# Applied Time-Series Analysis

Arun K. Tangirala

Department of Chemical Engineering, IIT Madras

## Estimation of spectral densities for stochastic signals

# Learning Goals

In this lecture, we focus on the estimation of spectral densities for stochastic signals. We shall learn the following:

- Blackman-Tukey and Daniell's method
- Welch's averaged periodogram
- Parametric methods

# Recap.

The lack of consistency in periodogram estimator can be explained from three viewpoints:

i. The infinitely-long ACVF is approximated by a finite-length **estimated** ACVF and the error in ACVF estimates increases with the lag (leads to Blackman-Tukey estimators).

ii. The true p.s.d. is a **smooth** function of frequency, whereas the estimated one is *erratically* fluctuating (leads to smoothers).

iii. The true p.s.d. is an **average** property, whereas the estimated one is from a single realization (leads to Welch's average periodogram estimators).

# Recap: Methods for improvement

| Method | Summary | Reference |
|---|---|---|
| Blackman-Tukey | Fourier transformation of smoothed, truncated autocovariance function | Chatfield, 1975 |
| Smoothed periodogram | Estimate periodogram by DFT of time series; Smooth periodogram with modified Daniell filter | Bloomfield, 2000 |
| Welch's method | Averaged periodograms of overlapped, windowed segments of a time series | Welch, 1967 |
| Multi-taper method (MTM) | Use orthogonal windows (ÒtapersÓ) to get approximately independent estimates of spectrum; combine estimates | Percival and Walden, 1993 |
| Singular spectrum analysis (SSA) | Eigenvector analysis of autocorrelation matrix to eliminate noise prior to transformation to spectral estimates | Vautard and Ghil, 1989 |
| Maximum entropy (MEM) | Parametric method: estimate acf and solve for AR model parameters; AR model has theoretical spectrum | Kay, 1988 |

# Non-parametric estimators

As remarked earlier, two classes of spectral estimators exist:

1. Non-parametric methods:
    - $\uparrow$ No model is required to be fit to data.
    - $\uparrow$ No major assumptions about the data sequence are made
        - Two assumptions are implicit though: (1) Autocorrelation dies out after $N$ and (2) Data is periodic with period $N$
    - An important goal is to obtain a consistent estimate
    - Reduction in variance is achieved at the cost of decrease in resolution
    - $\downarrow$ Spectral leakage and limited resolution (1/N) are issues

# Parametric estimators

Parametric estimators offer certain striking advantages, but caution has to be exercised.

2. Parametric methods:
   - $\uparrow$ Eliminate the need for windows. Therefore, no leakage issues arise
   - $\uparrow$ Provide better frequency resolution than FFT-based methods
   - $\downarrow$ Require a model-based description of the stochastic process
   - Autocorrelation need not go to zero after $N$. The auto-correlation sequence can be computed from the model

# Non-parametric estimators of PSD

# Method of smoothing the spectrum

## Basic Idea

At each frequency $\omega_i$, estimate the true density by smoothing the periodogram $\hat{\gamma}(.)$ in the vicinity of $\omega_i$.

**Underlying philosophy:** The *true* p.s.d. is constant over a small band of frequencies.

# Method of smoothing

Assuming that the spectral density is constant over a band of $L = 2M + 1$ frequencies (bandwidth of $B_w \simeq L/N$), the optimal estimate is the simple average,

$$\boxed{\hat{\gamma}^{\mathsf{D}}_{vv}(\omega_n) = \frac{1}{L}\sum_{m=-M}^{M} \mathbb{P}_{vv}(\omega_n - \omega_m) = \frac{1}{L}\sum_{m=-M}^{M} \mathbb{P}_{vv}[n-m]} \tag{1}$$

## Taking it further

The foregoing ideas can be generalized to the weighted-averaged smoother.

$$\hat{\gamma}^{\mathsf{D}}_{vv}(\omega_n) = \sum_{m=-M}^{M} W(\omega_m)\mathbb{P}_{vv}(\omega_n - \omega_m) = \sum_{m=-M}^{M} W[m]\mathbb{P}_{vv}[n-m] \quad \text{s.t.} \sum_m W[m] = 1$$

