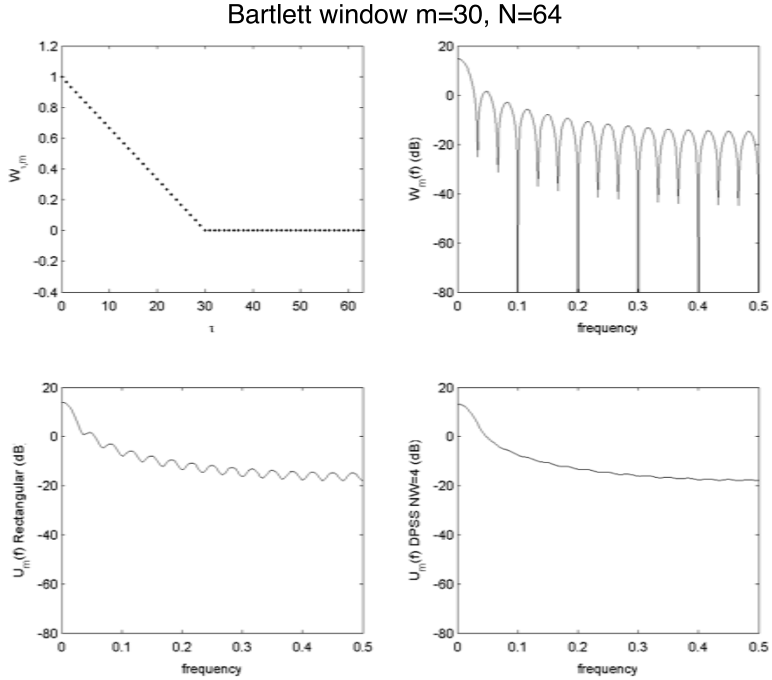
$$\left[\frac{\nu\hat{S}^{(lw)}(f)}{Q_\nu(1-p)}, \frac{\nu\hat{S}^{(lw)}(f)}{Q_\nu(p)}\right]$$

