SPECTRUM ESTIMATION AND SYSTEM IDENTIFICATION RELYING ON A FOURIER TRANSFORM AND SUPPLEMENT

BY

DAVID R. BRILLINGER[1] AND JOHN W. TUKEY[2]

TECHNICAL REPORT NO. 5
MARCH 1982

RESEARCH SUPPORTED IN PART BY
[1]NATIONAL SCIENCE FOUNDATION GRANT PFR-7901642
AND [2]AROD DAAG29-79-C-0205

DEPARTMENT OF STATISTICS
UNIVERSITY OF CALIFORNIA, BERKELEY

SPECTRUM ANALYSIS IN THE PRESENCE OF NOISE: SOME ISSUES AND EXAMPLES

David R. Brillinger

Statistics Department
University of California
Berkeley, California

John W. Tukey

Statistics Department
Bell Laboratories and Princeton University
Murray Hill, New Jersey and Princeton, New Jersey

ABSTRACT. The Fourier transforms of stretches of data are fundamental quantities in the estimation of spectra and the identification of systems. This paper describes and discusses the use of empirical Fourier transforms, in a critical fashion. Philosophy, techniques, formulas, experience, enlightening examples are all presented. The focus is on experimental design, scientific discovery and interpretation rather than signal processing. The presentation and notation attempt to unify areas such as types of process, parameters of varying orders, spectrum and system estimation. It is an overview rather than a detailed presentation, emphasizing philosophy and some of the success stories of time series analysis.

Keywords: System identification, time series, noise, signal spectra, Fourier transform, inference, stationarity, cepstrum, point process, bispectrum, spatial process, missing values, robust/resistant.

CONTENTS

INTRODUCTION

The authors of this paper learned separately, early on in their work, that a rigid "assumptions, then mathematics, then conclusions" approach to *techniques for the analysis of data* was neither safe nor effective. Nor was it the way that the discoveries of science seem to have been made. What seems the effective approach is being flexible at the beginning and then stepping into, *if and when appropriate*, some hypothetic-deductive structure (perhaps involving a few-parameter description).

This seems especially true of many -- probably most -- applications of spectrum analysis. The initial investigation needs to be flexible, considering fairly general patterns for the spectrum, after which we *may* -- or may not -- be able to go over to some popular few-parameter description of our spectrum. One of the early successes of modern spectrum analysis some 35 years ago, was the re-analysis of autocorrelations of a tracking radar, published, with an elementary-theory-based fit, in James *et al.* (1947). When re-analyzed flexibly in spectrum terms, this data showed (a) the low-frequency hump corresponding to the simple theory, which was well-enough fitted by the few constants, and (b) a smaller but quite definite peak near 2 Hertz, which the few-constant analysis had entirely overlooked. There have been many parallel instances in the intervening decades. As a result of this experience we have not written this review paper, particularly in its opening sections, in an assumptions-mathematics (possibly only sketched)-conclusions style.

As analysts of data we are often concerned with phenomena, with qualitative aspects of what the data shows, as well as with its quantitative aspects. This has helped us to be concerned with the phenomena of our techniques, with their qualitative and semiqualitative characteristics, which usually generalize much further than the narrowly quantitative ones. These phenomena are often both vague and important, so we will often be deliberately vague.

Many man-made signals, and a few environmental ones, are relatively non-statistical. When a single subject says the same vowel, again and again, the waveforms can be remarkably similar. The same can be true of successive pulses from a single radar, or for the traces seen at the same distant point due to earthquakes occurring on the same fault, years apart. These are the nice cases for the analyst. IF he KNOWS this is going on, he or she can apply special methods of spectrum analysis, which often do boil what is fitted down to a few parameters. This does not mean, however, that these techniques are either generally safe or suitable for use in a new field.

All such techniques have to include strong hidden assumptions, because they focus on the one particular realization before us. If that realization has been sampled, as today it will so often have been, so that it consists of T values, $x(1),...,x(T)$ and if we choose $m > T/2$ distinct frequencies, $\omega_1, \ldots, \omega_m$, no two of which are aliases (section 6a) of each other, then we can always represent the given data as

$$X(t) = \sum_j c_j \cos(\omega_j t + \psi_j)$$

as a linear combination of just these frequency components. This can be done, in particular, for all m ω's near zero, or all near $\pi/\sqrt{3}$, or all near $.98\pi$. If we look at individual realizations, one by one, and do not add (often tacitly) restrictive assumptions, the whole notion of a spectrum is in trouble. So we should not be surprised that special methods involve tacit assumptions -- assumptions that often do not hold.

The other extreme, where most of your authors' clients have found their data, is one in which we have one -- or perhaps a few -- realizations, but where our interest is in the ensemble or process from which the realization(s) come. In studying the WT4 wave guide (section 5c), our interest is not in the particular 12 kilometers laid for a test, but in understanding what irregularities are likely to occur, if and when such guide is manufactured and laid for regular service. In studying the background noise in a specific environment (perhaps with a view to the design of a special telephone set), our interest is not in the few minutes of record that we sampled, digitized, and analyzed, but in the hours to years of background noise yet to come. In looking at the noise characteristics of alternative frequency bands, when designing a radio communication system, our concern is not with the past we have analyzed, but with the future we must face.

The best paradigm we have for looking ahead in such "noise-like" situation is a realization -- ensemble one, a generalization to time-functions of the sample-population paradigm. As always, this paradigm may underestimate our uncertainty, because, indeed, "the future may be different". But, in "noise-like" cases, it rarely overestimates our uncertainty.

It is important to separate such "noise-like" situations -- optimistically modelled with Gaussian (or for point-processes, Markov) ensembles -- from the "signal-like" ones (e.g. FDR's vowels, the pulse from radar #115, the seismic return from station 7 when a one-pound charge is detonated at station 0). Any approach to spectrum estimation works for "signal-like" ensembles, but only care and trepidation can give us useful answers for "noise-like" ones.

This paper is a mixture of philosophy, techniques, formulas, experience, and enlightning examples (with a bias to those of discovery of phenomena). The approach and notation seek to unify areas: i) ordinary time series, point processes and spatial series and ii) second and higher order spectra, for example. The reader need not feel he must go straight through the paper, but should examine the Table of Contents for topics of special interest to himself. It may not be completely apparent from the specifics of the paper; however it is our deepest experience that the practitioner of spectrum analysis cannot know too much of the physical background of the situation of concern. Blind use of formulas and computer packages leads many astray. HU "* The spectrum *" Conceptually, the spectrum of a time function -- whether discrete time or continuous time -- would be defined by taking a very long stretch of the time function (much longer than we actually have), filtering it with a sharp, narrow-band filter, looking at a piece of the result (as long, perhaps, as our actual record), and asking how much energy (power, variance) is present in this segment. At least two extreme cases must be clearly understood:

Case 1. The very long stretch of the time function (a) is the only time function we want to consider and (b) consists of a superposition (a summation) of not too many sinusoidal terms of substantially different frequencies. Here the narrow-band filter will select only one frequency at a trial, and the segment we are to assess will be a perfect sinusoid. This we have a good chance of doing with high precision.

Case 2. The time function is to be regarded as a realization of a Gaussian process (which for convenience we could, but need not, assume to be ergodic). It is one of many possible time functions; our analyses are directed toward the properties of the ensemble of these functions, not toward those of a specific realization. The narrow-band filter will produce a narrow-band Gaussian noise, most likely -- especially if we are fortunate -- with a more detailed spectrum matching reasonably closely the result of passing white Gaussian noise through the given filter. In particular, the amplitude of the result will change little over time intervals short compared to the reciprocal of the filter width, and will be almost uncorrelated over time intervals long compared to the reciprocal of the filter width.