# Procedure

1. **Subtract mean and detrend time series**
   - *Removal of mean* is essential to remove the generally dominant DC component
   - *Detrending* (removal of trends) is necessary to remove any seasonal / linear trends
2. **Compute discrete Fourier transform (DFT)**
3. **Compute the (raw) periodogram**
4. **Smooth the periodogram to obtain** $\hat{\gamma}_{vv}^{D}(\omega_n)$
   - Use an appropriate window to smooth the spectrum

# Choosing the window

Window should be *symmetric*, *non-increasing* and have *unity sum of coefficients*.

- The *modifed* Daniell-kernel places half weights at the end points, while the remaining are distributed uniformly. For example, when $M = 1$, the weights are

$$W[1] = \frac{1}{4} = W[3]; \quad W[2] = \frac{1}{2} \tag{2}$$

- Excessive smoothing (larger $L$) can be detrimental to the spectral resolution while insufficient smoothing can result in higher variance
- A weighted average can be formed by applying the kernel repeatedly to obtain smoother estimates $\equiv$ convolution of the periodogram with convolution of kernels.

# Properties of the smoothed estimator

Asymptotic properties of the smoothed periodogram under two important assumptions (i) the bandwidth $B$ is small, i.e., $L \ll N$ and (ii) the underlying spectral density is roughly constant over this bandwidth.

1. *Variance*:

$$L_M^{-1}\text{cov}(\hat{\gamma}^{\mathsf{D}}(\omega_n), \hat{\gamma}^{\mathsf{D}}(\omega_m)) = \begin{cases} \gamma^2(\omega), & \omega_n = \omega_m = \omega \neq 0, 1/2 \\ 2\gamma^2(\omega), & \omega_n = \omega_m = 0, 1/2 \\ 0, & |n-m| > L \end{cases} \tag{3}$$

# Properties of Daniell smoother . . . contd.

2. *Distribution*:

$$\frac{2L_M \hat{\gamma}^{\mathsf{D}}(\omega)}{\gamma(\omega)} \sim \text{As}\, \chi^2(2L_M) \tag{4}$$

where $L_M = \left( \sum_{m=-M}^{M} W^2[m] \right)^{-1}$.

3. Consistency is guaranteed, i.e., as $L \to \infty, N \to \infty, L/N \to 0$, var$(\gamma^{\mathsf{D}}\hat{(}\omega)) \to 0$.

# Properties of Daniell smoother ... contd.

- The *bandwidth* of the spectral window can be defined as

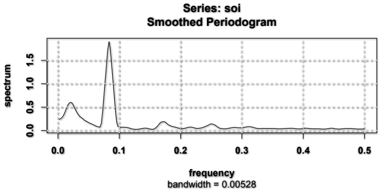
$$B_w = \frac{L_M}{N} \tag{5}$$

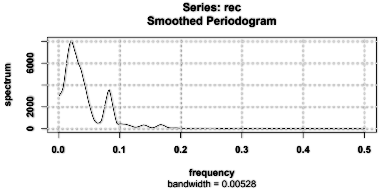
  Thus, the Daniell smoother has $\nu = 2NB_w$ degrees of freedom.

- The window length $L = 2M + 1$ is usually chosen such that $L \ll N/2$.
- The *decrease in variance comes at the price of increase in bias* due to flattening of the spectral density. Pure-tone or very narrowband spectra are smeared as a result.
- The modified Daniell-kernel has less spectral leakage compared to the regular rectangular kernel (simple averaging).

