Chapter 6

# SPECTRAL AND CORRELATION ANALYSIS WITH APPLICATIONS TO MIDDLE-ATMOSPHERE RADARS

Prabhat K. Rastogi

Electrical Engineering and Applied Physics Department
Case Western Reserve University, Cleveland, Ohio 44106

## 1. Introduction

The first Doppler radar observations of waves and turbulence in the stratosphere and mesosphere were reported in VHF experiments conducted at Jicamarca, Perú by Woodman and Guillén [1974]. Doppler radars at frequencies near 450 and 50 MHz, and lately even at 2-3 MHz, continue to be used in extensive studies of middle-atmosphere dynamics. They are collectively called MST radars in view of their ability to probe parts of the Mesosphere-Stratosphere-Troposphere region [Balsley, 1981; Röttger, 1987]. Information about the dynamics of the medium - in terms of its bulk velocity (v) along the radar axis, spread ($\sigma_v$) in this velocity due to turbulence and background wind shears, and on the intensity of refractivity fluctuations ($C_n^2$) induced by turbulence - is obtained from the low-order moments of the power spectrum density of radar signals. The moments of the power spectrum density may also be obtained equivalently from its Fourier transform, the autocorrelation function, often with reduced computations. Indeed, the latter method was used in the early experiments at Jicamarca.

Nearly simultaneous Doppler observations along three or more beams allow measurements of the bulk velocity vector. The measured velocity perturbations are indicative of atmospheric wave-like phenomena. Velocities along coplanar beams, symmetrically offset from the vertical, provide a direct measurement of the vertical momentum flux in the middle atmosphere [Vincent and Reid, 1983]. Power spectrum density is once again of interest in data analysis of time series {v[k]; k=1,2,3...K} of velocity components v, as it yields information about gravity-wave events [Rastogi and Woodman, 1974]

and on the almost turbulence-like ensemble of atmospheric waves [Balsley and Carter, 1982].

In this lecture we review the correlation and spectral analysis methods for uniformly-sampled stationary random signals, estimation of their spectral moments, and briefly address the problems arising due to nonstationarity. Some of these methods are already in routine use in atmospheric radar experiments. Others methods based on the maximum-entropy principle and time-series models have been used in analyzing data, but are just beginning to receive attention in the analysis of radar signals [Klostermeyer, 1986]. These methods are also briefly discussed.

We begin with a recapitulation of random signals (or processes) in Section 2. Several definitions used in the later sections are also introduced here. The nature of radar signals, with several different sampling time scales, and the contribution of unwanted components e.g. system noise and ground clutter, is outlined in Section 3. In Section 4, white Gaussian noise is used as a prototype to illustrate the salient statistical properties of the periodogram, obtained via the squared discrete Fourier transform (DFT). Use of the time-averaged periodogram to estimate the power spectrum density (PSD or power spectrum) of a wide-sense stationary signal is also discussed. In Section 5, methods for estimating the autocorrelation function (ACF) as lagged-product sums, and indirectly through the DFT, are introduced. We emphasize in Section 6 that, for nonstationary signals, the time-averaged periodogram may give a severely distorted estimate of the power spectrum and is not simply related to the true ACF via the Fourier transform. Use of windows or normalized weighting functions to improve the statistical properties of the PSD estimates is discussed in Section 7. The need for windowing and trend removal in spectral analysis of nonstationary signals, and the consequences of coherent integration are also discussed. Spectral parameters or moments can be estimated either directly, or by fitting an assumed shape (e.g. Gaussian or Lorenzian) to the spectral components by using a minimum mean squared error criterion. These fitting methods are discussed in Section 8. An efficient way of estimating the spectral moments from derivatives of the ACF at zero lag is discussed in Section 9. Limitations of this two-pulse technique, so called as a sequence of two closely-sapced

pulses suffices for obtaining the ACF derivatives, are also noted. Finally, high-resolution spectral-analysis methods based on maximizing the entropy for given ACF or data values, and through autoregressive moving-average models of the time series, are briefly introduced in Section 10.

## 2. Random Signals: Recapitulation and Definitions

In this section we review the salient concepts for wide-sense stationary random signals and introduce the definitions of the autocorrelation function (ACF), the power spectrum density (PSD) and spectral moments, and the notion of an estimate. An overall familiarity with the material of this section is assumed. The following recapitulation serves also the purpose of introducing the notation and other definitions used later. Further details may be found in standard engineering texts on random processes [e.g. Davenport and Root, 1958; Papoulis, 1983] and signal analysis [e.g. Stearns, 1975; Oppenheim and Willsky, 1983; Brigham, 1988].

*Random Signals* Suppose we perform some chance experiment $E$ with outcomes and events defined as points ($\zeta$) and subsets in a sample space $S$. A random signal or process $g(t,\zeta)$ is a mapping of these points $\zeta$ to real functions of some independent variable, usually taken as the time (t) or some spatial coordinate. The dependence on $\zeta$ is usually implied, hence $g(t,\zeta)$ is often written as $g(t)$. By a random process $g(t)$ we mean the ensemble of all time functions $\{g(t,\zeta)\}$ with chance outcomes $\zeta$ in the sample space $S$ [see Fig. 2.1]. For a given t, $g(t)$ is merely a random variable. Associated with the random process $g(t)$ are the joint probability density functions of successive orders at times $(t_1)$, $(t_1,t_2)$, $(t_1,t_2,t_3)$ etc.. This allows one to form statistical averages or moments of various products such as $g(t_1)$, $g(t_1)g(t_2)$, $g(t_1)g(t_2)g(t_3)$ etc.. Statistical averaging implies averaging over the entire sample space, i.e. over the ensemble $\{g(t,\zeta)\}$, with respect to an appropriate probability density function.

*Stationarity* An important class of processes that we deal with has joint densities and averages that do not depend on the choice of the time origin. Such random signals are called statistically stationary, or simply stationary. The statistical average or expectation $E[g(t)]=\mu_g(t)=\mu_g$ of a stationary process

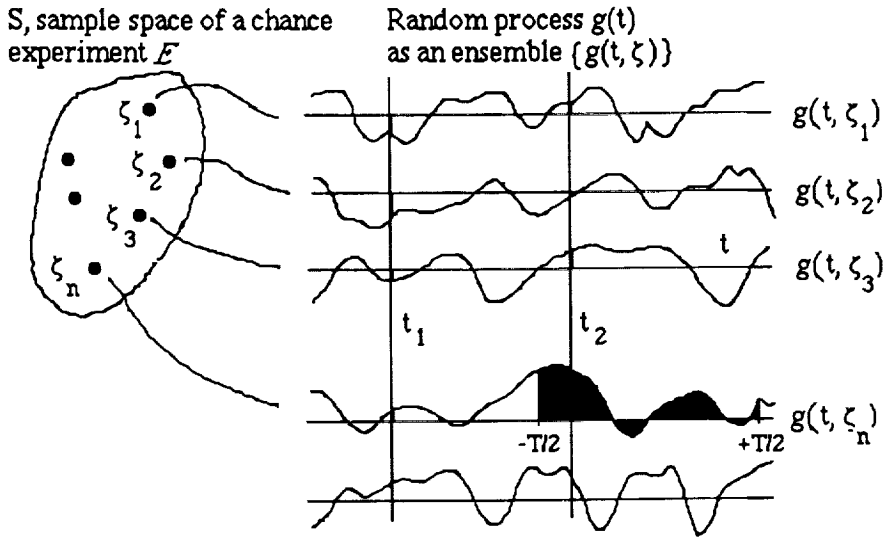


FIGURE 2.1. A random process g(t) as an ensemble of time functions corresponding to the outcomes ($\zeta$) in a sample space (S) for some chance experiment *E*. A suitable probability assignment is defined over S. Averages may be defined in two different ways as discussed further in the text. The time average $m(\zeta_n)$ of a realization $g(t,\zeta_n)$ is obtained by averaging it over a time window (-T/2,T/2) which is eventually made infinitely wide. The ensemble average $\mu(t)$ is obtained by statistical averaging at some fixed time t over all realizations. If the process is stationary and ergodic, then $\mu(t)$ is independent of t, $m(\zeta_n)$ is independent of n, and the two averages are equal.

g(t), evaluated with respect to the density function associated with it at time t, does not depend on t. Its ACF is the second moment defined as the expectation $R_g(t_1,t_2)=E[g(t_1)g(t_2)]=R_g(t_2-t_1)=R_g(\tau)$ of the product $g(t_1)g(t_2)$ of its values at times $t_1$ and $t_2=t_1+\tau$, and it depends only on the time lag $\tau=t_2-t_1$. In a strict sense, stationarity requires that similar conditions should hold for the joint probability densities and moments (or correlations) of all orders. We limit ourselves only to wide-sense stationary processes for which stationarity holds for any two times $(t_1,t_2)$, the average value $\mu_g$ is a constant, and the ACF $R_g(\tau)$ depends only on the time lag $\tau$.

*Time Averages and Ergodicity* A single realization or sample function $g(t,\zeta)$ may be averaged in time over an interval (-T/2,+T/2) or (0,T) of duration T. In a time averaged sense, the mean value of $g(t,\zeta)$ may be obtained as $m_{g,T}(\zeta) = \langle g(t,\zeta)\rangle_T$ and its ACF as $r_{g,T}(\tau) = \langle g(t,\zeta)g(t+\tau,\zeta)\rangle_T$. Higher order averages may be similarly defined. The dependence on the interval duration T is removed by letting it become infinitely wide in the limit. In this limit, $\langle\ \rangle_T$ is denoted by $\langle\ \rangle$. We then find that the time averages $m_g(\zeta)$ and $r_g(\tau,\zeta)$ depend on the identity $\zeta$ of the realization. Do time averages equal statistical averages? Usually not, but if they do then we say that g(t) is an ergodic process. An ergodic process must also be stationary. For an ergodic process, moments can be obtained as time averages over just one long (ideally, infinitely long) realization, as though different segments of the realization correspond to different members in the ensemble. The concept of ergodicity originated in statistical mechanics where it holds well for systems with a large number of molecules. Ergodicity is a useful assumption for atmospheric radar signals, but it is often quite difficult to verify.

*Gaussian Processes* A Gaussian process is one for which the first, second, and higher order probability density functions are jointly Gaussian . These processes are of interest for several reasons. First, it follows from the central limit theorem that a linear combination of many statistically independent identically distributed random variables tends to become Gaussian. In atmospheric radar experiments the scattered signal often arises from many small independent scatterers, hence its probability density functions approaches Gaussian. Exceptions occur when there are only few dominant components, due e.g. to coherent reflections from facets of turbulent layers

or from irregular terrain. Second, the joint probability density functions of any order for a Gaussian process can be expressed in terms of a correlation matrix **R**, i.e. from a complete knowledge of its ACF. Finally, uncorrelated Gaussian variables are also statistically independent. This implies that if the ACF $R_g(\tau)$ of a zero-mean Gaussian random process g(t) vanishes for $\tau > \tau_a$, then successive segments of a realization $g(t,\zeta)$ over windows (0,T), (T,2T),... etc of duration $T >> \tau_a$ become uncorrelated, therefore statistically independent. In essence, a Gaussian process whose ACF has a finite support is also ergodic. Uncorrelatedness does not usually imply independence for non-Gaussian random variables and processes.