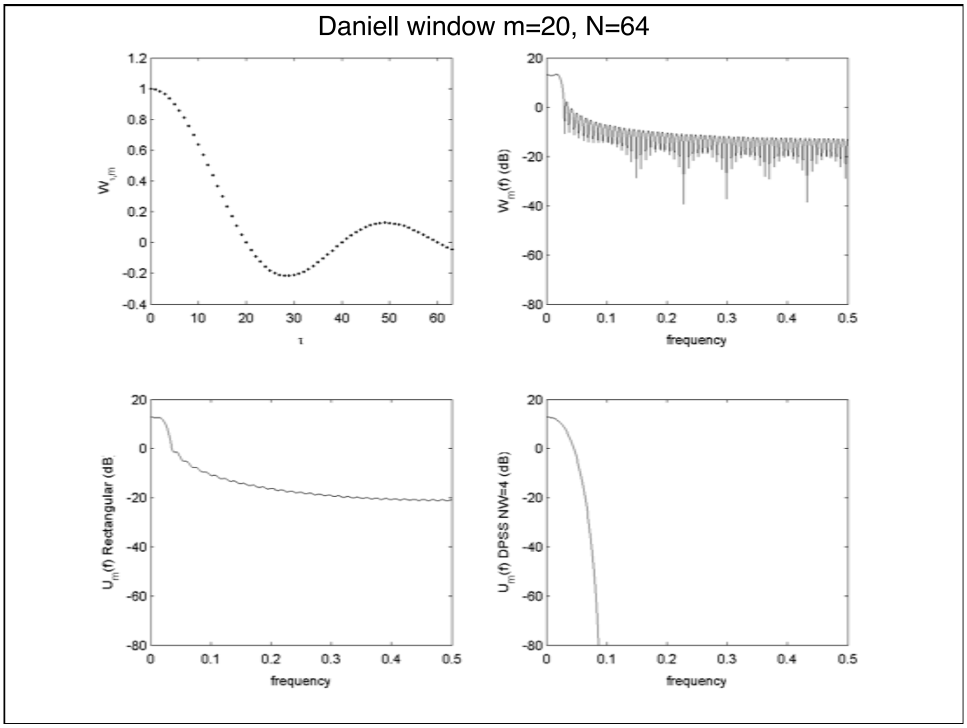
Examples of lag windows

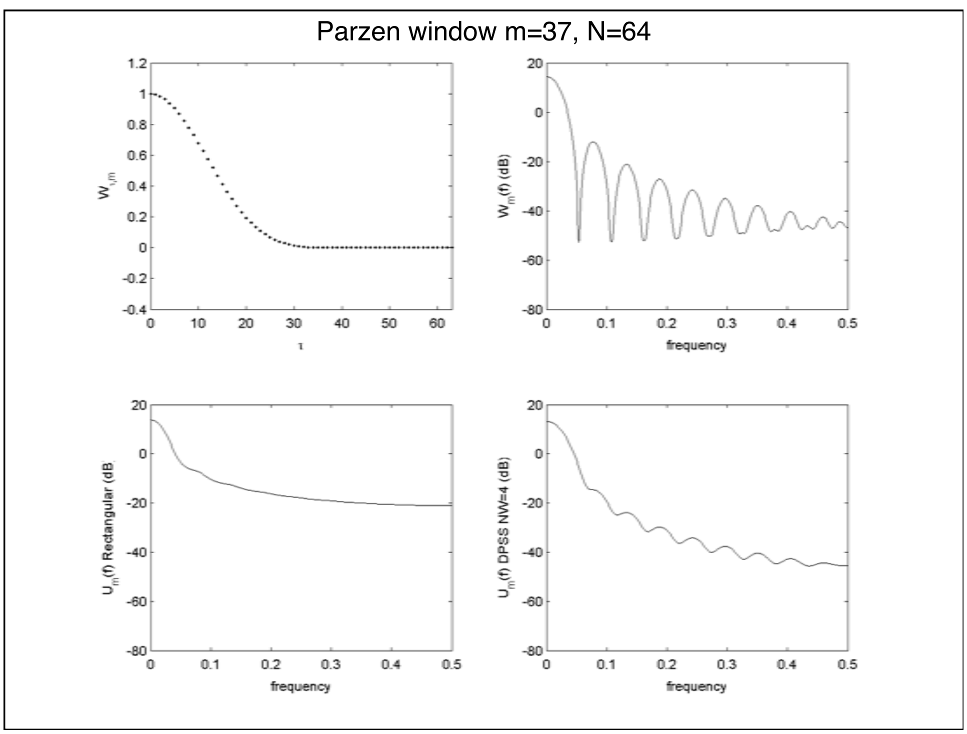
- We show examples for four lag windows. For each window, the lag window versus lag for specific m and N=64, the smoothing window $W_m(.)$, and the spectral windows for a rectangular taper and dpss NW4 taper.
- The lag windows shown are

Bartlett	$w_{\tau,m} = 1 - \lvert\tau\rvert/m \quad \lvert\tau\rvert < m$
Daniell	$w_{\tau,m} = \dfrac{\sin(\pi\tau/m)}{\pi\tau/m} \quad \lvert\tau\rvert < N$
Parzen	$w_{\tau,m} = \begin{cases} 1 - 6(\tau/m)^2 + 6(\lvert\tau\rvert/m)^3 & \lvert\tau\rvert \le m/2 \\ 2(1-\lvert\tau\rvert/m)^3 & m/2 < \lvert\tau\rvert \le m \end{cases}$
Papoulis	$w_{\tau,m} = \dfrac{1}{\pi}\lvert\sin(\pi\tau/m)\rvert + (1-\lvert\tau\rvert/m)\cos(\pi\tau/m) \quad \lvert\tau\rvert < m$