# Example in R

```
# Data from 'astsa' library by Shumway and Stoffer
soi = scan("data/soi.dat")
# Use modified Daniell's filter for smoothing
soi.per = spec.pgram(soi,span=c(7,7),taper=0,log='no')
# Can also use kernel routine to generate filter
kf = kernel('modified.daniell',c(3,3))
# Repeat for recruitment series
```




- The modified Daniell filter applies the regular Daniell filter (of equal weights) twice.
- Observe how the peaks in the SOI and recruitment series have smoothed out.
  - Smoothing comes at the cost of loss in resolution (bandwidth)

# Blackman-Tukey Method: Smoothing the ACF

## Basic Idea

Window the sample ACVF before applying the Fourier transform in (**??**)

## B-T estimate

A weighted ACVF-based estimator is therefore used:

$$\boxed{\hat{\gamma}_{vv}^{(\text{BT})}(\omega_n) = \frac{1}{2\pi} \sum_{l=-(N-1)}^{N-1} w[l]\hat{\sigma}_{vv}[l]e^{-\omega_n l}} \tag{6}$$

where $w[.]$ is the *lag window* function used for smoothing the ACF.

# Remarks

$$\hat{\gamma}^{(BT)}(\omega) = \sum_{l=-(N-1)}^{N-1} w[l]\hat{\sigma}_{vv}[l]e^{-\omega_n l} = \int_{-\pi}^{\pi} \mathbb{P}(\xi)W(\omega - \xi)\, d\xi \tag{7}$$

where $W(.)$ is the Fourier transform of the lag window $w[l]$

$$W(\omega) = \sum_{l=-M}^{M} w[l]e^{-j2\pi l\omega}$$

- Weighting the ACF with the lag window is equivalent to convolution of the "raw" periodogram with the spectrum of the window.
- Establishes the equivalence between B-T estimator and Daniell's smoother!

# Window function for B-T estimator

The lag window $w[l]$ is required to satisfy a set of conditions for the power spectral estimate to be real and non-negative valued

1. *Symmetric*: $w[-l] = w[l]$. Necessary for obtaining real-valued p.s.d. estimate.
2. *Non-increasing with* $|l|$: $0 \leq w[l] \leq w[0]$.
3. *Unity at center*: $w[0] = 1$. From its relationship to the smoothing window, this is akin to the unity sum requirement on $W(\omega)$ stated in (2).
4. *Non-negative DTFT*: $W(\omega) \geq 0, \; |\omega| \leq \pi$.

The fourth requirement is only a *sufficient, but not a necessary* condition.

A rectangular window satisfies all the three conditions but violates the fourth one. On the other hand, a Bartlett (triangular) window satisfies all the four requirements.

# Asymptotic properties of the B-T estimator

Under the conditions $M \to \infty$ as $N \to \infty$ and $M/N \to 0$, we have the following:

1. Asymptotically unbiased estimator: $E(\hat{\gamma}^{(\text{BT})}(\omega)) \approx \gamma(\omega)$
2. Variance of the estimator $\propto 1/N$ (**consistency**),

$$\text{cov}(\hat{\gamma}^{\text{BT}}(\omega_n), \hat{\gamma}_n{}^{\text{D}}(\omega_p)) = \begin{cases} \dfrac{\gamma^2(\omega)}{N\tilde{B}_w}, & \omega_n = \omega_p = \omega \neq 0, 1/2 \\ 2\dfrac{\gamma^2(\omega)}{N\tilde{B}_w}, & \omega_n = \omega_p = 0, 1/2 \\ 0, & |n-p| > L \end{cases}$$

# Asymptotic properties of the B-T estimator

3. Asymptotic distribution:

$$\frac{\nu\hat{\gamma}^{(\text{BT})}(\omega)}{\gamma(\omega)} \sim \text{As}\ \chi^2(\nu), \qquad \text{where } \nu = 2N/\left(\sum_{m=-M}^{M} w^2[m]\right) = 2N\tilde{B}_w \tag{8a}$$

4. As $M$ increases, variance $\downarrow$, but frequency resolution $\downarrow$ as well!

# Bartlett's Method: Averaged Periodogram

## Basic idea

Divide the data of length $N$ into $K$ *non-overlapping* segments of length $L$, i.e., $N = KL$, and take the simple average of the respective $K$ periodograms.

$$\boxed{\hat{\gamma}^B(\omega_n) = \frac{1}{K}\sum_{i=1}^{K}\mathbb{P}^{(i)}(\omega_n)} \tag{9}$$

where $\mathbb{P}^{(i)}(\omega)$ is the periodogram estimate in the $i^{\text{th}}$ segment.