If we look at a stretch of such a filtered record that provides only a few, say 5 to 50, portions whose length is the reciprocal of the filter width, we will have only a poor idea of how variable -- on the average -- a much longer narrow-band record would be -- how much power such a narrow-band record would contain. That is to say, how large the spectrum of the Gaussian process was at (and near) the frequency of our filter. And what we can do when allowed to narrow-band filter a very long record (and snip out a piece as long as our data) is at least somewhat better than we can possibly do when we have only the actual data.

* *SPECTRAL PRECISION FOR GAUSSIAN DATA* *

This lower bound on variability is easily evaluated. We know from the Sampling Theorem (Nyquist 1928), also known as the Cardinal Theorem of Interpolatory Theory (Whittaker 1935), that if we knew the whole of a narrow-band signal at equispaced points, separated by intervals of $1/2\Delta f$, where Δf is the bandwidth, we could recover the whole narrow-band signal by interpolation.

Except for "leakage of information" across the ends of our longer narrow band record (in one direction or the other), which is ordinarily negligible, then, the information in our narrow band signal is contained in these discrete points, of which there are, (within $\pm$ 1), $k = T/(1/2\Delta f)$ $= 2T\ \Delta f$ in a record of length T.

These k points will, if we started with a realization of a Gaussian process, have a joint Gaussian distribution. They will tell us as much as possible about the power or variance associated with the narrow-band Gaussian process when they are independent. If we look at the sum of their squares, essentially the best we can do in this case to estimate the power in this narrow band, its average will be $k\sigma^2$ and its variance will be $2k\sigma^4$ (as a multiple of chi-square on k degrees of freedom). The ratio of standard deviation to mean will be $(2/k)^{1/2}$, which falls to 1/12 for k about 300.

Thus, the power in our narrow band is to be known to *one significant figure* (conveniently taken as standard deviation/average = 1/12) we require $k \geq 300$ and hence

$$T \geq 300/2\Delta f = 150/\Delta f$$

For narrowish bands, we rarely analyze anything like this much data. So we rarely know even the *average power*, over narrowish bands, of any *Gaussian* process to as much as *one* significant figure. When we study phenomena that are at all like Gaussian processes, we need to

1) work with estimates of average spectra over bandwidths that are as great as will be useful, and to

2) plan for substantial variability in these averages, usually affecting the *first* significant figure seriously.

Those who must be concerned with spectra of Gaussian processes -- or with spectra of even crude imitations of Gaussian processes -- have a very much more difficult task than those who deal with superpositions of well-, or even moderately-well separated sinusoids, particularly when only a quite short portion of the sinusoid of interest is available.

The methods described below are intended to be effective when we seek information about spectra of noise-plus-signal or pure-noise processes. They will work -- giving *estimates* of *averages across bands* -- for almost any sort of data. When we are so fortunate -- usually either because we deal with consequences of the motion of heavenly bodies (e.g. the tides) or of human activity (e.g. radar applied to man-made targets) -- as to have most of the energy at discrete, well-separated frequencies, other methods may well be more effective for routine use. However, as we now discuss, the methods described below are still likely to be needed.

* *WHAT CAN BE ESTIMATED?* *

If we have the sum of a few perfect sinusoids (of constant amplitude) embedded in a very low-level noise, we do not need a lot of data to recover the frequencies, phases, and amplitudes of the sinusoids. But if the noise might be roughly Gaussian, then the limitations just discussed apply at least as strongly to learning about the *noise*, even if it is at a high level. Without very long records we will only learn a little about the noise.

If, by contrast, the process is Gaussian, we can only learn less, since the limitations just discussed apply now to all frequency components. Unless we are in fact correct, when we set about "modelling" the real situation with some description in terms of a small finite number of constants, we will be unable to do anything appreciably better than estimate *averages* of the spectrum over bands of well-chosen width -- neither narrow enough to make us really happy about our frequency resolution or broad enough to make us really happy about our sampling fluctuations.

These general statements are not confined to the approach to spectrum analysis we discuss in this paper. No procedure can do significantly better,

1) in holding down sampling fluctuations, or

2) in avoiding averaging over appropriate frequency ranges

provided that the data is a realization of a stationary Gaussian process, since, as is well known, the realization values of its mean and the quadratic functions of its observations are a set of sufficient statistics.

* *THE ROLE AND PLACE OF MODELLING* *

As statisticians, we are concerned with the real world from which our client's data come, with all its uncomfortable aspects. We can -- and do -- make use of models but in the spirit of guidance rather than trust. The theoretical study of oversimplified models, in some areas called "toys", can be of great help in suggesting what sorts of data analysis to try. It can be a source of equally great danger, if we forget how tenuous our initial assumptions were, and start to treat them as fact. In almost every case, any conveniently simple model has both (i) to be empirically tested to see in what ways it fails to be correct and by how much and (ii) to have the theoretical consequences of these failures at least crudely assessed. After all, no conveniently simple model is exactly correct. And we take our life in our hands if we trust it without good reason.

How are we to test empirically those aspects of our models that give a few-constant description of some spectrum? Only by analyzing real data by techniques that will work, in useful and understood ways, when the model is inadequate. However fortunate we may be in having few-constant models that give working-quality approximations to real situations, we can only know that they do this when we can test them. And the simplest adequate test is often the analysis of some spectra by methods of general application.

Whether or not they are able to design and operate as if some few-constant model were the truth, anyone deeply concerned with spectra who uses few constant models in spectrum analysis, we believe, needs to be able to check the adequacy of approximation these models give him or her. Part of this checking is likely to call for more generally applicable procedures of spectrum analysis. Often -- we would suspect almost always -- these will be the sorts of procedures we are about to discuss.

You may, if you like, regard our approach as *exploratory*. In doing so, however, you will need to think of exploration as *not only* what is done in the beginning *but also* what is done every so often, particularly to check whatever aspects of quality of approximation by the current model have been newly recognized as a critical importance.

LIST OF NOTATION

i	=	$\sqrt{-1}$
ω	=	frequency in radians per unit time
$\bar{x}$	=	complex conjugate of x
t	=	time
$X(t), Y(t)$	=	time series
ave X	=	average or expected value of the random variable X
$c_{XX}(u)$	=	covariance function at lag u of the stationary series $X(t)$
$c_{XY}(u)$	=	cross-covariance function of the stationary series $X(t)$ and $Y(t)$
$f_{XX}(\omega)$	=	$(2\pi)^{-1} \sum_{u=-\infty}^{\infty} c_{XX}(u) \exp\{-i\omega u\}$ = power spectrum at frequency ω of the series $X(t)$
$f_{XY}(\omega)$	=	$(2\pi)^{-1} \sum_{u=-\infty}^{\infty} c_{XY}(u) \exp\{-i\omega u\}$ = cross-spectrum of the series $X(t)$ and $Y(t)$
$R(\omega)$	=	$f_{XY}(\omega)/\sqrt{f_{XX}(\omega)f_{YY}(\omega)}$ = coherency of the series $X(t)$ and $Y(t)$
$\lvert R(\omega)\rvert^2$	=	coherence of the series $X(t)$ and $Y(t)$
$d_X^T(\omega)$	=	$\sum_{t=0}^{T-1} X(t) \exp\{-i\omega t\}$ = sample Fourier transform
$I_{XX}^T(\omega)$	=	$(2\pi T)^{-1} \lvert d_X^T(\omega)\rvert^2$ = periodogram of the sample
$I_{XY}^T(\omega)$	=	$(2\pi T)^{-1} d_X^T(\omega)\overline{d_Y^T(\omega)}$ = cross-periodogram of the sample
T	=	length of record (continuous time) number of data points (discrete time)

A. OPENING

1. VITAL ISSUES

A number of issues need to be understood in nearly every application of spectrum estimation and system identification. We begin by treating a few of the most important.