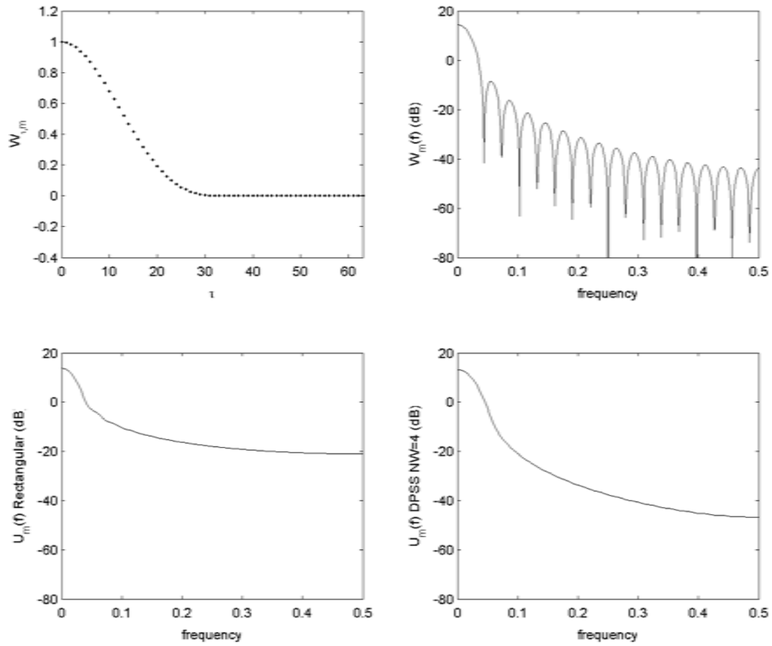
Bartlett window m=30, N=64




Daniell window m=20, N=64
$W_{\tau,m}$
τ
$W_m(f)$ (dB)
frequency
$U_m(f)$ Rectangular (dB)
frequency
$U_m(f)$ DPSS NW=4 (dB)
frequency


Parzen window m=37, N=64
$W_{\tau,m}$
τ
$W_m(f)$ (dB)
frequency
$U_m(f)$ Rectangular (dB)
frequency
$U_m(f)$ DPSS NW=4 (dB)
frequency

Papoulius window m=34, N=64



Characteristics of lag windows

Estimator	Asymptotic variance	ν	B_w
Bartlett	0.67 m $\Im$	3 $\Re$	1.5/(mΔt)
Daniell	m $\Im$	2 $\Re$	1/(mΔt)
Parzen	0.54 m $\Im$	3.71 $\Re$	1.85/(mΔt)
Papoulis	0.59 m $\Im$	3.41 $\Re$	1.70/(mΔt)

$$\Im = C_h S^2(f)/N \quad \Re = N/(mC_h)$$

Welch's Overlapped Segment Averaging

- In lag-window spectrum we smooth the direct estimate spectrum or the autocovariance sequence. In the Welch method, data is broken down into segments of a given length, the spectrum for each block is computed and an average taken of these spectrum.
- In the lag-window approach we loose resolution by smoothing over frequencies; In the Welch approach we lose resolution because the data spans are shorter.
- Two main ideas:
  - Using tapers to reduce leakage and
  - overlapping the blocks for improved variance properties. Overlapping helps because is partly compensates for the down weighting of data at ends of blocks (e.g., Hanning and dpss NW4 type tapers).

WOSA

- Method now know as WOSA (Welch/Weighted Overlapped Segment Averaging).
- The WOSA estimate is given by

$$\hat{S}_l^{(d)}(f) \equiv \Delta t \left| \sum_{t=1}^{N_S} h_t X_{t+l-1} e^{-i2\pi f t \Delta t} \right|^2 \qquad 1 \le l \le N+1-N_S$$

$$\hat{S}^{(wosa)}(f) \equiv \frac{1}{N_B} \sum_{j=0}^{N_B-1} \hat{S}_{jn+1}^{(d)}(f)$$

N is total data; N_S is block size and N_B is number of blocks

n is a shift factor such that $0 < n \le N_S$ and $n(N_B - 1) = N - N_S$

- h_t is the data taper being used.

Statistics of $S^{(wosa)}$

- The expectation is given by

$$E\{\hat{S}^{(d)}_{jn+1}(f)\} = \int_{-f_{(N)}}^{f_{(N)}} \mathrm{H}(f-f')S(f')df' = E\{\hat{S}^{(wosa)}(f)\}$$

- Note the expectation depends only on the block size and not on the total number of data (it also depends on the data taper and the true spectrum).
- The expectation does not depend on the number of blocks or the shift factor.
- When using WOSA it is important that the block size be large enough to capture the features in S(f).

Variance of $S^{(wosa)}$

- The variance of the WOSA estimate is given by

$$\mathrm{var}\{\hat{S}^{(wosa)}(f)\} = \frac{1}{N_B}\sum_{j=0}^{N_B-1}\mathrm{var}\{S^{(d)}_{jn+1}(f)\} + \frac{2}{N_B^2}\sum_{j<k}\mathrm{cov}\{S^{(d)}_{jn+1}(f), S^{(d)}_{kn+1}(f)\}$$

$$\mathrm{var}\{S^{(d)}_{jn+1}(f)\} \approx S^2(f) \text{ and}$$

$$\mathrm{cov}\{S^{(d)}_{jn+1}(f), S^{(d)}_{kn+1}(f)\} \approx S^2(f)\left|\sum_{t=1}^{N_S} h_t h_{t+|k-j|n}\right|^2$$

$$\mathrm{var}\{\hat{S}^{(wosa)}(f)\} \approx \frac{S^2(f)}{N_B}\left(1 + 2\sum_{m=1}^{N_B-1}\left(1-\frac{m}{N_B}\right)\left|\sum_{t=1}^{N_S} h_t h_{t+mn}\right|^2\right)$$