*Complex Processes* In radar experiments, the low-pass receiver output z(t) following coherent detection is a complex signal in the following sense. It comprises an in-phase part x(t) after demodulation the received signal with a reference carrier $\cos(2\pi f_o t)$, and a quadrature component y(t) after a similar demodulation with the orthogonal reference $-\sin(2\pi f_o t)$. Since both x(t) and y(t) exhibit random fading, the signal $z(t)=x(t)+\iota y(t)$, where $\iota=\sqrt{-1}$, can be regarded as a complex random process [see e.g. Papoulis, 1983, or Miller, 1974]. The probability density of z(t) is simply the joint density function of {x(t),y(t)}. Higher-order densities are similarly defined as joint densities of x and y at times $(t_1,t_2)$, $(t_1,t_2,t_3)$, etc. Statistical averages of a complex random process are defined with respect to these densities, but may also be evaluated as time averages under the ergodic assumption for a stationary process. Then the mean or average of the process z(t) is a complex constant $(\xi+\iota\eta)$. The autocorrelation function may then be obtained in either of the following equivalent ways

$$R_z(\tau) = E\{z(t)z^*(t+\tau)\} = r_z(\tau) = \langle z(t,\zeta)z^*(t+\tau,\zeta)\rangle \qquad [2.1]$$

$R_z$ ensemble average (independent of t), $r_z$ time average (independent of $\zeta$)

where * denotes the complex conjugate. Different ordering of the lagged term and conjugation gives three other forms, but we use the one above. The signal power $P_z$ defined as $\langle z(t)z^*(t)\rangle$ is real, but the autocorrelation function $R_z(\tau)$ is generally complex. It may be expressed in the cartesian form as $R_z(\tau)=R_{zx}(\tau)+\iota R_{zy}(\tau)$, or in the polar form as $R_z(\tau) = |R_z(\tau)| \exp\{\iota\phi_z(\tau)\}$. It is readily seen that $R_z(\tau)$ has a Hermitian symmetry, i.e.

$$R_z(\tau) = R_z^*(-\tau) \qquad [2.2]$$

which implies that its real part $R_{zx}(\tau)$ and magnitude $|R_z(\tau)|$ are even, but the imaginary part $R_{zy}(\tau)$ and phase $\phi_z(\tau)$ are odd in the time lag $\tau$.

The Wiener-Khintchine theorem relates the ACF $R_z(\tau)$ and the PSD $S_z(f)$ of z(t) as a Fourier transform pair (see e.g. Whalen, 1971; Miller, 1974),

$$S_z(f) = \Im\{R_z(\tau)\} = \int_{-\infty}^{\infty} R_z(\tau) \exp(-\imath 2\pi f\tau)\, d\tau \qquad [2.3]$$

$$R_z(\tau) = \Im^{-1}\{S_z(f)\} = \int_{-\infty}^{\infty} S_z(f) \exp(\imath 2\pi f\tau)\, df \qquad [2.4]$$

The signal power or variance $P_z=\langle z(t)z^*(t)\rangle=R_z(0)$ is obtained by integrating the PSD $S_z(f)$ over the entire frequency range. Since the power in each frequency band (f,f+δf) must be real and non-negative, we infer that the PSD $S_z(f)$ must also be real and non-negative everywhere.

*Periodogram* Each realization of the complex random signal z(t) is a deterministic signal. We assume that it has a Fourier transform Z(f). Its energy spectrum is obtained as $E_z(f) = |Z(f)|^2$. By the Rayleigh energy theorem, the signal energy can be obtained either as the time integral of $|z(t)|^2$ or as the frequency integral of $|Z(f)|^2$. It follows that for signals of finite power $P_z$, the PSD $S_z(f)$ may alternatively be obtained as the time average of $|Z(f)|^2$ over an interval (0,T) as T becomes infinite. Signals with infinite energy or power may be handled by including generalized functions e.g. the Dirac impulse. Consider now a truncated signal $z_T(t)$ which is zero outside the interval (0,T). Then

$$T^{-1}|\Im[z_T(t)]|^2 = T^{-1}|Z_T(f)|^2$$

and the right hand side has properties similar to the PSD $S_z(f)$. It is called the periodogram or sample spectrum. The time-averaged periodogram is often used as an estimate of the power spectrum. The importance of periodogram in power-spectrum estimation of uniformly sampled signals is due mainly to

the availability of efficient Fast Fourier Transform (FFT) algorithms for computing the DFT [see e.g. Cooley et al. 1977; Brigham, 1988]. As we see later, the use of time-averaged periodogram as a power-spectrum estimate requires several assumptions which do not always hold for atmospheric radar signals and data.

*Spectral moments* Radar signals scattered form the atmosphere are slightly Doppler shifted due to bulk atmospheric motions, and also undergo a Doppler broadening due to local fluctuations in the bulk velocity. In the absence of other components in the complex signal z(t) at the receiver output, the PSD $S_z(f)$ has a symmetric off-center peak. The area under the peak corresponds to the signal power $P_z$, its location or center frequency $f_{cz}$ to the Doppler shift $f_d$, and its width $\sigma_{fz}$ about the center frequency $f_{cz}$ to the Doppler frequency spread $\sigma_w$. We note that, except for normalization to unit area, the PSD $S_z(f)$ shares all the properties of a probability density function. Hence the location parameters that we seek may be derived from spectral moments, defined almost identically to the moments $E\{Q^k\}$ of a random variable Q, with respect to its probability density function $f_Q(q)$.

The first few noncentral spectral moments of z(t), denoted here by $s_z^{(0)}$, $s_z^{(1)}$, $s_z^{(2)}$ are are obtained by averaging $f^0$, $f^1$, and $f^2$ with respect to its PSD $S_z(f)$ over all frequencies. The zeroth moment $s_z^{(0)}$ is the same as signal power $P_z$. $S_z(f)/P_z$ is then a probability density function. The location parameters $f_{cz}$ and squared width $(\sigma_{fz})^2$ are obtained in the sense of mean and variance (or the second central moment) of $S_z(f)/P_z$. These may also be derived by transforming $s_z^{(1)}$ and $s_z^{(2)}$ as follows. First, $s_z^{(1)}$ and $s_z^{(2)}$ are normalized by dividing with $s_z^{(0)}$ i.e.

$$s_z^{(1)} \rightarrow s_z^{(1)}/s_z^{(0)} = f_{cz} \text{ and } s_z^{(2)} \rightarrow s_z^{(2)}/s_z^{(0)}.$$

Next $s_z^{(2)}$ is modified as

$$s_z^{(2)} \rightarrow [s_z^{(2)} - \{s_z^{(1)}\}2] = (\sigma_{fz})^2.$$

A Doppler shifted peak of Gaussian shape $P_z\, N(f_{cz},\sigma_{fz}^2)$ is fully specified by the (central) spectral moments $P_z$, $f_{cz}$, and $\sigma_{fz}^2$ as shown in Fig. 2.2.

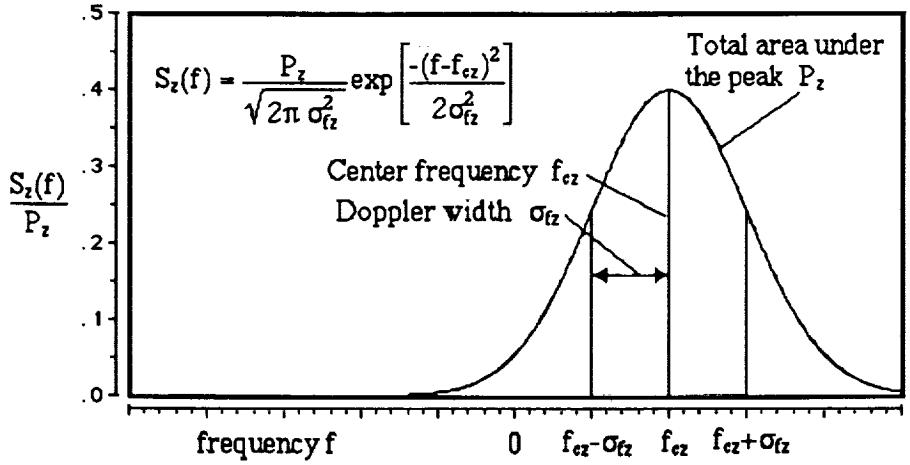


FIGURE 2.2. Power spectrum density and the corresponding spectral moments for an off-center Gaussian spectral peak. Parameters $P_z$, $f_{cz}$, and $\sigma_{cz}$ define the shape of the peak through its area, center frequency and standard deviation. These parameters also correspond to the zeroth, first and second order normalized spectral moments $s_z^{(0)}$, $s_z^{(1)}$, and $s_z^{(2)}$ interpreted as signal power, Doppler frequency shift and Doppler frequency spread. Note that the frequency spread is $\sigma_{cz}$, whereas $s_z^{(2)}$ equals $(\sigma_{cz})^2$.

If the signal z(t) at the receiver output contains components other than the scattered atmospheric signal, then extra steps may be necessary to relate $P_z$, $f_{cz}$, and $\sigma_{fz}^2$ to the signal power, Doppler shift and Doppler spread of the scattered signal. Finally, just as the moments of a random variable may be obtained from successive derivatives of its characteristic function at the origin, it is possible to infer the spectral moments via the autocorrelation function.

*Estimation* In statistical signal and data analysis we frequently estimate a random quantity $\theta$ by some function $\underline{\theta}(\theta_1,\theta_2,.. \theta_n)$ of n data points $\theta_1,\theta_2,.. \theta_n$. There can be many possible estimates of $\theta$, e.g. $\underline{\theta}_1,\underline{\theta}_2,.. \underline{\theta}_m$ etc. We prefer those that satisfy some reasonable properties viz. unbiasedness, minimum variance, and consistency. An estimate $\underline{\theta}$ of $\theta$ is unbiased if the statistical average $E[\theta-\underline{\theta}]$ of the bias or error $e=\theta-\underline{\theta}$ is zero. An unbiased estimate $\underline{\theta}$, on the average, neither overestimates nor underestimates $\theta$ i.e. $E[\theta] = E[\underline{\theta}]$. Of all the available estimates, we also prefer the one(s) whose variance var $\underline{\theta}$ is minimum. It may often be justifiable to use a biased estimate, if it has lower variance. Finally, when the number m of data points is made infinite, we should expect var $\underline{\theta}$ to approach zero, otherwise taking more observations would be futile. In that case we say that the estimate $\underline{\theta}$ of $\theta$ is consistent. It is often possible to obtain a theoretical lower bound on the variance of an estimator using the Cramer-Rao inequality of statistics. An estimator that meets this bound is called an efficient estimator.

## 3. Nature of Radar Signals and Radar Data.

Essential statistical characteristics of sampled radar signals and time series of derived velocity data are summarized in this section. Choice of a suitable spectral-analysis scheme depends critically on these characteristics and the sampling time scales. We also take a first look at the rudiments of spectral-analysis methods using the DFT.

In radar experiments, an amplitude and/or phase modulated pulse train is transmitted in which each pulse has the form $p(t)\exp(\iota 2\pi f_o t)$ at a carrier frequency $f_o$. The carrier term is removed in coherent demodulation, in

which the received signal is effectively multiplied with $\exp(-\iota 2\pi f_o t)$. The receiver should optimally have a bandpass frequency response to match the modulated pulse shape p(t). Hence the receiver bandwidth B about $f_o$ is decided primarily by the the correlation width $T_p$ of the pulse shape p(t). A simple way of defining $T_p$ is as the distance between points at which the magnitude of the ACF $R_p(\tau)$, defined as $\langle p(t)p^*(t+\tau)\rangle$, becomes 1/2 $R_p(0)$. Roughly, it corresponds to the smallest modulation time scale in p(t). Then $T_p$ is nearly equal to the pulse duration for amplitude-modulated pulse trains, but it is approximately equal to the baudlength $T_b$ for binary phase-coded pulses used in high-resolution experiments. The receiver output is sampled in range with a time resolution $T_r$, which should be somewhat less than $T_p$ to avoid undersampling. Typically, $T_r$ is 1-10 μs for a nominal range resolution of 0.15-1.5 km.