1a. NON-ENTANGLEMENT (FREQUENCY NONDISTORTION)

Frequencies gain their widespread importance because of what can be done to signals or noises without destroying the identities of the contents of frequency bands or mixing these contents together. Such things can be done (1) after the signal or noise is generated and before it is emitted, (2) between emission and reception, (3) in the reception process, and (4) in what we do after the signal or noise is safely received.

Our attempts, surprisingly often successful, to assess the frequency content of signals or noises depend, in practice, on the possibility of filtering out, directly or indirectly, portions of the signal or noise associated with frequency *bands* and then assessing how much power (energy per unit time, variance) is associated with each band. (In practice, however, the details of our computations may not make this process clear.) If filters could not extract certain frequency bands without too much disturbance from other bands, we could do much less to assess the frequency content of signals or noises. And in many other problems, where frequency techniques are used incidentally, we could not succeed as easily or as well.

Most signals or noises reach us through one or more media by propagation, be it electromagnetic, optical, acoustic, seismic, ocean-surface, etc. If the contents of frequency bands were not reasonably separated during the process of propagation, we would find it much harder to use received signals or noises to tell us about (1) transmitted signals or noises and (2) the characteristics of the propagation medium.

We formulate below, in Section 3, the conditions required for a "system" -- with one input and one output -- to be linear and time-invariant and hence to leave contents of frequency bands both unentangled and undisplaced. For examples, and further discussion of such systems and their generalizations, see Section 13a and b.

1b. KINDS OF TIME

We have, tacitly, assumed that time was either continuous or equispaced discrete. Situations where localized phenomena (earthquakes or lightning strokes), occur at unpredictable times can be treated by timing them in "continuous" time, or -- much more realistically -- in fine-grained equal-interval discrete time.

When, as is so often the case, the phenomenon is in continuous time, but our data is in equispaced discrete time, we face the problem of aliasing (see subsection 6a). The relationship between these two kinds of time is vital for all processes of analysis using digital -- or digitized -- data, which is rapidly becoming the most widespread and important case. We must always recognize, and distinguish, the kinds of time for the phenomenon, for data reception and recording, and for data analysis. Their differences can be vital.

1c. NOISE IS IMPORTANT

It is so much easier to avoid being realistic about noise. It is easiest to write equations as if there were no noise. Next after this comes attempts to cover noise's ills by adding on an $\epsilon(t)$ which is, more or less explicitly, assumed to: (1) have a flat spectrum, (2) follow a (joint) Gaussian distribution, (3) not be large enough to cause serious difficulties. Neither of these evasive approaches works for most meaningful problems.

In many important areas -- radio or radar background noise, ocean waves (tsunamis aside), seismic noise and the like -- our efforts at spectrum estimation are directed at the spectra of noise-like time functions. And the tools we have to identify systems are often noise-like inputs.

When noise does not dominate, it is still usually present to a meaningful degree. That means that *statistical* thinking is essential. Just because something is measured does not mean it is truth. Whether the noise is inserted at the source, in the analysis, in the receiver, or in the medium does not matter, when, as is so often the case, what we want to study is the noise-free source. We can rarely do this directly. As a consequence, we are almost always trying to look beneath our measurements or recordings, to make inferences down through some number of layers of noise. Even in dealing with the simplest human-created signals looking through noise is a common necessity. Those who study signals *carefully* must also be statisticians.

Detailed assumptions -- flat spectrum, Gaussian distribution, known total power -- about our corrupting noise apply no more frequently in studying signals than in other areas where statistical thinking can be used. The ideas of statistical mechanics usually apply when there are 10^{21} or more sources of comparable importance, naturally combined by addition using comparable weights. The unhappy state of affairs, in signal-chasing as well as in most other applications of statistical thinking or analysis, is the most common one, where there are 10^1 to 10^4 sources of comparable importance, -- enough so we cannot treat them source by source, too few to let the Central Limit Theorem convey us across the Styx to the Land of Gaussian Distributions.

Noise makes all of us unhappy; having to be statistical makes many unhappy; assumed *knowledge* of distributions is usually a broken crutch, guaranteed to leave us in trouble far too often.

2. THE VITAL DISTINCTION

No distinction in the whole area touched upon by this study is more vital than that among:

a) pure noise,

b) pure signal,

c) signal-plus-noise,

While (b) is ordinarily unrealistic, there are enough cases of (c) with nearly negligible noise that thinking about (b) is a helpful idealization -- if it is not allowed to be used where it should not.

The vital distinction is that between (a) and (b), and it is to this that the present section is devoted.

2a. NOISE-LIKE PROCESSES

The constancy of a noise-like process does not lie in the shapes of its individual time functions (its realizations). This constancy may well lie in the approximate frequency-content of each of its time functions, as would be the case for many instances of either white or colored Gaussian noise. It might lie in some other, more subtle characteristics, or it might be even grosser.

Consider, for instance, three different Gaussian noises, different in color (in frequency-content profile, in spectrum shape) as well as in size (in variance, in total power). A process which draws each entire realization from one of the three, drawing, say, from the first with probability 10%, from the second with probability 30%, and from the third with the remaining 60% probability, would be very noise-like. Different realizations are likely to appear quite different. We will have to draw and examine many realizations before we come to understand what is going on.

Many phenomena, as noted at the beginning of Section 1c, are noise-like. Yet, as phenomena, we need to treat them as "signals" to be captured. Their realizations may resemble one another more closely than in the example just given, but any two will be far from being identical.

We usually have a small sample -- all too often a sample of one -- for the sample size is the number of distinct realizations. Our need is to make as good inferences as we can to the population. Since we have a sample, our inferences are necessarily imprecise. We have to think statistically, and treat our conclusions -- like any statistical conclusions -- with both built-in grains of salt about the numbers and supplementary grains of salt about the wisdom of making the structural and stochastic (probabilistic) assumptions that reaching these conclusions really require. (See Section

17 for the distinction between conventional or optimum-generating assumptions and those we really need.)

It is for data near the noise-process end of the continuum that most of the sorts of Fourier methods of spectrum assessment discussed in this account were originally intended, and where they are most effective.

2b. SIGNAL-LIKE PROCESSES

At the other extreme come the signals recorded by geophones or other seismographs in the process of geophysical prospecting. It matters little whether the explosive charge was set off on Tuesday or Thursday, in one week or another, in the a.m. or the p.m., so long as the place of detonation is kept constant. The background noise is weak, the tracing will appear essentially the same, whenever the charge was fired.

This process is very "signal-like". Indeed, not only is there a single source, but the energy at the source is concentrated -- in this instance in time. Concentration in frequency also often happens for other kinds of single sources, as does concentration at or near a few times -- or a few frequencies -- or a few time-frequency combinations.

Studying pure signal-like processes may well be best done by quite different methods than those outlined in this paper. We direct the reader's attention to the other papers in this special issue.