Effective degrees of freedom

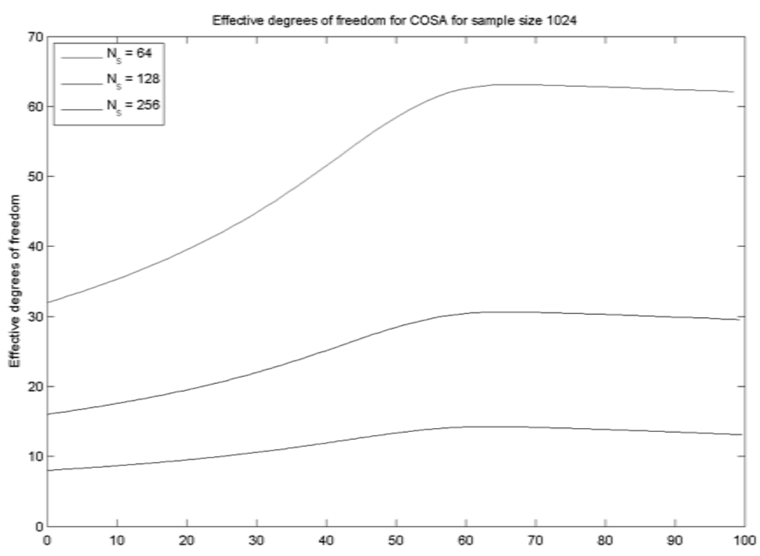
- Based on the variance estimates we can determine the effective degrees of freedom.
- Figure on next page shows results for a Hanning window given below.

$$\nu = \frac{2\left(E\{\hat{S}^{(wosa)}(f)\}\right)^2}{\mathrm{var}\{\hat{S}^{(wosa)}(f)\}} \approx \frac{2N_B}{\left(1+2\sum_{m=1}^{N_B-1}\left(1-\frac{m}{N_B}\right)\left|\sum_{t=1}^{N_S} h_t h_{t+mn}\right|^2\right)}$$

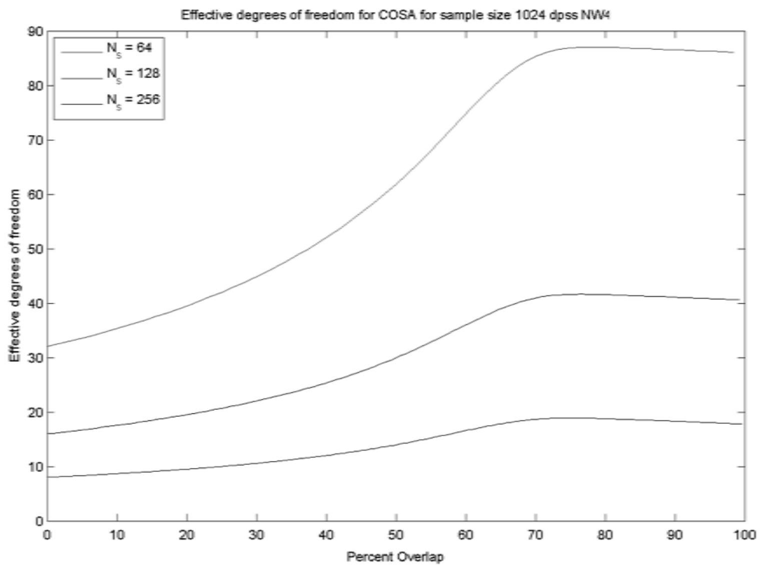
Hanning Window

$$h_t = \left(\frac{2}{3(N_s+1)}\right)^2\left[1-\cos\left(\frac{2\pi t}{N_S+1}\right)\right]$$