The pulses p(t) in the pulse train are repeated at an interval $T_I$, typically about one ms. The fading rate of the received signal is related to the nominal Doppler frequency shift. It is, nevertheless, very much smaller than the Nyquist frequency of ~500 Hz implied by $T_I$. The complex signal z(t) is therefore coherently accumulated or integrated, range by range, over I successive pulses to obtain an effective sampling time $T=I.T_I$. Typical value of I may be 100 in VHF experiments and 10 for the UHF case. The receiver output signal is thus sampled in time as the function of two indices, j and i denoting range and time. After coherent integration, the index i is changed to k corresponding to the coarser time scale $T=I.T_I$. As the signals are analyzed separately for each range, in our subsequent analysis we need only consider a single complex time series z[k]. A range index j and a sampling time T are then implicit.

The complex series z[k] not only includes the scattered atmospheric signal s[k], it also comprises a wide-band noise component n[k] due to the system and sky noise, a very slowly fading ground clutter term c[k] due to sidelobe returns from terrain, vegetation, weather etc., a sporadic interference component i[k] due to unwanted transmitters in the receiver passband, and possibly a residual d.c. or drift d[k] due to slow changes in the receiver circuits. The drift term d[k] is easily removed. Due to the intermittent and

sporadic nature of the unwanted interference, its identification and treatment is done on an ad-hoc basis. The only remaining terms are s, n, and c. The ground clutter component c[k] is the most problematic of these as it is often nonstationary over the measurement interval.

The signal z(t) is sampled in time as z[k] = z(kT). The frequency range for its PSD $S_z(f)$ is then limited to the Nyquist interval $F=(-0.5T^{-1},+0.5T^{-1})$. Any components of $S_z(f)$ outside F are aliased or folded back into it. The aliasing effect is most clear-cut for the wide-band noise component n(t), originally limited by the receiver bandwidth $B >> T^{-1}$. Hence the noise component is aliased many times over. The eventual effect is to impart a nearly flat or white-noise platform to $S_z(f)$, even when n(t) is nonwhite. The slowly-fading ground clutter component should be manifest in $S_z(f)$ as a near d.c. or very low-frequency spike. This would be true if the measurement interval were either too small or too large compared to the typical fading-time for the clutter. We see later in section 6, that the clutter component usually appears as an $f^{-2}$ platform in the PSD estimate.

Only a finite number K of signal samples z[k] is generally available for spectral analysis. The limitation on K is due to finite memory or storage in the on-line processor. An intermediate step in estimating $S_z(f)$ is the K-sample discrete Fourier transform (DFT) of z[k]. The DFT pair is defined as

$$Z[m] = F_K\{z[k]\} = \sum_{k=0}^{K-1} z[k]\, e^{-\iota\, 2\pi km/K} \quad \text{where} \quad m=0,1,...[K-1] \qquad [3.1]$$

$$z[k] = F_K^{-1}\{Z[m]\} = \frac{1}{K}\sum_{m=0}^{K-1} Z[m]\, e^{+\iota\, 2\pi km/K} \quad \text{where} \quad k=0,1..,[K-1] \qquad [3.2]$$

The DFT converts K time samples of z[k]=z(kT) to K samples of its Fourier transform Z(f), evaluated at equispaced frequency points in the Nyquist frequency range F as Z[m]=Z(m/KT). The effect of sampling in the time domain is to render Z(f) periodic outside the Nyquist range. Conversely, due to sampling in the frequency domain, z(t) is also treated as periodic, with a period KT. Thus both z[k] and Z[m] are periodic K-point sequences. Full implications of time and frequency sampling in the DFT pair, and its

equivalence to the continuous Fourier transform, has been discussed by Brigham(1988). The sampled signal z[k] has a finite power, but infinite energy. It can be shown that the following form of Parseval's relation holds for z[k] and Z[m],

$$\sum_{k=0}^{K-1} |z[k]|^2 = \frac{1}{K} \sum_{m=0}^{K-1} |Z[m]|^2 \qquad [3.3]$$

The use of DFT in estimating the PSD, $S_z(f)$, by time-averaged periodograms is examined in the next section.

Spectral analysis of derived parameters, e.g. the time series of a velocity component v[k], is also of interest here. We note that v[k] are samples of a real random process, and the index k denotes either the time or some spatial coordinate with a basic sampling interval. The power spectrum $S_v(f)$ often shows a power-law decay of the form $\alpha f^{-\beta}$ with a spectral index $\beta$. Here f may be a temporal or a spatial frequency. The power-law shape must be limited at the low-frequency end by some frequency $f_L$, else the power in v(t) may become infinite for some $\beta$. Unless the frequency $f_L$ is fully resolved, its effect is manifest in v[k] as a non-stationary trend, similar to the ground-clutter component c[k] in the radar signal z[k]. Implications of such trends in spectral analysis are discussed in Section 6.

## 4. Time Averaged Periodogram Analysis

The sample spectrum or periodogram $P_z(f)$ of a complex signal z(t) has been briefly discussed in section 2. Suppose the signal z(t) is first truncated over an interval of duration D, and $Z_D(f)=\Im\{z_D(t)\}$ is the Fourier transform of the truncated signal $z_D(t)$. Then the periodogram $P_z(f)$ is defined as

$$P_z(f) = \frac{1}{D} |Z_D(f)|^2 \qquad [4.1]$$

In the uniformly-sampled case, $z_D(t)$ is available at K sample points spaced an interval T apart over a total duration D=KT. For simplicity denote these sample values by the sequence $\{z[k], k=0,1....(K-1)\}$. The DFT $F\{z[k]\}$ of

this sequence is another complex sequence {Z[m], m=0,1....(K-1)}. The K points in Z[m] have a frequency spacing $(KT)^{-1}$ or $(D)^{-1}$ over the entire Nyquist frequency interval $\pm 1/(2T)$. The rightmost point Z[K] is excluded as it equals Z[0] by periodicity. The periodogram in the sampled case is defined in analogy with eqn. [4.1] as

$$P_z[m] = \frac{1}{K} | Z[m] |^2 \qquad [4.2]$$

The sum of $P_z[m]$ over all m, after scaling with the frequency spacing $(KT)^{-1}$, gives the signal power $P_z$. The distinction between the symbols used for the periodogram $P_z[m]$ and the signal power $P_z$ should be noted.

In the limiting case we expect that the statistical average of the periodogram will approach the PSD. This actually gives a physically reasonable alternative definition for the PSD,

$$S_z(f) = E \{ \lim_{D \to \infty} \frac{1}{D} |Z_D(f)|^2 \} \qquad [4.3]$$

The above asymptotic equality will not hold for periodogram estimated from samples of a single short realization. Hence we briefly state the statistical and sampling properties of the periodogram defined in equation [4.2] as a PSD estimator. Further details may be found in Blackman and Tukey(1958), Cooley et al. (1977), Koopmans (1974), Marple (1987), and Oppenheim and Schafer(1975).

The periodogram can be computed at any continuous frequency f. The signal z(t), however, has been truncated beyond the interval (0,D) or, in effect, a rectangular window has been applied to it. Hence $Z_D(f)$ is obtained by the convolution of Z(f) with the window transform D sinc(fD). Then $|Z_D(f)|^2$ is similarly obtained by convolving $|Z(f)|^2$ with $D^2\text{sinc}^2(fD)$. The convolving functions are modified slightly for K equispaced samples of z(t); the sinc function is now replaced with the Dirichlet kernel $\sin(\pi fTK)/\sin(\pi fT)$. Those frequency component in $|Z(f)|^2$ that fall exactly at a sampled frequency point, when convolved with $\sin^2(\pi fTK)/\sin^2(\pi fT)$, produce a null response at

all other sampled frequencies. Hence the periodogram values at the sampled frequencies tend to be uncorrelated, provided that the signal z(t) does not have significant intermediate frequency components that fall in-between two adjacent sampled frequencies. This fact has an important bearing in PSD estimation for signals with a strong clutter component, or with a power-law PSD. We also see later that this gives a singularly irregular appearance to the periodogram.

To simplify our discussion of the statistical properties of the periodogram, we assume that $z(t) = z_X(t) + \iota\, z_Y(t)$ is a zero-mean, complex Gaussian noise with variance $\sigma^2$ and a white or flat PSD. The signal power $P_z$ then equals the variance $\sigma^2$, and is divided evenly between the real and imaginary parts $z_X(t)$ and $z_Y(t)$ of z(t). With samples at time spacing T, the PSD $S_z[m]$ should equal $\sigma^2 T$. Since the DFT Z[m] is a linear combination of sample values z[k], it follows that Z[m] is also zero mean and Gaussian. From the definition of DFT given in eqn [3.1] and using uncorrelatedness of adjacent samples of white Gaussian noise, it can be verified that $\text{var}\{Z[m]\} = K\sigma^2$ and it is evenly divided between the real and imaginary parts $Z_X[m]$ and $Z_Y[m]$ of Z[m]. We are interested in the statistics of $|Z[m]|^2 = \{Z_X[m]\}^2 + \{Z_Y[m]\}^2$. We note that a chi-square random variable $\chi_n^2$ with n degrees of freedom (d.o.f.) is obtained by quadratically adding n statistically independent zero-mean Gaussian random variables from a density $N(0,s^2)$. It has a mean $ns^2$ and variance $2ns^4$. It follows then that $|Z[m]|^2$ has simply a chi-square density with two d.o.f.. An alternative and simpler way of arriving at the same result is to note that |Z[m]| has a Rayleigh density, hence $|Z[m]|^2$ has an exponential density, which is the same as the chi-square density with two d.o.f.. Hence $E\{|Z[m]|^2\}=K\sigma^2$ and $\text{var}\{|Z[m]|^2\}=K^2\sigma^4$.These results for the mean and variance of $|Z[m]|^2$ are only slightly modified when the convolutional effect of the Dirichlet kernel is properly considered, and are valid at least locally in the limit of large K.

With the sampling interval T included, and by noting that the area under the periodogram $P_z(f)$ must equal the signal variance $P_z=\sigma^2$, we see that $P_z[m]$ is an asymptotically unbiased estimate of PSD $S_z(f)$ at the sampled frequencies, with an average value $\sigma^2 T$ and a variance $\sigma^4 T^2$. As its variance remains

independent of the sample size, $P_z[m]$ is an inconsistent estimate of the PSD. The standard deviation of the periodogram is $\sigma^2 T$, same as its mean value. We recall that the periodogram values at adjacent sampled frequency points are nearly uncorrelated. However, as the sample size K increases, these points only come closer in frequency without any reduction in their standard deviation. Hence the periodogram usually shows large fluctuations, making it appear more and more jagged as the number K of sample points increases. Examples of this behavior may be found e.g. in Oppenheim and Schafer (1975) and Marple (1987). These results are approximately valid for non-Gaussian noise, as for even modest K the central limit theorem warrants Gaussian statistics for Z[m]. As our analysis is localized in frequency, these results also nearly correct for signals with colored PSD. Then the periodogram $P_z[m]$ has its mean value and its standard deviation approach the local PSD $S_z[m]$ for large K.