But few processes are purely signal-like. Some corruption with noise of various kinds, including (but not limited to) measurement noise (and, usually, digitization or quantization noise) is the norm, not the exception. As statisticians, your authors: (1) are ever suspicious, as to whether noise was important in each specific instance when techniques are either derived from or demonstrated with signal-only cases, and (2) look forward to the development of techniques apt for mixtures of signal and noise, at least. As engineers, we think you should be, too.

The most basic point is simple: what is a good technique for pure signal-like processes is *almost sure* NOT to be a good technique for pure noise-like processes, and *vice versa*.

2c. NUMBER OF SOURCES

As we have noted, and will note again, from time to time, one of the main reasons why some processes are noise-like and others are signal-like is the number of elementary sources. Ocean waves are an easy-to-understand example. Both swell and local surf involve the combination of effects from many individual atmosphere-ocean interactions, spread over wide ranges of space and time. So many indeed, as to have the Central Limit Theorem ensure that the wave process is very, nearly Gaussian (so long as we stay in deep water, and avoid shallow water nonlinearities).

Tsunamis (often miscalled "tidal waves") also move on the ocean surface. They typically come from earthquakes, which can often be treated as single sources. As a process tsunamis are close to being purely signal-like. (If there were no swell or local surf in the background, tsunamis would have been very nearly signal-like indeed.)

Roughly, then, we have the following scale of source multiplicity and process characterization:

a single elementary source: very signal-like

a few elementary sources: reasonably signal-like

moderately many elementary sources: noise-like, but with uncertain distribution,

very many elementary sources: noise-like and probably near Gaussian in distribution.

In using this classification we need to think of these "elementary sources" as *independent elementary sources*. The return of a radar pulse from a large corner reflector involves a diversity of ray paths, mirrored on different parts of the corner reflector. These are not to be taken as separate elementary sources as long as the corner reflector is rigid. (A flimsy corner reflector, heavily vibrating in each of many modes, would return a noise-like signal, involving many local sources of reflection (each slowly changing)). Scattering by a number of scatterers, or multipath transmission, can convert a single real source into several or many virtual sources.

Wherever we understand the source and propagation phenomena, we can usually count elementary sources well enough to tell whether to expect noise-like or signal-like properties in what we receive or measure, and record.

2d. STABILITY IN THE NOISE-LIKE CASE

We have already, in the introduction, illustrated the stability behavior in the Gaussian case, and given the relevant formulas. These correspond, to

$$\text{degrees of freedom} = (\#\text{ data points}) \frac{(\text{bandwidth})}{\pi/(\text{data spacing})} \tag{2.1}$$

in which the bandwidth is in radians; where one degree of freedom for each data point is spread uniformly over the frequencies (see Section 6a) from $\omega = 0$ to $\omega = \pi/(\text{data spacing})$. (Another way to put it is that the number of degrees of freedom is the product of bandwidth/π and observation time.)

If we have a non-Gaussian distribution of data points, as happens so frequently, the variability of our spectrum estimates is almost always (in practice) increased. There is also some increase when our pass-bands are not the ideal square-cornered boxcars that "ideal" filters would give us. As a result, (2.1) and its companion (exact for any multiple of chi-square with specified degrees of freedom)

$$\text{variance} = 2(\text{average})^2/(\text{degrees of freedom}) \tag{2.2}$$

are either satisfactorily close, or give *too small* a value for the true variance (which is then larger that (2.2) indicates).

There is *no way* to analyze any ordinary *noise-like* input (for which a few-parameter form is not guaranteed), and obtain a reasonable estimate of the underlying process's spectrum which is *more stable* than is suggested by (2.1) and (2.2).

Sometimes it will not be possible to do even approximately this well without redefining the process (as when we divide the 10%-30%-60% process of Section 2a into three subprocesses, and estimate only the spectrum of that subprocess from which we have a realization).

We also have, particularly in dealing with *noise-like* processes, to fear the appearance either of a few or many gigantic deviations. These may come from a stretched-tail distribution of values in the original phenomenon, or from corruption by lightning flashes or other low-duty-cycle high-amplitude perturbations. In the latter case, we are likely to want to estimate the spectrum of the input with the corruption removed. In the former, we may have to settle for estimation of the spectrum associated with forgetting (and smoothing over) the gigantic values. Robust/resistant methods of spectrum analysis have been developed to do these things. They will be discussed in Section 19.

With adequate use of process-subdivision and robust/resistant analysis we can reasonably plan to bring the variance of our spectrum estimates down close to that indicated by (2.1) or (2.2), close enough for most design considerations, for the planning of data accumulation. (True variabilities 150% or 200% of those given by (2.1) or (2.2) need not surprise us.)

Higher stability of spectrum estimates can ONLY be attained for *signal-like* processes -- or for cases where trustworthy physical laws (not unwarranted optimism) guarantee a few-parameter description.

2e. *EXAMPLES OF NON-GAUSSIANITY*

We offer now, briefly, some examples of noise-like processes that are not Gaussian, trying to illustrate individual modes of failure, emphasizing as hard as we can that REAL DATA OFTEN FAIL to be Gaussian IN MANY WAYS.

Example 1. Each realization can be thought of as a realization of Gaussian white noise with variance σ^2, but σ^2 varies from realization to realization.

Example 2. Each realization is of the form $X(t) = \alpha\cos(\omega t + \phi)$, with α, ω and ϕ all random, where ϕ is uniform on $[0, 2\pi]$, independent of α and ω, which follow some messy joint distribution.

Example 3. $X(t)$ is always either $+1$ or -1.

Example 4. $X(t)$ is the sum of a realization of some fixed Gaussian colored noise and a peppering of random (say Poisson) values which affect only a small fraction of the observations (the contributions of this component are elsewhere zero).

Example 5.

$$X^*(t) = X(t) + c[X(t)]^3$$

where $X(t)$ is a realization of some fixed Gaussian colored noise.

Example 6.

$$X^*(t) = [10 + \cos(\nu t + \phi)]X(t)$$

where $X(t)$ is a realization of some fixed Gaussian colored noise ν is fixed and ϕ is a uniformly distributed random variate.

It is easy to be -- or become -- non-Gaussian; it is hard to be -- or remain -- Gaussian. (Narrow-band filtration is an exception.)

3. *SUPERPOSITION AND LINEAR TIME INVARIANCE; NON-ENTANGLEMENT*

We return now to the question of how systems keep frequency bands separate, or mix them up. The essential requirements of a linear time-invariant system are:

- time-origin-shift invariance, which means that delaying the input by a given constant delays the output by that same constant, and
- superposability, which means that, if the input is a moment-by-moment sum of two inputs, the output is the moment-by-moment sum of the two corresponding outputs (often called "linearity" or "additivity").

In symbolic formulas these read:

- if $x(t) \rightarrow y(t)$, then, for any admissible h, $x(t+h) \rightarrow y(t+h)$, and
- if $x_1(t) \rightarrow y_1(t)$ and $x_2(t) \rightarrow y_2(t)$, then $x_1(t)+x_2(t) \rightarrow y_1(t)+y_2(t)$.

The concept is much discussed in signal processing texts, e.g. Oppenheim and Schafer (1975). Simple differencing (for discrete equispaced time) or differentiation (for continuous time) shows that if filter input is $X(t) = \cos(\omega t + \phi)$ then its output must be of the form $C \cdot \cos(\omega t + \phi + \psi)$ where C and ψ depend on ω alone and the complex-valued function of frequency

$$B(\omega) = Ce^{i\psi}$$

is called the transfer function of the system. This says that single-frequency signals are only changed in amplitude and phase.