# Bartlett's method

- Slicing of data amounts to generating artificial realizations keeping in view the requirement of the p.s.d. definition.
- Variance as well as frequency resolution drop by a factor of $K$,
- As $K$ increases, variance reduces, but once again at the cost of loss in frequency resolution!

# Welch's method

## Basic idea

Same as that of Bartlett's approach, but with two changes: (i) taper the segmented data and (ii) allow overlapping of segments.

$$\boxed{\hat{\gamma}^{W}(\omega_n) = \frac{1}{K}\sum_{i=1}^{K}\hat{\gamma}^{(i)}(\omega_n)} \tag{10}$$

# Welch's method ... contd.

The quantity $\hat{\gamma}^{(i)}(\omega_n)$ is the *modified* periodogram estimate of the $i^{\text{th}}$ segment,

$$\hat{\gamma}^{(i)}(\omega_n) = \text{PSD}(w[k]v^{(i)}[k]) = \frac{1}{C_w L}\left|\sum_{k=0}^{L-1} w[k]v^{(i)}[k]e^{-j\omega_n k}\right|^2 \tag{11}$$

and $C$ is a normalization factor to account for loss in power due to tapering

$$C_w = \frac{1}{L}\sum_{k=0}^{L-1} w^2[k] \tag{12}$$

# Remarks

- Windowing the segment improves the raw estimate in each segment.
- Overlapping of segments provides two advantages over Bartlett method: (i) for a fixed length $L$, *more segments* can be obtained, (ii) for a fixed number of segments $K$, we can have *longer segments*.
- In principle, $O$ can take on values between 0 (Bartlett method) to $(L-1)$. Welch, (1967) recommends a 50% overlap, $O = L/2$, for which $K = 2N/L - 1$.
- For 50% overlap, the asymptotic variance of the estimator is:

$$\text{var}(\hat{\gamma}^{\text{W}}(\omega_n)) \simeq \frac{9}{8K}\gamma^2(\omega_n) = \frac{16N}{9L}\gamma^2(\omega_n) \tag{13}$$

- Welch recommends the use of Welch (Hann like) or a Parzen window.

# Example: PSD estimation using N-P methods

## PSD Estimation of WN

A series consisting of $N = 1024$ observations of a Gaussian white-noise is generated.

Power spectral density estimates using four different estimators (i) raw periodogram, (ii) WOSA, (iii) B-T and (iv) MTM are computed and shown in Figure 1.

i. *Welch's method:* $L = 128$, $O = L/2$; Hann window; 256-point FFT.
ii. *B-T estimator:* ACVF truncated to $|l| = 127$, Bartlett window; 256-point FFT
iii. *Multi-tapering method:* $NW = 4$; 256-point FFT.

The averaged / modified periodogram estimates are a clear improvement over the periodogram. The dashed line represents the theoretical p.s.d. of a unit variance GWN.

# Example: PSD estimation ...contd.

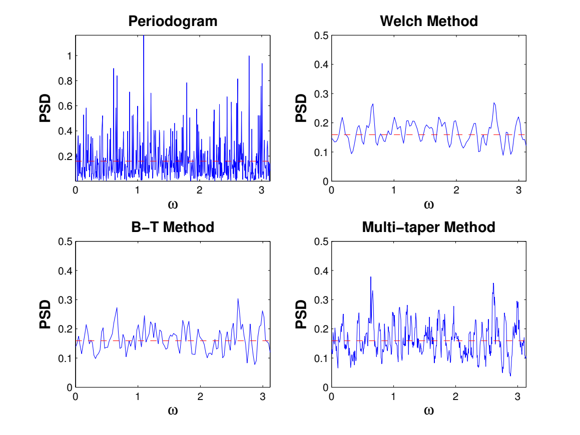



Figure 1: Estimation of PSD of a GWN process using non-parametric methods

- The performance of Welch and B-T estimators are comparable. Incidentally these estimates are better than those offered by MTM.
- However, these observations should not be used to draw any general conclusions.