Effective dof for Hanning data taper



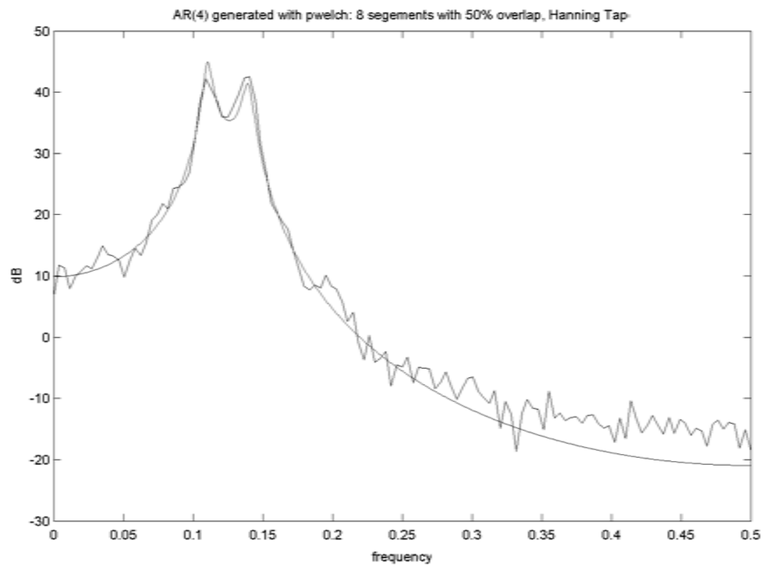
Effective dof for dpss NW=4 taper



Use of WOSA

- The WOSA spectral estimator is widely used because:
  - In can be implemented with fixed length FFTs.
  - Long sequences of data can be handled.
  - Commercial spectrum analyzers have this method built in.
  - A robust sdf estimator can be devised such that individual S(d) are combined weighted so that blocks with outliers are down weighted (Chave et al, 1987)
- Biggest problem can be bias from too small a block size.

Example WOSA



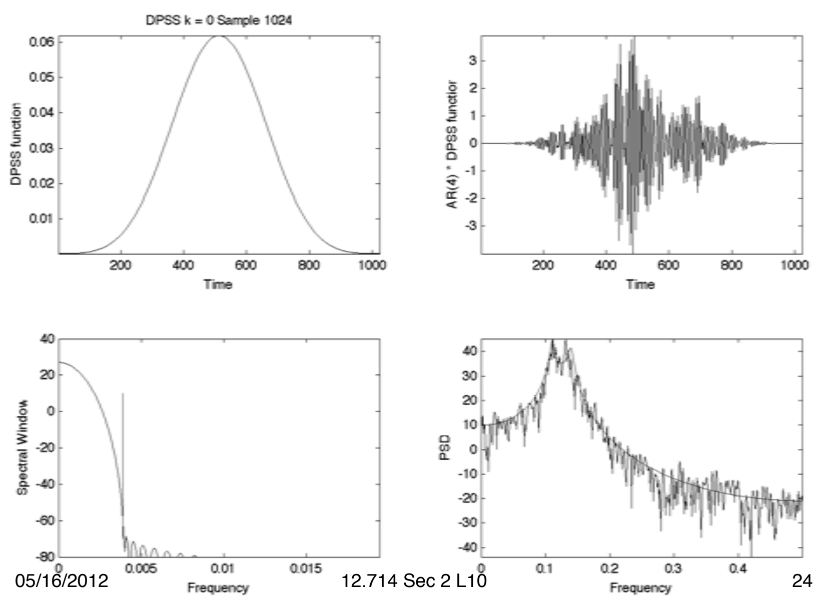
Multi-taper Methods

- The multi-taper methods were developed by Thompson [1982] and address the issue of information lost in the WOSA approach.
  - Multi-taper methods with orthogonal tapers can be used in a reasonably automatic fashion without the design needed for pre-whitening or leakage with WOSA approaches. (Can be used in instrumentation).
  - Bias can be separated into two quantifiable components: local due to band pass and broad-band leakage
  - Considered to generate spectral estimates with more than 2-degrees of freedom.
- Method and example discussed in PW Chapter 7.