For reasons discussed above the periodogram is perhaps the most maligned PSD estimator. Yet, the ease and efficiency with which it can be implemented through FFT algorithms also make it the most frequently used technique for spectral analysis. The FFT algorithms can work in place without additional storage, require only $\sim K \log_2 K$ complex multiply-adds instead of $\sim K^2$ for direct DFT evaluation, and are modular so that repetitive and computation intensive tasks such as bit reversal and sine-cosine computations can be detached from the main program (see e.g. Cooley et al., 1977). The periodogram $P_z[m]$ becomes a usable PSD estimate only after time averaging over many independent sequences of z[m] of length K. We show below that its standard deviation is substantially reduced through averaging.

When M independent periodograms $P_{z[q]}[m]$ for q=1,2.. are averaged, then each point in the averaged periodogram $Q_z[m]$ is obtained by quadratically adding 2M zero-mean Gaussian random variables with density $N(0,0.5K\sigma^2)$. The sum is normalized by division with KM, and then multiplying it with T, to conserve the area under $Q_z[m]$ as the signal variance $\sigma^2$. Hence the mean of the averaged periodogram $Q_z[m]$ becomes $\sigma^2 T$ and its variance, $\sigma^4 T^2/M$. The standard deviation of the time averaged periodogram is then just $\sigma^2 T/\sqrt{M}$. These results are approximate, but the approximations improve for larger K. For an arbitrary PSD $S_z[m]$, it follows that the averaged periodogram $Q_z[m]$

has a mean ~$S_z[m]$ and a standard deviation ~$S_z[m]/\sqrt{M}$. The variance of $Q_z[m]$ tends to vanish as the number M of periodograms averaged together increases. Hence the time-averaged periodogram *is* a consistent estimator of the PSD.

We close our discussion with a relevant example. In a typical UHF radar experiment with a 1 ms pulse repetition interval and coherent integration over 10 pulses, 64 complex samples may be gathered in 640 ms. About 64 sec of observations suffice for averaging over 100 periodograms. The time-averaged periodogram has a standard deviation that is $1/\sqrt{100}$ or 10% of the local PSD value. A Doppler shifted peak which occupies a sixth of the available frequency window, and is 50% above the background noise level, can be readily detected in the averaged periodogram. The total signal power is only about 0.04 of the total noise power for a hypothetical triangular peak. This corresponds to a detectable signal to noise ratio (SNR) of -14 dB with one minute of observations. This detectability criterion may often be difficult to attain in the presence of other dominant components. But the example does illustrate the basic considerations.

## 5. Estimation of the Autocorrelation Function

An alternative approach to estimating the PSD is through the ACF, using the Wiener-Khintchine relations stated in Section 2. These are readily modified for the discrete case using the DFT. The ACF cannot be usually recovered from the time-averaged periodogram estimates of PSD if the signal z(t) has a nonstationary component, and if it does not satisfy certain ergodic conditions that constrain the ACF to a finite support [Papoulis, 1977 and 1983; Marple, 1987] . These conditions are further examined in Section 6 for the MST radar signals. Here we outline a direct and an indirect method of estimating the ACF from data. The use of these estimates in PSD estimation is discussed in Section 7.

As before, suppose z(t) is a realization of a complex, ergodic, wide-sense stationary signal. Its samples z[k] are available at times kT for k=1,2..K. Under the ergodic assumption, the ACF $R_z(\tau)$ can be estimated as a time average. Its estimates $R_z[n]$ are obtained at discrete time lags nT, for indices

$|n| \leq N < K$. The estimate $R_z[n]$ is evaluated as averaged lagged-products of the form $z[i]\, z^*[i+n]$, provided that the indices [i] and [i+n] do not exceed the bounds on k. We consider two different estimates $R^{[1]}[n]$ and $R^{[2]}[n]$ that differ only in the normalization :

$$R^{[1]}[n] = \frac{1}{K-n} \sum_{i=1}^{K-n} z[i]\, z^*[i+n] \;,\; n=0,1,..N < K \qquad [5.1]$$

$$R^{[2]}[n] = \frac{1}{K} \sum_{i=1}^{K-n} z[i]\, z^*[i+n] \;,\; n=0,1,..N < K \qquad [5.2]$$

The estimates for negative n may be obtained either by inter-changing the order of products in the summations, or by using the Hermitian symmetry (see eqn. 2.2) that $R_z[n] = R_z^*[n]$.

Only [K-n] lagged products can be formed at a lag n. The estimate $R^{[1]}[n]$ normalizes the lagged-product sums by the their actual count [K-n]. The second estimate $R^{[2]}[n]$ normalizes these sums by the number K of data points. We may surmise that $R^{[1]}[n]$ should be an unbiased estimate of $R_z[n]$. Though $R^{[2]}[n]$ is biased, it becomes asymptotically unbiased as K becomes infinite. The variance of the unbiased estimate $R^{[1]}[n]$ increases with index n as there are fewer products averaged. For both the estimates, the variance decreases with increasing number K of data points, and eventually vanishes. Hence both the estimates are consistent. To ensure that a sufficient number of products has been averaged at each lag, we require $N/K \ll 1$, with the ratio K/N of ~10 or more usually desirable [Blackman and Tukey, 1958]. The two estimates have nearly identical properties under these conditions. The biased estimate $R^{[2]}[n]$ puts a triangular weight $1-|n|/K$ on the estimated values. This warrants for $R^{[2]}[n]$ the very desirable ACF property that $|R^{[2]}[n]| \leq R^{[2]}[0]$. The unbiased estimate $R^{[1]}[n]$ does not always satisfy the condition $|R^{[1]}[n]| \leq R^{[1]}[0]$. This condition may be readily violated for small K as the variance of $R^{[1]}[n]$ increases with n.

For a given maximum lag index N, the lagged product-sum scheme can be automated using two buffers of size N. New data is sequentially stored in a data buffer, at an address which wraps around the buffer. For each new data

point, all possible N lagged product sums are updated in the second buffer. Normalization can be done to obtain the ACF estimate, once the data buffer has been filled several times around. This scheme is readily adapted for real-time multi-channel signal processing. It was first used by R. M. Harper in 1974 for real-time data acquisition with the Jicamarca radar. The scheme has also been found quite effective for analysis of irregularly spaced data. Since an N-point history of the time-series is always available in the data buffer, the scheme is readily adapted for editing bad data points or outliers using e.g. mean, variance, median, and order statistics of the data.

An alternative and faster method of estimating the ACF is through the use of DFTs [see e.g. Cooley et al. 1977; Oppenheim and Schafer, 1975, Press et al. 1986]. We recall that the DFT of a K-point sequence z[k] is another K-point sequence Z[m], and convolution in time domain is equivalent to a product in the frequency domain. We also notice the similarity of ACF R[n] with the discrete self-convolution $R_{\otimes}[n]$ of z[n]

$$R[n] = < z[i]\, z^*[i+n] >$$
$$R_{\otimes}[n] = z[n] \otimes z[n] = < z[i]\, z[n-i] >.$$

These operations yield (2K+1)-point sequences with zero end values. The only difference between R[n] and $R_{\otimes}[n]$ is that in convolution one of the terms is folded in the time index i, and in ACF one of the terms is conjugated. Hence ACF may be obtained as $R[n] = z[n] \otimes z^*[-n]$ using the convolution. In the frequency domain, the DFT of R[n] is merely the product of Z[m] with Z*[m]. The only caution that needs be exercized is that R[n], hence its DFT must be at least 2K-points long. The method then is to augment or extend the K-point sequence z[n] with K zeros. The 2K-point DFT's of the extended 2K-point sequences $z_e[n]$ and $z_e^*[-n]$ are then multiplied point by point. Finally, the 2K-point inverse DFT gives the 2K-point periodic sequence R[n]. The estimate thus obtained is weighted by a triangle as for $R^{[1]}[n]$. The method can be readily extended to the cross-correlation function (CCF) $R_{xy}[n]$ of two complex K-point sequences x[k] and y[k]. We merely note the following relations

$$R_{x\otimes y}[n] = x[n] \otimes y[n] = < x[i]\, y[n-i] >.$$
$$R_{xy}[n] = < x[i]\, y^*[i+n] > = x[n] \otimes y^*[-n] = F^{-1}\ [\, X[m]\ Y^*[m]\, ]$$

which suggest that the K-point sequences x[n], y[n] must first be augmented with K zeros to get the 2K-point sequences $x_e[n]$ and $y_e[n]$, one of which is conjugated and inverted in time to get $y_e^*[-n]$, as was also tacitly done for the ACF. The CCF is obtained as the inverse DFT of the point by point product X[m] Y*[m] of the DFTs of these sequences. Averaging over several K-point data sequences is desirable to reduce the variance of ACF and CCF estimates.

This method has several advantages over the direct ACF estimation using lagged-product sums. The DFT (or FFT) computations can be carried out in-situ. When 2K is of the form $2^\kappa$, the number of complex multiplies and adds in the FFT can be made as small as $\sim 2K\kappa$. This computational advantage becomes quite significant even for short data sequences. The PSD estimate, moreover, is available as an intermediate step and it *is* related to the ACF estimate $R^{[2]}[n]$ by the DFT. However, augmenting the data sequence with zeros also doubles the storage requirements. It is perhaps for this reason that this method has not been used in real-time MST radar signal processing. The declining cost of computer memory certainly favors its use.

## 6. Nonstationarity and Spectral Distortion

In the foregoing discussion we have assumed that the complex signal z(t) is a wide-sense stationary and ergodic random process. Usually several sets of K equispaced samples z[k] at sample spacing T are available from a single realization $z(t,\zeta_0)$ . The assumption of wide-sense stationarity implies that the low-order moments viz. the mean $\mu_z$ and the variance $\sigma_z^2$ of the process are constant, and its ACF $R_z(\tau)$ depends only on the time lag $\tau$, irrespective of the time origin. The ergodic hypothesis is invoked to circumvent statistical averaging, by estimating these quantities as time averages over many statistically independent sub-sets from a single realization.

A constant mean value $\mu_z$ contributes a platform of fixed height $\mu_z\mu_z^*$ to the ACF, and a single spike of height $(TK)\mu_z\mu_z^*$ exactly at the zero frequency in the K-point PSD estimate. If the mean $\mu_z$ is indeed a constant, then it can be

effectively removed from z[k], rendering it a zero-mean process in further analysis. The assumption of stationarity of mean is merely a convenient model for the signal time series z[m]. It is readily violated in situations described below making $\mu_z(t)$ a slowly varying function of time with discernible trends over the observation interval.

The ground clutter component c(t) in radar experiments arises due to multiple paths to terrain seen through the antenna sidelobes. Its fading time varies from fraction of a second to minutes due to atmospheric refraction along the paths. When the same path is not traced back due to multiple reflections, c(t) also has a very small Doppler shift. Fading time and Doppler shift of c(t) critically depend on the radar frequency, radar location and on severe weather conditions. Nonstationarity of c(t) is most serious for the ~450 MHz UHF radars. The same refractive multipath effects are nearly an order less severe and nearly insignificant for the ~50 MHz VHF radars. Coherent reflections at near vertical incidence from planar or slightly curved turbulent layers also produce a slowly fading component. Non-stationarity is also evident in the velocity data v(t), especially when these are indicative of a power-law PSD, as slow trends at time scale of several hours to several days.