Superposition now tells us both the parallel result for each of several frequencies applied at the same time, and that what happens to each frequency is unaffected by the presence of the others. Thus linear time-invariant systems do leave frequency bands unshifted and unblurred. If we see the world through such a linear time-invariant system we can know something of what went on -- in frequency terms -- at its input. Thinking and working in frequency terms can help us do both this and its generalizations.

For more discussion and generalization of linear time-invariant systems see Section 13.

4. *TWO ATTEMPTS TO ESCAPE REALITY*

Both statisticians and physicists have shown a marked tendency to over-simplification, specifically by assuming that the probabilistic (stochastic) parts of their data could be described in terms of a few parameters, and that these parameters could be reasonably chosen in such a way as to be estimable from the data. The last 50 years have seen the statisticians struggling to free themselves from such straight-jackets, to develop procedures that were known to work respectably across a wide variety of stochastic (probabilistic) situations.

This great increase in realism came first, quite incompletely, through the development of so-called nonparametric or distribution-free methods and later, in the last decade or so, much more completely though the development of so-called robust/resistant methods. (There remain a variety of ways in which realism needs to be still further increased, so that there will be important work for the decades ahead, but it seems likely that the most important steps have been taken.)

Some would say that when they, or others, "assume" they are really only talking about approximations. This is safe for those who understand, but dangerous for innocent readers. To say that b in $\hat{y} = a + bx$ is "the slope of the curve $y = f(x)$" is similarly dangerous when $f(x)$ is not a straight line. Perhaps the basic difficulty is unfamiliarity with the idea of a leading case (see Section 17), and a consequent belief that the only place one can start from is an assumption.

The most unfortunate consequence is the sequence: (a) this might be a good approximation, (b) I'll "assume" it and see if it works, (c) "ah ha", it works well enough for my problem, (d) since it works (well enough) it must be the truth, (e) so I must be estimating the spectrum, (f) so other methods of spectrum estimation are no good. In this sequence, (a), (b), and (c), except for the choice of the word "assume" are frequently practical and often adequate. The trouble comes from the illogicality of attaching (d), (e) and (f). (Even (c) fails, if the trial is not made on sufficiently real and sufficiently diverse examples.)

There are two ways in which this "unrealism through narrow specification" is prominent in the literature of spectrum analysis. We all need to recognize both, and understand where it is important to be concerned about each.

One is within the spectrum framework, where it has seemed natural to some to represent spectra by functional forms involving just a few parameters.

The other is outside that framework, where it has been thought that it was enough to know the spectrum, that there was never any need to go further. This simplistic view is often put forward under the shallow concealment of the assumption that the process is Gaussian and *hence* that we need go no further than the spectrum.

4a. *THE FEW-PARAMETER FUNKHOLE*

The idea that we can do well by assuming a specific, few-parameter form for the spectrum is seductive. But does the evidence suggest that we should allow ourselves to be seduced?

There are a variety of physical phenomena which are fairly well represented by line (or band) spectra -- speech (for a short time interval), the tides, atomic and molecular spectra, Milankovich's analysis of the changes of the earth's orbit over time. These have localized-parameter descriptions (two or three parameters per line) but *not* few-parameter descriptions.

For every case that we know well, a few-parameter description is only an approximate one. This means that if we fit a few-parameter model, we ought, almost always, to think of it as an approximation -- to think of what the parameters tell us as *averaged information*, as being given by suitable averages over the spectrum, or some other function of frequency. We ought to believe that the interpretation of these few parameters has to be consistent with much more diverse and flexible sorts of spectra than can be represented explicitly in few-parameter forms. After all, when we take the mean of successive observations, we need not believe in some underlying *exact* constancy; when we fit a straight line to (y,x) data, the true dependence of y on x is rarely *exactly* a straight line. (For related discussion, see Section 17.)

From this point of view, moving from few-parameter models to estimates of averages over each of moderate number of bands is a step toward reality, a step from few "parameters" to moderate number of "parameters". (The idea that one can, starting from samples of conventional lengths, go very much further for noise-like processes is, of course, like the Lorelei, a route to danger and destruction, though these do not occur every time.)

4b. THE GAUSSIAN FUNKHOLE

The view that, since we are only going to estimate the spectrum, we must assume that our processes, actual and potential, are such that (together with a mean) the spectrum will tell us all about the process -- either all there is to know, if we really knew the spectrum, or all there is to estimate, if we are working from data -- is a parallel funkhole. Again it tries to define or assume away all the detail or diversity we have decided, often for good reason, not to try to capture. Again it is dangerous at best.

A good way to make clear why this is only a funkhole is to spend a little time on a variety of processes with the same spectrum. It may as well be the simplest spectrum -- so let us turn to "white noise".

4c. GAUSSIAN WHITE NOISE

Here again, in thinking of any white noise, some care concerning the relation of discrete and continuous time is needed. In discrete time, if $X(t)$ is a realization and $ave\ X(t) = \mu(t)$ is the mean across the process, we require

$$ave\{(X(t)-\mu(t))(X(t+s)-\mu(t+s))\} = 0, \quad \text{for } s \neq 0$$

where "*ave*" denotes an average either/or both "across the process" (over different realizations) and over t. (Stationarity is a third funkhole.) This is a relatively precise concept, and corresponds to a discrete-time spectrum which is nearly flat over $0 < \omega < \pi/\Delta$, where Δ is the time spacing.

What does it mean to be a white noise in continuous time? To have an effectively flat spectrum, presumably, which means to have a spectrum nearly enough flat over a sufficiently broader baseband so that, when aliased on to a reasonable baseband by recording in discrete time -- or when low-passed with an extremely sharp-cutoff filter -- the result will have a very nearly flat spectrum.

For the discrete-time case, it suffices to say, as the ideal, that

$$ave(X(t)) = \mu, \text{ for all } t \tag{4.1}$$

$$var(X(t)) = \sigma^2, \text{ for all } t \tag{4.2}$$

but that

$$ave((X(t)-\mu)(X(u)-\mu)) = 0, \text{ for all } t \neq u \tag{4.3}$$

which describe discrete-time white noise, and that the distribution of $X(t)$, for each t, follows the Gaussian shape. (Very familiar to the statistician and, in his other contexts, usually overidealized.)

4d. ON-OFF WHITE NOISE

We can combine (4.1) to (4.3) with the condition that all realizations have $X(t) = 0$ or $X(t) = 1$ for all t. The result is properly called on-off white noise, and comes in very different flavors, depending upon

$$\frac{\text{prob}\{X(t)=1\}}{\text{prob}\{X(t)=0\}} = \frac{\gamma}{1-\gamma}$$

If $\gamma = 10^{-4}$, roughly one digitization in 10,000 will be unity, the others will be zero (very dilute single-data-point spikes). If $\gamma = 1/2$, we have a process that might have arisen as $\phi(Y(t))$ where $Y(t)$ is Gaussian (on other symmetrically distributed) white noise and ϕ only considers the sign of $Y(t) - \mu$.

These noises -- they are a one-parameter family -- are very different from Gaussian white noise, yet they, too, are white, sharing a flat spectrum on $(0, \pi/\Delta)$.