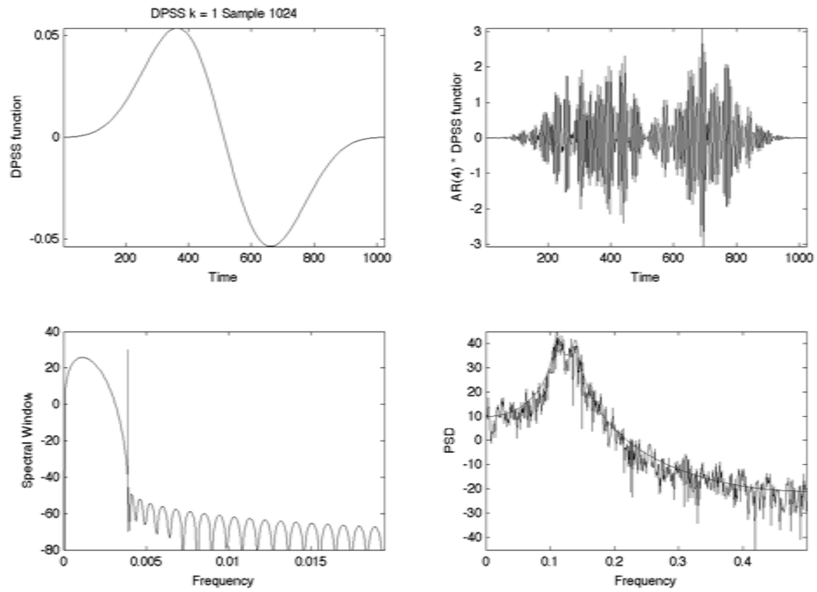
Example

- Following figures and matlab code for lecture show the use of orthogonal DPSS functions as multitapers
- Results from each taper are shown and then the average of the sequences of tapers.
- The noise characteristics of the average to kth order is chi-squared with 2k degrees of freedom divided by 2k (PW p 374 and section 7.1)
- As discussed in PW chapter 7 it is possible to generate an adaptive estimate of the spectral density by weighting the average.