Removal of a nonstationary trend $\mu_z[k]$ from a single K-point sequence z[k] is difficult unless K is very large or many contiguous K-point sequences are available. Subtracting the mean value <z> from the points z[k] in a sequence does not remove the trend. Gottman(1985) describes simple methods for identifying and removing trends. These methods use averaging and differencing at several time scales to estimate parameters of an ad-hoc linear or quadratic trend model. Alternatively, the parameters of a low-order polynomial that models trend can be found by computation-intensive least-square methods [see e.g. Hamming, 1973, Press et al., 1986].

Nonstationary trends produce a severe distortion of time-averaged periodogram estimates obtained by DFT methods as convincingly discussed by Sato and Woodman (1982). Due to this distortion, ACF cannot generally be recovered from time-averaged periodogram estimates $Q_z[m]$ of the PSD. Suppose the N-point periodogram $P_z[m]$ is formed from an N-point sequence z[k] using its N-point DFT Z[m]. From the same sequence a (2N-1)-point

ACF estimate $R_z[n]$, for n ranging over $\pm(N-1)$, can be formed as $\langle z[i]z^*[i+n]\rangle$. A zero value can be added at either end. Now both $P_z[m]$ and its inverse DFT $P_z^{-1}[n]$ are periodic N-point sequences. We expect the periodic N-point sequence $P_z^{-1}[n]$ and the aperiodic 2N-point sequence $R_z[n]$ to be related. Thus $P_z^{-1}[n]$ is derived from $R_z[n]$ by wrapping it around a circle with N points indexed from 0 to (N-1). If $R_z[n]$ is constant at all lags, or if it is zero for $|n| > N/2$, then $P_z^{-1}[n]$ unambiguously contains all the information about $R_z[n]$. However, if the support of $R_z[n]$ exceeds $\pm N/2$, then $P_z^{-1}[n]$ is severely distorted by wrap-around and its DFT, the periodogram $P_z[m]$, is no longer a reasonable PSD estimate. The problem can be alleviated with the use of a 2N-point DFT with N-point data (extended by zero-padding) to estimate both the PSD and triangular-weighted ACF estimate $R^{[2]}[n]$ as outlined in the previous section.

An alternative way to explain the periodogram distortion is to realize that the true PSD of the trends is a narrow spectral spike near, but not exactly at, the zero frequency. The use of a uniformly weighted N-point sequence z[n] in periodogram estimation smooths this spike by convolution with a squared Dirichlet Kernel which can be approximated with $\text{sinc}^2(fT)$ for the continuous case. The contribution of the spike thus leaks or spills over all frequencies, and is evident at the sample points of the periodogram as an $\sim f^{-2}$ platform. Due to sampling in time at spacing T, tails of the $\sim f^{-2}$ platform are also aliased into the Nyquist window (-0.5/T,+0.5/T). We discuss some ways of containing this leakage in the next section.

## 7. Windowing and Coherent Integration

The PSD $S_z[m]$ of an N-point sequence z[n] sampled at time steps T can be estimated either directly from the N-point DFT Z[m] via the periodogram $P_z[m]$, or as the DFT of an ACF estimate $R_z[n]$. Use of uniform weights or the default rectangular time window is equivalent to a circular convolution of Z[m] or $R_z[n]$ with the Dirichlet kernel $\sin(\pi NfT)/\sin(\pi fT)$. A sinusoid of frequency f ' is seen to leak at other frequencies f in the periodogram $P_z[m]$ as

$$\sin^2(\pi N(f'-f)T\}/\sin^2\{\pi(f'-f)T\}.$$

This leakage eventually decays only as $\sim(f'-f)^{-2}$. In PSD estimates obtained as DFT of the ACF, the Dirichlet kernel produces undesirable negative ripples whose magnitude decreases as $\sim|(f'-f)|^{-1}$. These effects are similar to the familiar Gibbs phenomena in the Fourier reconstruction of signals near discontinuities. The PSD estimates can be improved by shaping the data z[n] or ACF $R_z[n]$ with a suitable window. Since the sampling and aliasing effects in PSD estimation have already been considered in detail, window properties are discussed below in terms of the continuous variables t, $\tau$, and f. The subscript z is also dropped for clarity.

In their classical monograph, Blackman and Tukey(1958) advocated the use of shaping the ACF $R(\tau)$ by multiplication with a window or weighting function $w(\tau)$ that depends on the lag $\tau$. The windowed PSD estimate $S_W(f)$ is obtained by convolving the true PSD S(f) with the window transform $W(f) = \Im\{w(\tau)\}$. $S_W(f)$ has better statistical properties due to smoothing in frequency by W(f). To conserve the signal power R(0), lag windows $w(\tau)$ used with the ACF are normalized to have w(0)=1. Other desirable attributes of $w(\tau)$ are a smooth decaying shape as a function of time lag $\tau$, an even symmetry about $\tau$=0, and negligible negative sidelobes in the transform W(f). Good ACF windows are further selected to be well-behaved in frequency by requiring that the transform magnitude |W(f)| has a small width, and a low sidelobe level that decays sufficiently steeply with f.

A data window d(t) can be directly applied as a weighting function to the signal z(t) before periodogram analysis. The windowed periodogram estimate $P_D(f)$ is now obtained by convolving S(f) with the squared window-transform $|D(f)|^2$. Data windows share nearly all the properties of ACF windows, now stated in terms of d(t) and $|D(f)|^2$. The only major differences are that d(0) need not be 1, the PSD estimates with data windowing are always non-negative, and the signal power is modified because z(t) is scaled by d(t). Due to peaked shape of a data window d(t), the values of z(t) near the end points are not fully utilized. For this reason, as much as half of one set of K points of z[m] can be used with the next set. This method of data windowing with partially overlapping data segments has been described by Welch (1967), who also discusses the statistical properties of the windowed time-averaged periodogram.

A very complete description of many windows, their transform properties, and criteria for their selection has been compiled by Harris (1978). Corrections to some of these are given by Nuttal(1981) who also discusses sidelobe properties of some preferred windows. Rabiner et al. (1979) give code for generating a few frequently used windows, including von Hann, Hamming, Kaiser and Dolph-Chebyshev. The Dolph-Chebyshev window attains a uniform sidelobe level and is described through its transform W(f). The Kaiser window is a time-domain approximation to this window in terms of the modified Bessel function $I_0(x)$ of zeroth order. These windows are nearly ideal for data-processing applications.

Some of the simpler windows are given below as lag windows w(t) for a support (-0.5,+0.5) of t. The rate at which their sidelobes in |W(f)| eventually decay with f is also indicated.

| | | |
|---|---|---|
| Hamming | $w(t) = 0.54 + 0.46 \cos(2\pi t)$ | $\sim f^{-1}$ |
| von Hann (or Hanning) | $w(t) = 0.50 + 0.50 \cos(2\pi t)$ | $\sim f^{-3}$ |
| Approximate Blackman | $w(t) = 0.42 + 0.50 \cos(2\pi t) + 0.08 \cos(4\pi t)$ | $\sim f^{-3}$ |

The Hamming window minimizes the first sidelobe for a simple cosine shape but its transform decays as $\sim f^{-1}$ due to the rectangular platform of height 0.08. The von Hann and the approximate Blackman windows have a better sidelobe behavior. In the analysis of power-law PSD's, it may be desirable to use windows with a steeper side-lobe decay. The Blackman window can be modified by including higher-order cosine terms. The coefficients can be selected in such a way that with m cosinusoids, the frequency response decays at the rate $|f|^{-(2m+1)}$. Two examples of modified Blackman windows are given below.

| | |
|---|---|
| Modified Blackman : order 2, highest term $\cos(4\pi t)$ | |
| Coefficients (0.375, 0.500, 0.125) | $\sim f^{-5}$ |
| Modified Blackman : order 4, highest term $\cos(8\pi t)$ | |
| Coefficients(0.2734375, 0.4375000, 0.2187500, 0.0620000, 0.0078125) | $\sim f^{-9}$ |

The time-domain shape of these windows is shown in Fig. 7.1. The response of the modified fourth-order Blackman window is shown in Fig. 7.2 with its

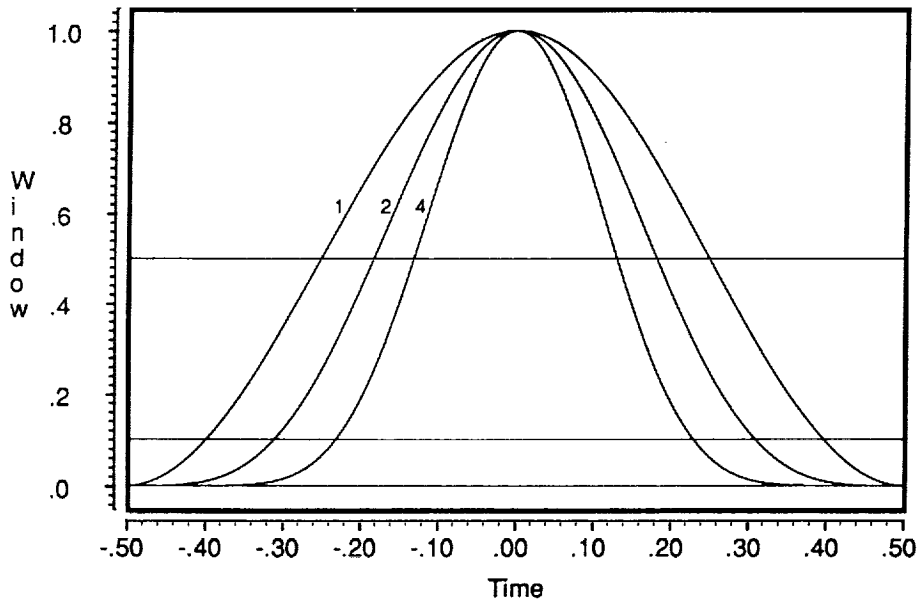


FIGURE 7.1. Time windows of order 1, 2 and 4 with good sidelobe behavior derived from the Blackman window are shown on a support (-0.5,0.5). The order 1 window is just the von Hann or Hanning window with a frequency response decaying at 60 dB/decade. The order 2 and 4 windows have a response decaying at 100 and 180 dB/decade respectively. The effective temporal width of these windows is one-half to one-fourth of their support. For a frequency resolution comparable to the rectangular, data length should then be two to four times longer.

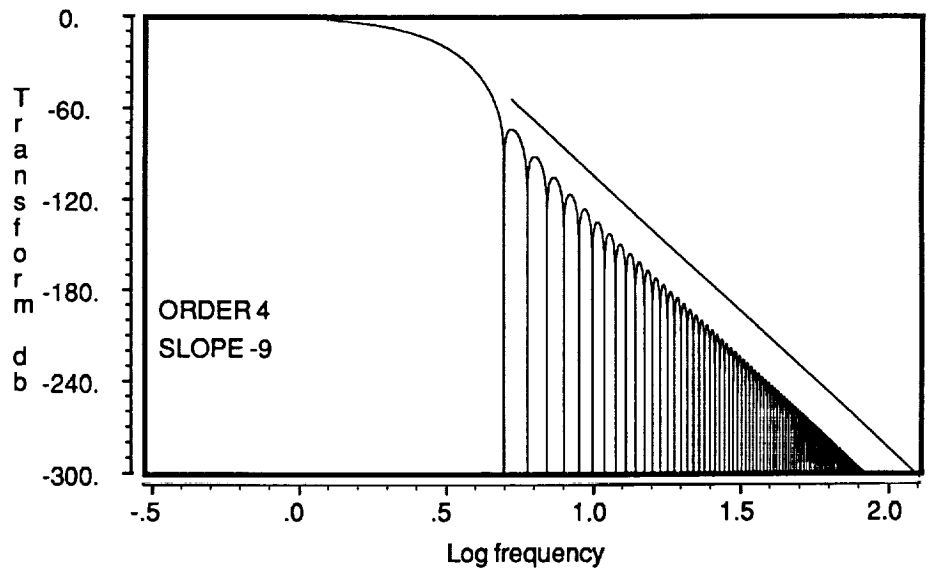


FIGURE 7.2. Frequency response of the fourth order window shown in Fig. 7.1 is shown to decay at 180 dB/decade. This and other windows with well-constrained side-lobe behavior may be useful in spectral analysis of velocity data with power-law spectra, and in suppressing the smearing of ground clutter in radar signal spectra by using longer record lengths.