4e. MULTICHIRP WHITE NOISE

Let $a(t)$ be any (generalized) chirp, vanishing for $t < 0$ and decaying toward zero as $t \to \infty$. (It will be convenient if the integrals of $|a(t)|$ and $(a(t))^2$ are finite.) Let $\tau_1, \tau_2, \ldots, \tau_{N(T)}$ be the times of occurrence of random events, and let

$$X(t) = \sum_j a(t-\tau_j)$$

Then, as discussed in Section 11, the spectrum of X will be $|A(\omega)|^2 S(\omega)$ where $A(\omega)$ corresponds to $a(t)$ and $S(\omega)$ is the spectrum of the point process that generated the τ's.

In discrete time, we can suppose that our choice of $a(t)$ and $S(\omega)$ compensate one another, after aliasing, so that the resulting spectrum is white. The rate at which our random events occur is now important. If they occur only rarely, then our time history (continuous or discrete) shows infrequent chirps, rarely overlapping one another. Such realizations, too, can be realizations of *white noise*.

White noises come in very many types -- we have only shown a few.

4f. COLORED NOISES ARE EQUALLY DIVERSE

However many kinds of white noises we may wish to think of, we can convert them into equally many kinds of noise of some other prescribed color by passing them through an appropriate filter.

B. OTHER BASICS

5. ENLIGHTENING EXAMPLES

Time series analyses have various kinds of aims. These include: discovery of phenomena, modeling, preparation for further inquiry, reaching conclusions, assessment of predictability and description of variability. Tukey (1980a) expands on these aims in some detail.

Some studies exemplifying these aims will now be described.

5a. AN EXAMPLE: SPECTROSCOPY

In general terms spectroscopy is the study of the characteristic frequencies at which resonating particles of matter absorb and emit energy. It is a powerful tool, for probing the microscopic details of the structure of matter. The Fourier-transform pulsed-nuclear-magnetic-resonance technique proceeds by irradiating a sample with an intense pulse of radio-frequency energy for a short time (1 to 100 microseconds) and Fourier analyzing the resulting electrical signal. The introduction of this Fourier technique resulted in a dramatic improvement in detection capability. (See Becker and Farran (1972) and Levy and Craik (1981).) (Previously the spectra were recorded in continuous time with the magnetic field or the frequency swept slowly.) The goal of the analysis is to achieve good resolution of the peaks in the spectrum. Levy and Craik (1981) provide a number of examples of advances in interpretation that have resulted from the use of Fourier-transform nuclear magnetic resonance in chemistry and biology.

In a situation involving matter far from a laboratory, by means of Fourier spectroscopy Connes *et al* (1967) were able to detect the presence of HF in the atmosphere of Venus (at an abundance ratio of a few parts per billion) and to estimate its amount. They were further able (Connes et al.,

1968) using the fine details of the carbon monoxide (CO) spectrum (only uncertainly detected, if at all, by previous measurements), to calculate the pressures and temperatures at which CO exists.

It is staggering to consider the resolution thus achieved. Connes and Michel (1974) remark (about spectra of stars): "However, it will no longer be practical to publish the complete data in atlas form: one spectrum on the scale of figure 1 (which is inadequate to show line profiles) is 170m long."

5b. AN EXAMPLE: THE NEGATIVE IMPORTANCE OF STATIONARITY

A striking example of the detection capability of spectrum analysis is given in Munk and Snodgrass (1957). Ocean-wave spectra were estimated from data collected by a pressure recorder sitting on the ocean floor off Guadalupe Island. Spectra were estimated for contiguous four hour periods. A small peak, at a frequency that increased as time passed, was noted. The frequency of the peak increased with time as longer waves travel more quickly. From the rate of increase of the frequency of the peak the distance (14,000 km) of the storm was able to be estimated. From the amplitude of the spectrum, the wave height (1 mm!) could be estimated. The wave length was approximately 1 km. The source of this peak was then determined, on the basis of the rate at which its frequency *changed*, to be a storm in the Indian Ocean.

Clearly this discovery could not have been made if the recorded wave process had been stationary. At times nonstationarity is a great advantage.

5c. AN EXAMPLE: THE WT4-WAVEGUIDE

Waveguides for the transmission of wide-band high frequency radio signals have to be of very constant cross-section and to be laid exceedingly straight in order that energy loss not be great and that multiple modes not occur. Thomson (1977a,b) presents a spectral analysis of data collected by sending a mechanical mouse through 12 kilometers of WT4 waveguide measuring its curvature at frequent intervals (nearly every centimeter.) The dynamic range of the (spatial) spectrum in this case is substantial, up to 16 decades (160 db).

Thomson dealt with these difficulties by employing a high quality taper (prolate spheroidal function) and by employing a resistant prewhitening filter. Further details are given in Kleiner *et al* (1979).

In particular this study found that, in these rather extreme circumstances: one dust speck could flatten the low portion of a naive spectrum estimate, concealing that portion of the true spectrum, while two such could introduce an obliterating ripple, and that not tapering could result in overwhelming bias.

5d. AN EXAMPLE: CELL MEMBRANE NOISE

Muscle cells are electro-chemical devices. In the presence of the chemical acetylcholine, a background electrical noise is produced by a cell. Bevan et al (1979) recorded this noise and estimated its spectrum under a variety of conditions.

Physical reasoning suggested the form $\alpha/(1+(\omega^2/\beta^2))$ with the parameters α, β having physical interpretations, $1/\beta$ being the fraction of the time certain channels were open, and $\alpha\beta$ being the channel conductance. 60 Hz noise was found to be present and had to be filtered out. Initially, a ripple appeared on the spectrum. This seemed unlikely to be a physiological phenomenon. Eventually its cause was tracked down as feedback from the noise of a pen recorder graphing the series. When the pen noise was eliminated, the ripple went away. The above functional form was found to be reasonable except for some extra power at the low frequencies. The dependence of the parameters on factors that the experimenters altered was also studied.

6. TYPES OF FOURIER TRANSFORM

Suppose, for the moment, that the series of interest is continuous and defined for $-\infty < t < \infty$. If the piece $X(t)$, $0 \le t < T$ is available for analysis, then one might compute the finite Fourier transform

$$\int_0^T X(t)e^{-i\omega t}dt \ . \tag{6.1}$$

When, for example, that $X(t) = \alpha\cos(\beta t+\gamma)$, the absolute value of this Fourier transform will have maxima at $\omega = \pm\beta$. (This is the reason that computing the Fourier transform was proposed for estimating hidden periodicities by Stokes many years ago.) The Fourier transform itself is seen to be proportional to the sample covariance between the segment of the series X and the series $\exp\{-i\omega t\}$.

When only $X(t)$, $t = 0,\dots,T-1$ is available, one may compute the discrete Fourier transform

$$d_X^T(\omega) = \sum_{u=0}^{T-1} X(u)e^{-i\omega u} \ . \tag{6.2}$$

For the case of $X(t) = \alpha\cos(\beta t+\gamma)$, the absolute value of this transform has maxima at $\omega = \pm\beta$ as before, and also, equally large, at $\omega = \pm(\beta \pm 2\pi), \pm(\beta \pm 4\pi),\dots$. (These latter frequencies are referred to as aliases (or images) of the frequency β.) The Fourier transform (6.2) may also be of use in estimating the frequency β; however additional information will be needed to rule out the aliases. (See the next Section).