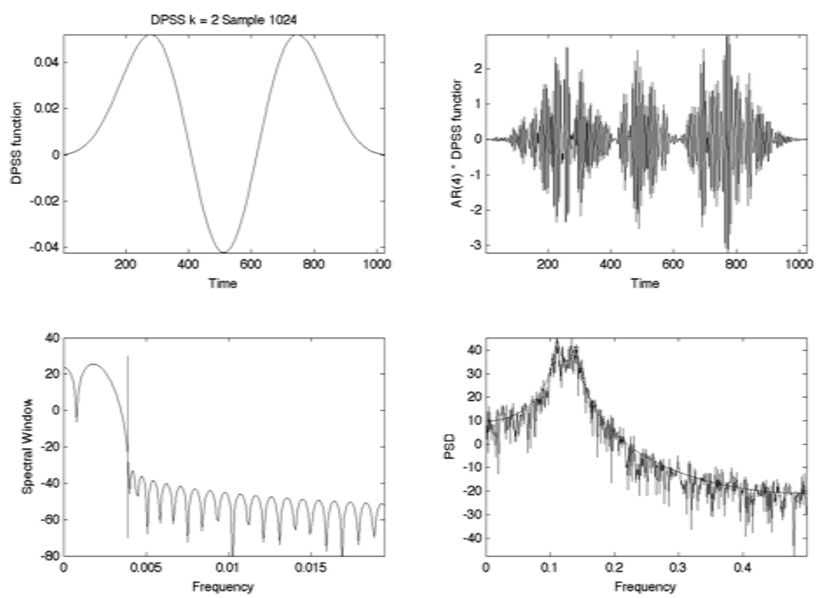
DPSS k = 0



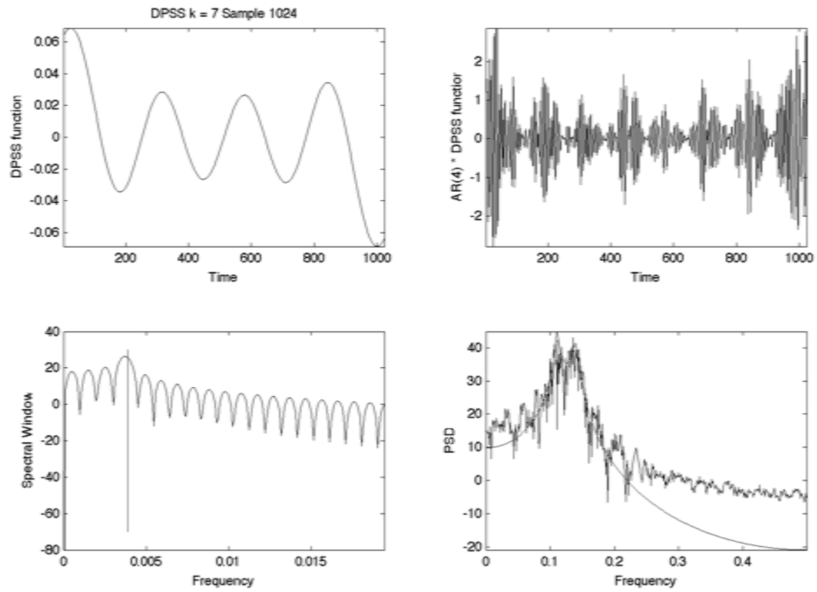
DPSS k = 1



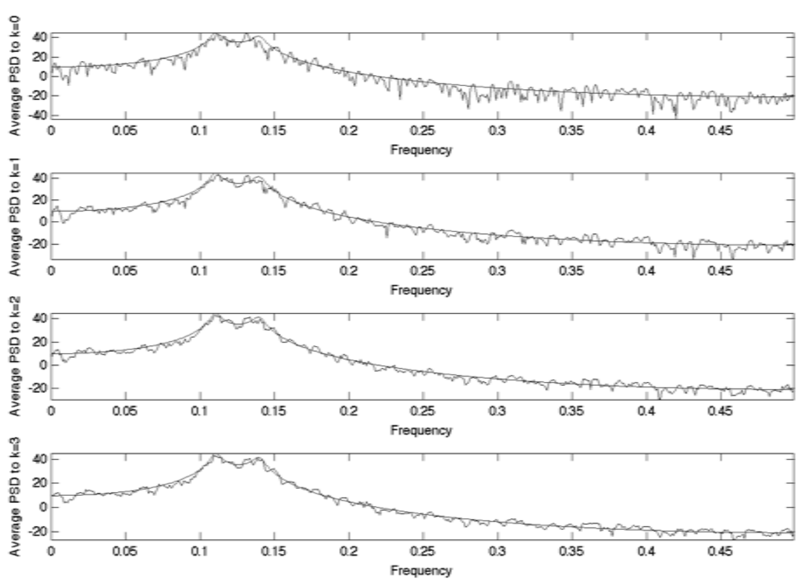
DPSS k = 2



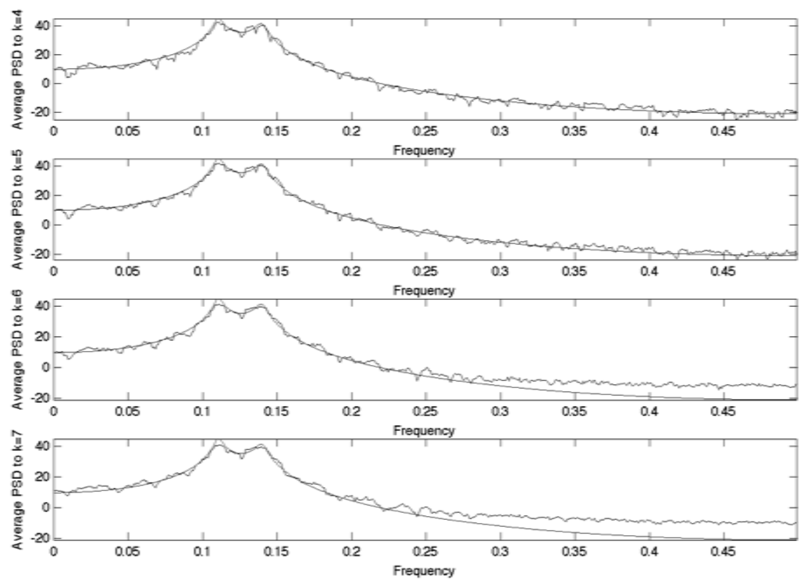
DPSS k = 7



Averaging the multi-taper results



Averaging more (and bias)



Summary of today's class

- Asymptotic distribution of lag window estimators: Allows variance, effective degrees of freedom, and confidence intervals to be computed (valid for large samples)
- Examples of lag window estimators, Bartlett, Daniell, Parzen and Papoulis. Each has its own properties (with Bartlett being closest to wosa)
- Welch's overlapping segment averaging (wosa): Divide data in blocks and then average direct spectra from blocks.
- Example using Matlab pwelch routine
- Exampe of multi-taper applications of Slepian functions
- Remember: High dynamic range needs to be carefully considered when spectral density functions are computed.