$\sim f^{-9}$ decay rate. Simulations indicate that windows with well constrained sidelobes are effective in reducing the influence of trends, but require at least two to four times longer data segments. It may be surmised that the use of the modified Blackman windows, or any other suitable window, in the time-averaged periodogram method can contain the effect of fading ground clutter to near zero-frequencies.

We now briefly mention the effect of coherent integration of radar signals in PSD estimation with periodograms [Rastogi, 1983]. In coherent integration, I successive samples of z[i] at a time spacing $T_I$ are averaged with uniform weights (1/I) and the averaged sequence y[i] is re-sampled with time spacing $T=IT_I$. The periodogram $P_y(f)$ of $y[k]=y(kIT_I)=y(kT)$ is formed at K frequencies in the Nyquist interval $\pm 0.5(KT)^{-1}$ using the DFT.

The consequence of time averaging is to multiply the original periodogram $P_z(f)$ with a filter weighting function

$$|H(f)|^2 = \frac{1}{I^2} \frac{\sin^2(\pi f T_I I)}{\sin^2(\pi f T_I)} \qquad [7.1]$$

This filter function has maxima at multiples of $1/T_I$. Between any two maxima, there are (I-2) secondary peaks with nulls at multiples of $1/(IT_I)$. The principle lobe at zero frequency, with adjacent nulls at $\pm 1/(IT_I)$, is twice as wide as the Nyquist interval. Echoes with Doppler shifts near the end points of the Nyquist interval are weighted down by nearly -4dB. A correction for this effect must be applied in spectral-moment estimation. Any components of $P_z(f)$ outside the Nyquist interval are weighted by the filter function of equation [7.1], and would then appear aliased in $P_y(f)$. Hence the coherent integration scheme is not very successful as an anti-aliasing filter.

Coherent integration does provides a computationally efficient means, through simple accumulation, of implementing a 'poor' matched filter for radar signals. Its principal advantage is in reducing the overall data rate by a factor I. The received signal z(t) is originally constrained by the receiver bandwidth B. Sampling at interval $T_I >> B^{-1}$ aliases the entire received signal,

including noise and interference, into the frequency interval $\pm 0.5(T_I)^{-1}$. This frequency interval is further reduced to the Nyquist interval $\pm 0.5(IT_I)^{-1}$ through coherent integration. Obviously, the white-noise power outside the Nyquist interval is rejected by weighting with the filter function and its contribution is reduced by ~1/I. But within the Nyquist interval, the Doppler shifted signal peaks and white noise component are both weighted by the same filter function. Hence the detectability of spectral peaks, as discussed in Sec. 4, is not improved in any tangible way through coherent integration and there is definitely some impairment near the ends of the Nyquist interval.

## 8. Least Squares Estimation and Spectral Parameters

The general problem of estimating parameters from observations or data can only be examined within the frame work of a model. For any choice of parameter values, the model produces an output, which generally differs from observations. That choice of parameter values for which the model output matches the observations, in some statistical sense e.g. by minimizing the mean squared error (m.s.e.), can be said to agree with or derived from the observations. The behavior of m.s.e. as a function of model parameters may be visualized as an error surface. The best choice of parameters corresponds to the true or global minimum on this surface. An exhaustive search for the true minimum is impractical, so an acceptable local minimum is sought only within a limited region of parameter values.

With an initial guess of parameter values, it is possible to seek a local minimum in m.s.e. by using any of the several adaptive search strategies e.g. by changing parameters in the direction in which the m.s.e. changes most steeply. Excellent discussion of least mean square (l.m.s.) algorithms may be found in Alexander(1986), Bard(1974), and Widrow and Stearns(1985). Sato and Woodman(1982) have adapted Bard's formulation to spectral parameter estimation in radar experiments at Arecibo. Their approach is discussed below.

Suppose the observations $\mathbf{X} = x(k)$ represent an N-point vector. The model input is a parameter vector $\mathbf{P} = p(j)$ with J points. The model output $\mathbf{Y(P)} = y(k,\mathbf{P})$ is an N-point vector that depends on $\mathbf{P}$. The error vector $\varepsilon(\mathbf{P}) =$

$\varepsilon(k,\mathbf{P})$ varies with observation index k and depends on the choice of **P**. Since $\varepsilon(k,\mathbf{P})$ may be either positive or negative, we seek to minimize its accumulated square value (which divided by K is the m.s.e.)

$$e(\mathbf{P}) = \sum_{k=1}^{N} [\, y(k,\mathbf{P}) - x(k) \,]^2 \qquad [8.1]$$

with respect to **P**. Equating the derivative of e(**P**) with respect to **P** gives J conditions for each of its component p(j); j=1,2...J

$$\sum_{k=1}^{N} [\, y(k,\mathbf{P}) - x(k) \,] \frac{\partial y(k,\mathbf{P})}{\partial p(j)} = 0 \quad \text{for } j=1,2..J \qquad [8.2]$$

Now J linear equations in as many unknowns can be solved by matrix methods, but eqns [8.2] contain nonlinear terms of the form $y\,\partial y/\partial p$. The equations may be linearized locally, about a parameter vector $\mathbf{P}_0$, through a simple perturbation scheme. Then retaining linear terms in a Taylor series about $\mathbf{P}_0$, gives $\mathbf{P} = \mathbf{P}_0 + \delta\mathbf{P}$. The model output y(k,**P**) can now be written as

$$y(k,\mathbf{P}) = y(k,\mathbf{P}_0) + \sum_{i=1}^{J} \frac{\partial y(k,\mathbf{P}_0)}{\partial p(i)} \delta p(i) = 0 \quad \text{for } k=1,2..N \qquad [8.3]$$

Substituting for y(k,**P**) in the condition [8.2] for minimum m.s.e., we obtain the following J equations for each j=1,2,..J

$$C(j) + \sum_{k=1}^{N} \sum_{i=1}^{J} \frac{\partial y(k,\mathbf{P}_0)}{\partial p(j)} \frac{\partial y(k,\mathbf{P}_0)}{\partial p(i)} \delta p(i) = 0 \quad \text{where } j=1.2..J \qquad [8.4]$$

where the J constant terms C(j) are given by

$$C(j) = \sum_{k=1}^{N} [\, y(k,\mathbf{P}_0) - x(k) \,] \frac{\partial y(k,\mathbf{P}_0)}{\partial p(j)} \quad \text{where } j=1.2..J \qquad [8.5]$$

Eqn [8.4] can be more effectively written in the matrix form

$$C + D\,\delta P = 0 \qquad [8.6]$$

$$\text{or } c(j) = \sum_{j=1}^{K} d(i,j)\,\delta p(i) \text{ where } j=1,2..J$$

Here C is a [Jx1] matrix defined in eqn [8.5], D is a [JxJ] matrix denoting the product of derivatives of model output y in eqn [8.4], and $\delta P$ is a [Jx1] matrix which denotes the desired change in $\mathbf{P}$ about $\mathbf{P}_0$ to locally minimize the m.s.e. This equation can be inverted to yield,

$$\delta P = -D^{-1}C \qquad [8.7]$$

where $D^{-1}$ is the inverse of the [JxJ] matrix D evaluated through any of the conventional numerical methods [see Press et al, 1986], since D does not have any special properties.

This gives the perturbation $\delta\mathbf{P}$ about $\mathbf{P}_0$ to minimize the m.s.e. We are now at a new value of $\mathbf{P}_0$ and the process can be iterated to find a parameter vector which either stabilizes the m.s.e. near a local bottom of the error surface, or brings it below an acceptable threshold corresponding to a 'good' estimate of parameter vector. It should be emphasized that the above scheme does not warrant a solution, though it often gives one for a reasonable initial guess $\mathbf{P}_0$, and it is extremely computation intensive.

In the m.s.e. spectral parameter scheme implemented for the 430 MHz Arecibo radar by Sato and Woodman (1982), the observation vector is the DFT of the time averaged periodogram sequence. The model output vector is then in the form of a distorted ACF sequence. In the model, MST radar signals s(t) have one or two Doppler shifted components, each with three ACF or PSD parameters for an assumed Gaussian shape in the PSD. Fading ground clutter c(t) also has three similar parameters. But due to its narrow, symmetric, and possibly unknown shape in the PSD, it is overspecified by the coefficients of a third order polynomial in $(\tau)^2$ and a small Doppler shift. With a noise platform included, the parameter vector has a length of 7(10) for 1(2) Doppler peaks. The distortion of ACF and PSD has been outlined in

Sec.6. The m.s.e. search is set about an initial guess of parameters obtained either by an ad-hoc analysis of spectra, or using the ACF method discussed in the next section. The m.s.e. implementation can routinely detect signals up to 50 dB below ground clutter, with a typical radial velocity uncertainty of 0.1-0.2 m/s.

The ad-hoc analysis, instead of estimating the parameters of ground clutter, merely removes it on the basis of its approximate symmetry in PSD estimates about zero Doppler shift. Estimates of Doppler shift and other parameters can be considerably improved by using time and range continuity of measured velocity, statistical editing of spectra, and by a statistical analysis of all the available data in several passes (Rastogi, 1984). These steps can be used to set a narrow range of parameters **P** for the m.s.e. method. Adaptive processing of spectral records using the available prior statistical information, e.g. tracking Doppler peaks in range, searching for parameters near a median Doppler-shift profile, and even using 'future' data, may speed up spectral-moment processing.

## 9. Spectral Moment Estimation via Correlation Function

Consider a complex wide sense stationary process $z(t)$ with power P, PSD $S(f)$ and ACF $R(\tau)$. For simplicity z is omitted as a suffix. In as much as $S(f)/P$ has all the properties of a probability density function, and $S(f) = \mathfrak{F}\{R(\tau)\}$, the non-central moments of $S(f)$ and parameters derived from these are simply related to the successive derivatives of $R(\tau)$ at $\tau=0$. This method was originally used at Jicamarca for measuring the vertical motions in the F-region using the incoherent-scatter radar technique and later applied to the first middle-atmospheric radar experiments by Woodman and Guillén (1974). A complete statistical analysis of this approach has been independently given by Miller and Rochwarger (1972).