(6.1) and (6.2) are two types of Fourier transform that prove of use in dealing with random process data. In the case of a (2-dimensional) spatial process, $X(t_1,t_2)$ the transforms

$$\int_0^{T_1}\int_0^{T_2} X(t_1,t_2)e^{-i(\omega_1 t_1+\omega_2 t_2)}dt_1dt_2, \qquad \sum_{t_1=0}^{T_1-1}\sum_{t_2=0}^{T_2-1} X(t_1,t_2)e^{-i(\omega_1 t_1+\omega_2 t_2)} \ ,$$

prove useful. For point processes, whose realizations are sequences of times τ_j at which some event of interest took place, the transform

$$\sum_{0\le\tau_j<T} e^{-i\omega\tau_j}$$

is relevant. For marked point processes, point processes where a value M_j is associated with each time τ_j, a key transform is provided by

$$\sum_{0\le\tau_j<T} M_j e^{-i\omega\tau_j} \ .$$

In classical Fourier analysis it has been found convenient, in many circumstances, to insert convergence factors into the transforms, yielding expressions such as

$$\int_{-T}^{T}\left(1-\frac{|t|}{T}\right)X(t)e^{-i\omega t}dt$$

for example. (This is a simple example, not a preferred window.) Such modifications of the Fourier transform prove essential in the time series situation as well, and go under the names of "tapering" or "windowing". (cp. Blackman & Tukey 1959, Kaiser 1966.)

We next discuss key aspects of empirical Fourier transforms such as those defined above. It is worthwhile to add the comment that using a Fourier transform does not require that a periodic phenomenon be present — its usefulness is much broader.

6a. CONTINUOUS TIME VERSUS DISCRETE TIME: ALIASING

Numerous processes whose realizations can be seen to develop or endure are best thought of in continuous time, a few are meaningful for discrete time alone. Various computational aspects however, are best dealt with in terms of discrete time. In consequence it is important that the effects of switching between continuous and discrete time (for example by sampling the continuous series at discrete times) be well understood.

Suppose measurements of the continuous series are made at equally spaced points in time, Δ time units apart. If the sampling instants are $t_n = n\Delta$, then because of the periodicity property $\cos(\omega+2\pi) = \cos(\omega)$ one sees that sampling the series $\alpha \cos([\beta+\frac{2\pi}{\Delta}]t+\gamma)$ would lead to exactly the same set of observed values as found by sampling $\alpha \cos(\beta t+\gamma)$. The frequencies β and $\beta + 2\pi/\Delta$ cannot be distinguished from the data. Similarly the frequencies $\beta + 2\pi k/\Delta$, $k = \pm 1, \pm 2, \ldots$ cannot be distinguished from β. Further the frequency $-\beta$ cannot be distinguished from the frequency β as it too corresponds to a cosine wave of frequency β. In consequence the frequencies $-\beta + 2\pi k/\Delta$ are all indistinguishable, if values of the series are available only every Δ time units. The frequencies $\pm \beta + 2\pi k/\Delta$, $k = 0, \pm 1, \pm 2, \ldots$ are called aliases of each other. There is but one in the interval $[0, \pi/\Delta]$, it is called the principal alias.

Aliasing cannot be avoided totally in practice; however its effects can be much reduced by prefiltering the series in such a way that virtually all the power above the highest frequency of interest is removed, and then choosing Δ so small that one has ≥ 2 points (often ≥ 3 and possibly as many as 7) per cycle at this cutoff.

The aliasing of frequencies due to operating with equi-spaced data is bad, but without it one would have to spread the information in the data over $0 \leq \omega < \infty$, without the advantages of knowing what is aliased with what. This would be far worse.

In the interpretation of spectra estimated from equi-spaced data one must never forget the effects of aliasing. (And it is best not to forget to give thanks for aliasing's simplicity.)

Recently Tukey (1980a) emphasized that procedures which work with sums of squares of Fourier coefficients, $a_j^2 + b_j^2$ ($= \bar{d}_Y^T(\omega)\, d_Y^T(\omega)$), in the notation of (6.2) above, rather than with the $a_j + ib_j$'s (the $d_Y^T(\omega)$'s) (or their real and imaginary parts) also alias *time separations*. In particular, two spikes nearly at the ends of the data segment being Fourier transformed can yield the same sequence of values $a_j^2 + b_j^2$ (where $a_j + ib_j$ is the complex Fourier coefficient) as two spikes close together. This can be avoided by adding enough zero's to the data sequence (which still needs to be first tapered, see Section 6d below) before carrying out the Fourier transform.

6b. *CONTINUOUS VERSUS DISCRETE FREQUENCY*

The previous section indicated some of the effects that discrete time has on analysis and interpretation. Frequency may also be discrete.

The Fourier transform (6.2) may be inverted via the relationship

$$X(t) = \frac{1}{T} \sum_{u=0}^{T-1} d_X^T \left(\frac{2\pi u}{T}\right) \exp\{i\, \frac{2\pi u t}{T}\} \tag{6.3}$$

$t = 0, \ldots, T-1$. This makes it appear that only the discrete frequencies $2\pi u/T$, $u = 0, \ldots, T-1$ are needed for a description of the time series. The issue here is however that, while these discrete frequencies do suffice to represent the data, they are almost never able to represent the phenomenon under study. The right hand side of expression (6.3) defines a function for all values of T. This function is seen to have period T, (because the frequencies are equispaced). It might correspond to the phenomenon under study if that phenomenon had period T. This is most unlikely, however, as T was, usually, merely the number of data values collected.

The user must beware that Fourier manipulations can insert periodic artifacts in such a way.

6c. *MIXTURES (BETWEEN DATA AND NOISE)*

Suppose that the series is a mixture of a signal of interest and noise, $X(t) = s(t) + \epsilon(t)$. As the Fourier transform is a linear operator, one has

$$d_X^T(\omega) = d_s^T(\omega) + d_\epsilon^T(\omega) \ .$$

T being the length of the time period of observation. The data transform is the sum of the signal transform and the noise transform. This has practical implications.

In many natural situations, the signal is concentrated in specific frequency bands. In these circumstances one need only deal with $d_X^T(\omega)$ for ω in those bands.

If the series X is passed through a linear time invariant system with transfer function $A(\omega)$ to obtain a series Y, then in terms of Fourier transforms one has

$$d_Y^T(\omega) \quad A(\omega)d_X^T(\omega)$$

$$A(\omega)d_s^T(\omega) + A(\omega)d_n^T(\omega)\ .$$

One can set about determining a filter such that $A(\omega)d_n^T(\omega)$ is generally small, while $A(\omega)d_s^T(\omega)$ is not, and may succeed.

6d. TAPERING

Suppose that $X(t) = \cos \beta t$, for $t = 0,1,...,T-1$ then $d_X^T(\omega) = \frac{1}{2}\Delta^T(\omega-\beta) + \frac{1}{2}\Delta^T(\omega+\beta)$ where

$$\Delta^T(\omega) = \sum_{t=0}^{T-1} e^{-i\omega t} = \exp\{i\omega(T-1)/2\}\,\frac{\sin \omega T/2}{\sin \omega/2}\ .$$

The function $\sin \omega T/2 \,/\, \sin \omega/2$, the Dirichlet kernel, has substantial ripples and appreciable size away from $\omega = 0$. This can cause substantial interference among components of different frequencies present in the series X. In practical examples leakage from one frequency to another can reach over substantial frequency intervals and can cause great difficulty of interpretation. The phenomenon reflects the fact that, with a finite amount of data, one can only look at particular frequencies with limited resolution.

An elementary procedure, variously called tapering, windowing, inserting convergence factors, or weighting, is available for reducing the effects of leakage.