# Parametric method for PSD estimation

Parametric method of estimating spectrum is a straightforward approach based on the time-series modelling of the given time-series.

## Procedure

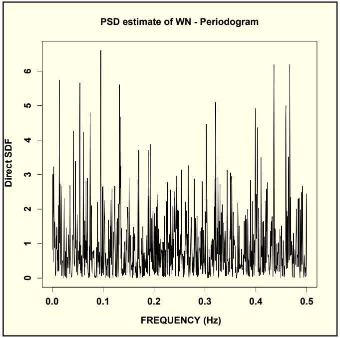
1. Estimate a time-series model $H(q^{-1})$ for the process from the given data.
2. Compute the spectral density function of the time-series $v[k]$ as

$$\gamma_{vv}(\omega) = |\hat{H}(e^{-j\omega})|^2 \frac{\sigma_e^2}{2\pi} \tag{14}$$

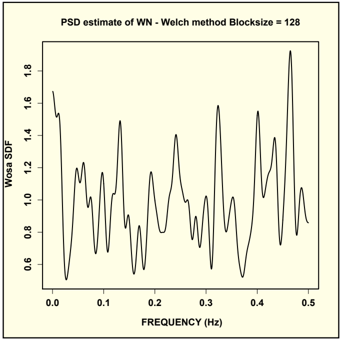
# Parametric method for PSD estimation ...contd.

- Usually, an AR model of suitably high-order is fit to the time-series.
- Remember that this AR model is only an intermediary - meaning, the ultimate aim is to arrive at a spectral estimate and hence we are not particularly concerned about the accuracy of the model
- AR models are the natural choice because they can be estimated easily whereas MA model estimation will invariably involve a non-linear optimization problem

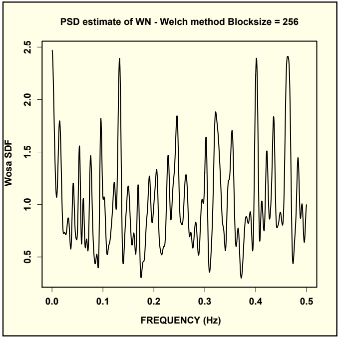
# Comparing methods: Example in R




(a) Periodogram

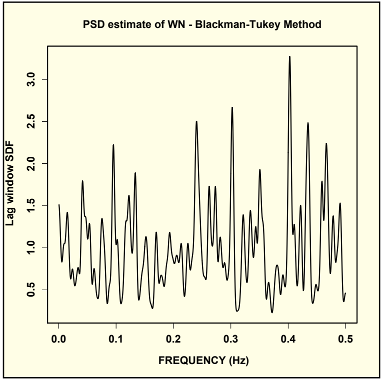



(b) WOSA, M=128

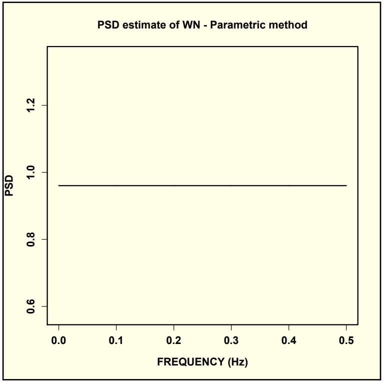



(c) WOSA, M=256

# Comparing methods: Example in R




(d) B-T




(e) Parametric

# Comparing methods: Example in R

- As the block size increases, Welch's method gives lesser smoothing
- When block size = sample size, WOSA returns the periodogram estimate
- In all non-parametric methods, there is an inherent trade-off between bias (frequency resolution) and variance (of p.s.d. estimates)
- Each method has a tuning parameter (block size, window type, order, etc.) Therefore, comparison should be done very carefully.
- Parametric estimators are superior to the non-parametric ones, but quite sensitive to the TS model.
- Always advisable to examine both parametric and non-parametric estimates.