Details can be seen by considering $R(\tau) = \mathfrak{F}^{-1}\{R(\tau)\}$ as in eqn. [2.4]. Using the series expansion of $\exp(\iota 2\pi f\tau)$ and evaluating the successive derivatives of $R(\tau)$ at $\tau=0$, we have

$$R(0) = \int_{-\infty}^{\infty} S(f)\, df = s^{(0)} \qquad [9.1]$$

$$R'(0) = (\imath\, 2\pi) \int_{-\infty}^{\infty} f\, S(f)\, df = (\imath\, 2\pi)\, s^{(1)} \qquad [9.2]$$

$$R''(0) = (\imath\, 2\pi)^2 \int_{-\infty}^{\infty} f^2\, S(f)\, df = (\imath\, 2\pi)^2\, s^{(2)} \qquad [9.3]$$

We find that these derivatives are related to the successive spectral moments, $s^{(0)}$, $s^{(1)}$ and $s^{(2)}$. $s^{(0)}$ is merely the signal power P. The other two spectral parameters of interest are the center frequency or the Doppler shift $f_c$, and the spread $\sigma_f$ of the PSD about it. As outlined in Sec. 2, these are related to the central moments of the PSD. In terms of the noncentral moments $s^{(1)}$ and $s^{(2)}$,

$$f_c = s^{(1)}/P \qquad [9.4]$$

$$\sigma_f^2 = s^{(2)}/P - f_c^2 \qquad [9.5]$$

which shows that uncertainties in a lower-order moment effects all higher-order parameters.

An interesting case arises when the Doppler-shifted component in the PSD is expressible through a simple shape such as the Gaussian. In terms of a normalized Gaussian function $N(f_c, \sigma_f^2)$ with mean $f_c$ and variance $\sigma_f^2$, the PSD becomes $S(f) = PN(f_c, \sigma_f^2)$. The ACF $R(\tau)$ is generally complex with a Hermitian symmetry. Its real part and magnitude are even, and the imaginary part and phase are odd functions of the lag $\tau$. For the Gaussian PSD,

$$R(\tau) = P \exp(\imath\, 2\pi f_c \tau) \exp\{-\tfrac{1}{2}(2\pi)^2 \tau^2 \sigma_f^2\} \qquad [9.6]$$

Comparing it with the polar form $|R(\tau)| \exp\{\imath\, \phi(\tau)\}$ of the ACF we see that the phase $\phi(\tau)$ increases linearly with lag $\tau$ and the mean frequency $f_c$. The magnitude $|R(\tau)|$ has a Gaussian shape which can be approximated by a

parabola for small $\tau$. From just two ACF values at zero lag and a small lag $\tau$, we find $P = R(0)$, and

$$(2\pi)\, f_c = \phi(\tau) / \tau \qquad [9.7]$$

$$(2\pi)^2\, \sigma_f^2 = 2\, \tau^{-2} \{ 1 - | R(\tau) | / P \} \qquad [9.8]$$

Fig. 9.1 shows how the spectral parameters are related for the ACF and PSD. The effect of two Gaussian components in signals scattered from two turbulent layers has been considered by Rastogi and Bowhill (1976).

The ACF approach provides a clever method for finding spectral parameters if z(t) contains only an atmospheric component s(t) conforming to the simple models just discussed. Otherwise spectral contributions to z(t) from noise n(t), ground clutter c(t) and interference i(t), are all included, by definition, in the ACF $R(\tau)$. We now use an appropriate suffix to identify these components. Corrections to remove their effect require ACF measurements at several lags.

An additive white noise n(t), merely adds a spike of size $P_n$ to $R_z(0)$ at zero lag. Then $P_s=P_z-P_n$. A correction for $P_n$ can be applied by using two or more small non-zero lags of $R(\tau)$ to estimate and remove the noise spike $R_z(0)$. Ground clutter c(t) has an effect on the estimation of $f_c$ only through the error it introduces in the power estimate. It contributes a nearly constant platform $R_c$ to $R_z(\tau)$ at small lags due to its long fading time. Its contribution may be effectively removed by d.c. subtraction from z(t) [See Fig. 9.2].

Statistical errors in parameter estimates obtained by the ACF method are discussed in detail by Miller and Rochwarger (1972). The following analysis of the uncertainty in Doppler estimation is, however, quite instructive. Consider K samples of a complex, zero-mean Gaussian process $z(t) = x(t) + j\, y(t)$ with a sampling interval T. If the variance of z(t) is $\sigma^2$, identified also as its power P, then the signal power $P_K$ estimated from K samples as $\langle z[k] z^*[k] \rangle$ has the statistics $E[P_K] = \sigma^2$ and $\mathrm{var}[P_K] = \sigma^4/K$ . Hence $P_K$ is unbiased and its statistical error $P/\sqrt{K}$ decreases with large K. Next we estimate $R(T)=R[1]$ at the first sampled lag index as $\langle z[k] z^*[k+1] \rangle$ using

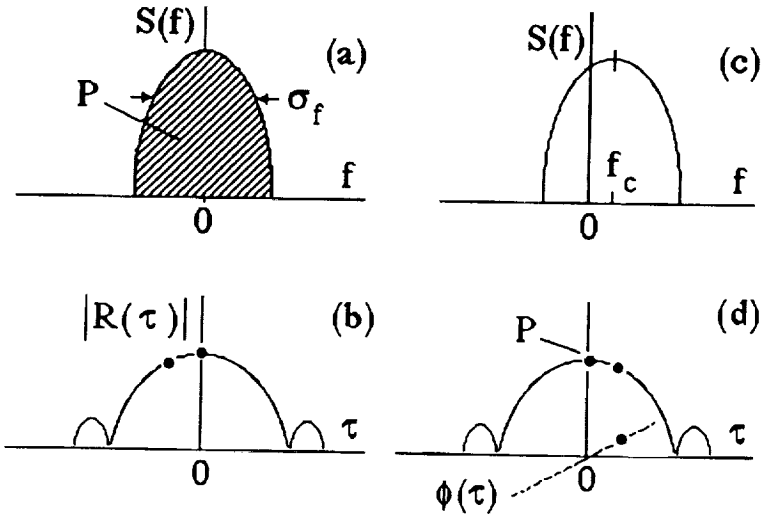


FIGURE 9.1 : A hypothetical PSD and the corresponding ACF for zero Doppler shift are shown in (a) and (b). The effect of a slight Doppler shift is shown in (c) and (d). The area under the PSD and the ACF at zero lag are equal to the signal power. The frequency width of the PSD and the relative value of ACF magnitude at a small lag are related. When the PSD is Doppler shifted by a small amount, the ACF becomes complex. Then the shift can be estimated from the ACF phase at a small lag.

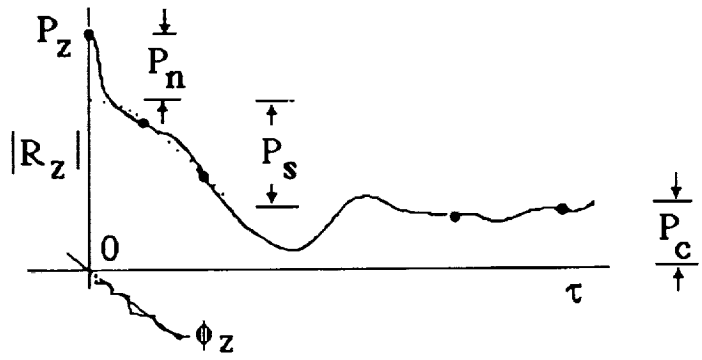


FIGURE 9.2 : The effect of noise and clutter on the shape of the ACF and their conritibution to the total power. The effect of noise and clutter can be effectively removed from the total power using the ACF values measured at a few key points. The ACF phase still remains linear at small lags, but the Doppler shift is underestimated unless noise and clutter are removed from the total signal power. The spectral width is overestimated from the ACF value at a small lag, unless the noise spike at zero lag is removed.

either of the two estimates $R^{[1]}$ or $R^{[2]}$ given in eqns [5.1] and [5.2]. The biased estimate $R^{[2]}$ is preferable for reasons discussed earlier, but the use of $R^{[1]}$ is more convenient. For small T, the K sample estimate $\phi_K$ of $\phi$ is

$$\phi_K(T) \approx \tan \phi_K[1] = \{[K-1]\, P\}^{-1} \sum_{k=1}^{K-1} y[k]\, x[k+1] - x[k]\, y[k+1] \qquad [9.9]$$

This estimate is unbiased due to the use of $R^{[1]}$. Its variance involves a moment of the form E[abcd] of four zero-mean Gaussian variables. Using a result due to Isserlis and Hotelling (see e.g. Papoulis, 1983) the fourth moment reduces to E[abcd] = E[ab] E[cd] + E[ac] E[bd] + E[ad] E[bc] . The final result ,

$$\mathrm{var}\{\phi_K(T)\} \approx \{ P^2 - |R(T)|^2 \} / 2K\ |R(T)|^2 \qquad [9.10]$$

shows that at small lags the uncertainty in phase estimate is quite sensitive to the relative magnitude of the ACF. The corresponding statistical error in the radial velocity for a radar wavelength $\lambda$ in terms of the normalized autocorrelation magnitude $\rho$ or $|R(T)/R(0)|$ is

$$\sigma_v \approx \frac{1}{\sqrt{2K}} \frac{\lambda_0}{4\pi T} \frac{\sqrt{1-\rho^2}}{\rho} \qquad [9.11]$$

For a 50 MHz radar, with T=0.25 sec, $\rho$=0.5, and K=100, we find that the radial velocity can be measured with a standard deviation of 0.23 m/s. With $\rho$=0.8 the figure improves to 0.1 m/s.

The ACF method provides a relatively fast means of estimating the spectral moments for clean radar signals. Due to the ease of its implementation, it is suited to real time estimation of spectral moments. Statistical averages of these moments may also serve as an initial guess in the m.m.s.e. approach.

## 10. Spectral Analysis by Time Series Models and Maximum Entropy Method

Methods discussed so far for estimating the PSD S(f) of a complex random process z(t), from its uniformly-spaced samples z[k], make some unrealistic

assumptions about extension of data or its ACF $R(\tau)$. The DFT assumes a periodic extension of data. In methods that use the ACF, windowing or truncation assumes zero correlation beyond a convenient maximum lag. J. P. Burg has proposed a method which circumvents these objections by seeking an extension of the ACF at measured lags that maximizes the entropy (in an Information-theoretic sense) of the observed process [see Childers, 1978]. Alternatively, one seeks to extend the process or its ACF from limited observations, using suitable time-series models. These are examined first.

Spectral analysis may be regarded as a filter design problem in which we seek coefficients h[k] of a feedback filter excited by white noise n[k], so that its output becomes the observed process z[k]. The filter output is taken as a linear combinations of the current input , q past inputs, and p past outputs. Such parametric representation of an observed process is called an autoregressive moving-average or ARMA model

$$z[k] = -\sum_{i=1}^{p} a[i]\, z[k-i] \quad + \quad \sum_{j=0}^{q} b[j]\, n[k-j] \quad \text{ARMA(p,q) model} \quad [10.1]$$

Recalling that shifting a signal s to the left by an interval iT amounts to multiplying its Fourier transform by $\exp(-\iota 2\pi i f T)$, the PSD S(f) can be represented in terms of two polynomials (with b[0]=1),

$$A(f) = 1 + \sum_{i=1}^{p} a[i] \exp(-\iota 2\pi i f T)] \quad \text{and} \quad B(f) = 1 + \sum_{j=1}^{q} b[j] \exp(-\iota 2\pi j f T)]$$

and using the sampling interval T and noise variance $\sigma^2$, as

$$S(f) = \sigma^2 T \frac{|B(f)|^2}{|A(f)|^2} \qquad [10.2]$$

This representation has q zeros and p poles. Hence we expect the AR model to be more suitable for representing a process with sharp peaks in the PSD, and the MA process for a PSD with flat peaks. The ground clutter component c(t) in radar experiments has a near-ideal representation as a pole. We surmise

that the Doppler shifted components should require an MA part. An ARMA(p,q) process can be overdefined in terms of an AR(p') or an MA(q') process with p'>>p, q'>>q. So a purely AR model, with q=0, may be adequate for representing PSD of radar signals z(t).

For an AR(p) process z(t), the ACF R[k] is related for lags 0,1,..p through the Yule-Walker normal equations.

$$\begin{bmatrix} R[0] & R[-1] & & R[-p] \\ R[1] & R[0] & & R[-p+1] \\ & R[1] & R[-1] & \\ & & R[0] & R[-1] \\ R[p] & R[p-1] & R[1] & R[0] \end{bmatrix} \begin{bmatrix} 1 \\ a[1] \\ a[2] \\ \\ a[p] \end{bmatrix} = \begin{bmatrix} \sigma^2 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \qquad [10.3]$$

These linear equations involve the (p+1) ACF values arranged as a Toeplitz matrix. In this matrix form, the same elements appear along a diagonal. In addition, the elements along cross diagonals have Hermitian symmetry. The matrix can be inverted through Levinson's recursion in ~$p^2$ operations. Programs for solving these equations may be found in Press et al. (1986) and Marple (1987). Note that the use of Wiener-Khintchine theorem to find the PSD S(f) would require the ACF values R[m] at all lags. But for an AR(p) process, the p coefficients suffice through eqn [10.2] for finding the PSD. The structure of these equations may also be discussed in terms of forward and backward linear-prediction filters, which given some values of data z[k] extend these in the both directions. Further discussion may be found in several excellent papers in Childers(1978), and Marple (1987).

The modified Yule-Walker equations for MA and ARMA models are nonlinear and inherently difficult to solve for filter coefficients.

Entropy H of a random variable X with a probability density function $f_x$ is defined as the expectation $E\{-\ln f_x(x)\}$. It is a measure of the randomness in the underlying chance experiment. Maximizing the entropy may yield *a* solution in some statistical situations. An interesting example is that of a

loaded die with an average face value of 3.5, instead of 4.5 for a fair one. There are infinitely many solutions to the probabilities $p_i$ for the six faces. Maximizing the entropy H under the constraint of the given average value can be set up as a nice variational problem. Using the method of Lagrange multipliers, this gives $p_i$'s as a geometric series with a ratio r. The resulting equations for $p_1$ and r are nonlinear, but can be solved recursively from an initial guess. The solution is $p_1 = 0.05435$, $r = 1.44926$. This is not a unique solution, since changing any two $p_i$'s by a small amount $\pm\delta$ is also a solution.

For (N+1) uniformly-spaced samples of a complex, zero-mean, Gaussian random process, the entropy H is obtained using the joint probability density function of of 2(N+1) real Gaussian variables. This density involves the Toeplitz ACF matrix form given in eqn. [10.3], albeit of size (N+1) instead of (p+1). We denote this matrix by $\mathbf{R}_N$ as it involves N distinct nonzero lags. It is also convenient to use the base $(2\pi)^{1/4}$ for the logarithm. Then the entropy H becomes $0.5 \log\{\det \mathbf{R}_N\}$ and it increases with N, eventually becoming infinite. We deal with the entropy rate h defined as $h = H/(N+1)$ which becomes $0.5 \log\{(\det \mathbf{R}_N)^{1/(N+1)}\}$. In the limiting case of infinite N, it can be shown from that for the Toeplitz form of $\mathbf{R}_N$, the entropy rate h reduces to

$$h = -0.5 \log T + 0.5T \int_{-0.5/T}^{0.5/T} \log S(f)\, df \qquad [10.4]$$

where the integral is over the Nyquist interval. Complicated details leading to this result may be found e.g. in Smylie et al.(1973). We may expand S(f) in a Fourier series using the ACF values R[k]. The entropy rate h may now be maximized with respect to the unknown ACF values R[k] for $|k|>N$ under the constraints that the first (N+1) values of ACF, including the zero lag, are known from the data. This difficult exercise, as in the loaded-die problem, does not warrant a unique solution. The final result expressed in the form of a PSD estimate is that

$$S(f) = \frac{T\sigma^2}{\left| 1 + \sum_{k=1}^{p} a[k] \exp(-\iota\, 2\pi\, fkT) \right|^2} \qquad [10.5]$$

This result is exactly the same as the PSD of an AR(p) process given in eqn [10.2] with the moving-average order q set to zero and the numerator polynomial $|B(f)|^2 = 1$. The p coefficients a[k] are the same as for the AR(p) model obtained by solving the Yule-Walker equations [10.3]. Hence the maximum entropy method (MEM) is equivalent to the AR(p) model for equispaced samples of a complex Gaussian process.

If the process is not Gaussian, then the final result in eqn [10.5] for entropy rate would not hold. MEM still will give a result, but it may not be a representative estimate of the PSD for the process. We also remark that though we have shown eqn [10.5] in the AR form, actual implementations of MEM are rather different and take the form of designing a linear-prediction filter. Computer programs for MEM may be found in Press et al. (1986) and in Marple (1987).

A fundamental problem in implementing the above methods is that of finding the order p of the process. Use of an incorrect order give larger statistical errors. The order must be found empirically for each class of processes. In a recent experimental and numerical study, Klostermeyer (1986) has compared the performance of periodogram, MEM and maximum likelihood method (MLM) for PSD estimation of ST signals observed with the 53.5 MHz SOUSY radar. It was found that for SNR of 0.3 to 10, MEM and MLM give better estimates of Doppler shift. The optimum order of the MEM filter is ~3±1 with a sampling time of 0.173 sec, and appears to decrease with the SNR. Similar studies with other atmospheric radar signals, and of their statistics, are needed for developing the use of MEM and AR PSD models.

## References

Alexander, S. T., *Adaptive signal processing - Theory and applications*, Springer-Verlag, New York, 197pp., 1986.

Balsley, B. B., The MST technique - a brief review, *J. Atmos. Terres. Phys.*, **43**, 495-509, 1981.

Balsley, B. B. and D. A. Carter, The spectrum of atmospheric velocity fluctuations at 8 km and 86 km, *Geophys. Res. Lett.*, **9**, 465-468, 1982.

Bard, Y., *Nonlinear parameter estimation*, Academic, NewYork, 1974.

Blackman, R. B. and J. W. Tukey, *The measurement of power spectra from the point of view of communications engineering*, Dover, New York, 190pp., 1958.

Brigham, E. O., *The fast Fourier transform and its applications*, Prentice-Hall, Englewood Cliffs, 448pp., 1988.

Childers, D. G.,Ed., *Modern Spectral Analysis*, IEEE Press, New York, 334pp., 1978

Cooley, J. W., P. A. W. Lewis, and P. D. Welch, *The fast Fourier transform and its application to time series analysis*, in *Statistical methods for digital computers*, Eds. K. Enslein et al., J. Wiley, New York, 377-423, 1977.

Davenport, W. B. and W. L. Root, *An introduction to the theory of random signals and noise*, McGraw-Hill, New York, 393pp., 1958.

Gottman, J. M., *Time-series analysis : A comprehensive introduction for social scientists*, Cambridge, New York, 400 pp., 1981.

Hamming, R. W., *Numerical methods for scientists and engineers*, 2nd edition, Dover, New York, 721pp., 1973.

Harris, F. J., On the use of windows for harmonic analysis with the discrete Fourier transform, *Proc. IEEE*, **66**, p.51-83, 1978.

Klostermeyer, J., Experiments with maximum entropy and maximum likelihood spectra of VHF radar signals, *Radio Sci.*, **21**, 731-736, 1986.

Koopmans, L. H., *The spectral analysis of time series*, Academic, New York, 366pp., 1974.

Marple, S. L., *Digital spectral analysis with applications*, Prentice-Hall, Englewood Cliffs, 492pp., 1987.

Miller, K. S., *Complex stochastic processes*, Addison-Wesley, Reading, Mass, 238pp., 1974.

Miller, K. S., and M. M. Rochwarger, A covarinace approach to spectral moment estimation, *IEEE trans. Inf. Th.*, **IT-18**, 588-596, 1972.

Nuttal, A. H., Some windows with very good sidelobe behavior, *IEEE trans. Audio, Speech and Sig. Proc.*, **ASSP-29**, p.84-91, 1981.

Oppenheim, A. V. and R. W. Schafer, *Digital Signal Processing*, Prentice-Hall, Englewood Cliffs, 585pp., 1975.

Oppenheim, A. V. and A. S. Willsky, *Signals and systems*, Prentice-Hall, Englewood Cliffs, 796pp., 1983.

Papoulis, A., *Signal analysis*, McGraw-Hill, NewYork, 431pp., 1977.

Papoulis, A., *Probability random variables and stochastic processes*, 2nd edition, McGraw-Hill, NewYork, 576pp., 1983.

Press, W. H., B. P. Flannery, S. A. Teukolsky and W. H. Vetterling, *Numerical recipes : The art of scientific computing*, Cambridge, New York, 818pp., 1986.

Rabiner, L. R., C. A. McGonegal and D. Paul, FIR windowed filter design program - WINDOW, in *Programs for Digital Signal Processing*, IEEE Press, New York, p. 5.2.1-5.2.19, 1979.

Rastogi, P. K., A note on the use of coherent integration in periodogram analysis of MST radar signals, *Handbook for MAP*, v. **9**, Ed. S. A. Bowhill and B. Edwards, SCOSTEP secretariat, Urbana, p. 509-512, 1983.

Rastogi, P. K., Criteria and algorithms for spectrum parametrization of MST radar signals, *Handbook for MAP*, v. **14**, Ed. S. A. Bowhill and B. Edwards, SCOSTEP secretariat, Urbana, p. 289-293, 1984.

Rastogi, P. K. and S. A. Bowhill, Scattering of radio waves from the mesosphere 2. Evidence for intermittent mesospheric turbulence, *J. Atmos. Terres. Phys.*, **38**, 449-462, 1976.

Rastogi, P. K. and R. F. Woodman, Mesospheric studies using the Jicamarca incoherent-scatter radar, *J. Atmos. Terres. Phys.*, **36**, 1217-1231, 1974.

Röttger, J., VHF radar measurements of small-scale and meso-scale dynamical processes in the middle atmosphere, *Phil. Trans. Royal Soc. London*, **A 323**, 611-628, 1987.

Sato, T. and R. F. Woodman, Spectral parameter estimation of CAT radar echoes in the presence of fading clutter, *Radio Sci.*, **17**, p. 817-826, 1982.

Smiley, D. E., G. K. C. Clarke, and T. J. Ulrych, Analysis of irregularities in the Earth's rotation, in *Methods in Computational Physics*, **v. 13**, eds. B. Alder et al, Academic, New York, p. 391-430, 1973.

Stearns, S. D., *Digital signal analysis*, Hayden, Rochell Park, 280pp., 1975

Vincent, R. A. and I, M. Reid, HF Doppler measurements of mesospheric gravity wave momentum fluxes, *J. Atmos. Sci.*, **40**, 1321-1333, 1983.

Whalen, A. D., *Detection of Signals in noise*, Academic Press, San Diego, 411pp, 1971.

Welch, P. D., The use of fast Fourier transform for the estimation of power spectra : A method based on time averaging over short, modified periodograms, *IEEE trans. Audio, Speech and Electroac.*, **AU-15**, p.70-73, 1967.

Widrow, B. and S. D. Stearns, *Adaptive signal processing*, Prentice-Hall, Englewood Cliffs, 474pp., 1985.

Woodman, R. F. and A. Guillén, Radar observations of winds and turbulence in the stratosphere and mesosphere, *J. Atmos. Sci.*, **31**, 493-505, 1974.