# Comparing methods: Example in R ...contd.

```
ek = rnorm(1000) # GWN
library(sapa)
ek.psd <- SDF(ek,method='direct') # Periodogram
plot(ek.psd,yscale='linear')
# WOSA, Block = 128
ek.psd <- SDF(ek,method='wosa',blocksize=128)
plot(ek.psd,yscale='linear')
# WOSA, Block = 256
ek.psd <- SDF(ek,method='wosa',blocksize=256)
plot(ek.psd,yscale='linear')
# B-T estimator
ek.psd <- SDF(ek,method='lag window',taper.=taper(type='parzen',n.sample=1000))
plot(ek.psd,yscale='linear')
# Parametric estimate of PSD
ek.psd <- spec.ar(plot=F)
plot(ek.psd$freq,ek.psd$spec)
```

# Estimation of cross-spectral density

The cross-spectral density, like the auto-spectral density, is estimated using averaged periodogram methods. A procedure for computing the CPSD follows:

1. Divide the sequences $\mathbf{y}_N$ and $\mathbf{u}_N$ into $K$ overlapping segments, each of length $L$.
2. Compute the c.p.s.d. for the respective pair of segments, $\mathbf{y}^{(i)}$ and $\mathbf{u}^{(i)}$ via DFT:

$$\hat{\gamma}_{yu}^{(i)}(\omega_n) = \frac{1}{L}\sum_{k=0}^{L-1} Y^{(i)}(\omega_n)U^{(i),\star}(\omega_n), \qquad n = 0, 1, \cdots, L-1 \tag{15}$$

where $Y^{(i)}(\omega_n)$ and $U^{(i)}(\omega_n)$ are the standard $L-$point DFTs of the respective *windowed* segments.

# Estimation of CPSD ...contd.

3. Compute the averaged c.p.s.d as in Welch's periodogram:

$$\boxed{\hat{\gamma}_{yu}(\omega_n) = \frac{1}{K}\sum_{i=1}^{K}\hat{\gamma}_{yu}^{(i)}(\omega_n)} \tag{16}$$

# Estimation of Coherence

The coherence function was defined previously as the *normalized* cross-spectral density. Estimation of coherence is, however, carried out by **averaging** estimates across segments.

$$\boxed{\hat{\kappa}_{yu}(\omega_n) = \frac{\hat{\gamma}_{yu}(\omega_n)}{\sqrt{\hat{\gamma}_{yy}(\omega_n)\hat{\gamma}_{uu}(\omega_n)}}} \tag{17}$$

where the cross- and auto-spectral densities using the averaged periodogram method.

*If a single segment is used, coherence estimates take on unity values at all frequencies,* which is clearly a misleading result. It is also necessary to use a *non-parametric method*. The use of a parametric method also results in unity coherence at all frequencies.

# Summary

- Periodogram is an inefficient estimator of the p.s.d. of a stochastic signal
  - For a deterministic signal, it is a very good and natural estimator.
- Spectral leakage is an issue that arises due to finite-length effects.
  - Remedy: Either use large sample sizes or apply tapered windows to data.
- Smoothed / Averaged periodogram methods induce the consistency property at the cost of losing out on the ability to resolve frequencies.

# Summary ... contd.

- Parametric methods have several advantages (e.g., do not suffer from spectral leakage and have very fine resolution) over their non-parametric counterparts. However, they should be used cautiously.
- Cross-spectral density and coherence for stochastic signals should only be estimated using averaged periodogram methods and not using the raw periodogram.
- Parametric estimators also exist for estimating CPSD and coherence, but are not discussed here.

# Bibliography


- Schuster, A. (1897). "On lunar and solar periodicities of earthquakes". In: *Proceedings of the Royal Society* 61, pp. 455–465.
- Tangirala, A. K. (2014). *Principles of System Identification: Theory and Practice*. Boca Raton, FL, USA: CRC Press, Taylor & Francis Group.
- Welch, P. (1967). "The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms". In: *IEEE Transactions on Audio Electroacoustics* AU-15, pp. 70–73.