Consider a function, $h^T(t)$, defined as 0 for $t < 0$ and $t > T-1$ and whose Fourier transform, $H^T(\omega) = \sum_t h^T(t)e^{-i\omega t}$, is concentrated near $\omega = 0$ and dies off rapidly as ω increases. (This will, in particular, be the case for functions h^T that are near 0 at the ends, rise to say 1 in the center, then fall to 0, and are very smooth.) The Fourier transform of the series $Y(t) = h^T(t)X(t)$ may be written

$$d_Y^T(\omega) = \sum_{t=0}^{T-1} h^T(t)X(t)e^{-i\omega t} = \frac{1}{T}\sum_{s=0}^{T-1} d_X^T\left(\frac{2\pi s}{T}\right)H^T\left(\omega - \frac{2\pi s}{T}\right) \tag{6.4}$$

For a series $X(t) = \cos \beta t$, $d_Y^T(\omega) = \frac{1}{2}H^T(\omega-\beta) + \frac{1}{2}H^T(\omega+\beta)$ and leakage is reduced to the extent that H^T has better resolution than Δ^T above.

In practice, as the dramatic examples of Thomson (1977) and Van Schooneveld and Frijling (1981) show, tapering is essential in the preliminary estimation of spectra of unknown form.

Harris (1978) presents quite a number of tapers with graphs of h^T and the corresponding Fourier transform H^T. See also Nuttall (1981).

While expression (6.4) shows that one could "taper" after forming the Fourier transform (at least if all calculations were to high enough precision), this is not computationally sensible in many circumstances. It is better to multiply by h^T prior to Fourier transforming, except in special circumstances.

In the direct method of spectrum estimation one smooths the periodogram

$$|d_Y^T(\omega)|^2 = \frac{1}{T^2} \sum_{s_1} \sum_{s_2} H^T\left(\omega - \frac{2\pi s_1}{T}\right) \overline{H^T\left(\omega - \frac{2\pi s_2}{T}\right)}\, d_X^T\left(\frac{2\pi s_1}{T}\right) \overline{d_X^T\left(\frac{2\pi s_2}{T}\right)} . \qquad (6.5)$$

The result is invariably NOT a smoothed version of the (untapered) periodogram $(2\pi T)^{-1}|d_X^T(\omega)|^2$. One should not miss the opportunity to taper prior to forming a periodogram in any circumstances where the spectrum is not KNOWN to be very, very nearly flat. Priestly (1981) Chapter 7 is a further reference.

6e. *PREFILTERING*

In the case that the series X is stationary with mean 0 and spectrum $f(\cdot)$, the expected value of the periodogram (6.5) is

$$\int |H^T(\omega-\alpha)|^2 f(\alpha) d\alpha . \qquad (6.6)$$

This expression makes it apparent that the more nearly constant the function $f(\cdot)$ is, the less biased the estimate. Now, if the series X is passed through a linear time-invariant filter with transfer function A the spectrum of the resulting series is $|A(\cdot)|^2 f(\omega)$, and the expected value of its periodogram is

$$\int |H^T(\omega-\alpha)|^2 |A(\alpha)|^2 f(\alpha) d\alpha .$$

In many circumstances it is possible to determine a filter, probably digital, such that the function $|A(\cdot)|^2 f(\cdot)$ is more nearly constant (has been partially prewhitened) than $f(\cdot)$ itself. These remarks suggest the estimate $|A(\omega)|^{-2} f^T(\omega)$, where $f^T(\omega)$ is the spectral estimate of the prewhitened series. In essence one filters the series of interest prior to estimating the spectrum and then compensates for what the filter did.

In practice it can be essential that some form of prefiltering be carried out in spectrum estimation. Prewhitening can also help in terms of the determination of the variability of a spectrum estimate.

6f. *QUADRATIC WINDOWING*

One class of spectrum estimates takes the form

$$(2\pi)^{-1} \sum_u w^T(u) c^T(u) e^{-i\omega u} = \int W^T(\omega-\alpha) I^T(\alpha) d\alpha \qquad (6.7)$$

where, in the case that the series has 0 mean

$$c^T(u) = T^{-1} \sum_{t=0}^{T-|u|} X(t+u) X(t) ,$$

is an estimate of the covariance function, $w^T(u)$; $u = 0, \pm 1, \pm 2, \ldots$ is a sequence of convergence factors with Fourier transform $W^T(\alpha) = \sum_u w^T(u) \exp\{i\alpha u\}$, and where $I^T(\alpha) = (2\pi T)^{-1}|d^T(\alpha)|^2$ is the periodogram of the data. The use of $w^T(u)$, or equivalently, of $W^T(\omega)$ is called using a quadratic window.

The expected value of the estimate (6.7) is

$$(2\pi T)^{-1} \int \left[\int W^T(\omega-\alpha-\beta) |\Delta^T(\beta)|^2 d\beta\right] f(\alpha) d\alpha$$

with Δ^T the Fourier transform of the boxcar of length T introduced above. This expression has the same form as (6.6), indeed as Carter and Nuttall (1980) and Van Schooneveld and Frijling (1981) remark one can derive a quadratic window such that the corresponding estimate has the same expected value as any tapered estimate. The variances of the two estimates, however, are often but not always quite different, as are the susceptibilities to roundoff (during the usual calculations).

Another difference is that the indirect estimate, via the quadratic window, can take on negative values.

6g. VARIANCE CONSIDERATIONS

Comparative discussion of spectral estimates cannot be carried out in terms of expected values alone. The techniques of tapering and prewhitening have been advanced as means of manipulating the bias of spectrum estimates. Of course, if their use makes the estimates less stable one can end up with a poorer estimate. It is crucial to also discuss the variance of a spectrum estimate and to provide an estimate of that variance.

All spectrum estimates we have considered are quadratic functions of the data and so can be represented as

$$\sum_{t,u} q^T(t,u)X(t)X(u) .$$

whose expected value is

$$\int Q^T(\alpha,\alpha)f(\alpha)d\alpha$$

where

$$Q^T(\alpha,\beta) = \sum_{t,u} q^T(t,u)e^{i(t\alpha-u\beta)} .$$

In the case that the series X is Gaussian its variance is

$$2\int\int |Q^T(\alpha,\beta)|^2 f(\alpha)f(\beta)d\alpha\, d\beta . \tag{6.8}$$

See for example Anderson (1971), p. 445. (In many non-Gaussian cases this will provide a close approximation as well.) Expression (6.8) may be employed to derive explicit expressions for the variances of the various spectrum estimates. One general characteristic that expression (6.8) makes clear is that the variance of an estimate depends on the whole course of the population spectrum $f(\omega)$. In cases where a spectrum falls off rapidly, substantial leakage into the low-power region may be expected to come from other frequencies. Tapering and prefiltering will be required to reduce the effects of this leakage.

7. COMPLEX-DEMODULATION AND SINGLE SIDE BAND TRANSMISSION

The basic ideas here are: narrow-band filtering to allow certain frequency dependent features to stand out more clearly, and frequency translation to slow down (or, possibly speed up) oscillating phenomena.

7a. MULTIPLICATION AND SMOOTHING

Consider a signal $X(t) = \cos(\beta t+\gamma)$. One has the trigonometric identities

$$2X(t)\cos(\omega t) = \cos([\beta-\omega]t+\gamma) + \cos([\beta+\omega]t+\gamma)$$

$$2X(t)\sin(\omega t) = -\sin([\beta-\omega]t+\gamma) + \sin([\beta+\omega]t+\gamma)$$

In each case, for ω near β the first term on the right-hand side is slowly varying, while the second oscillates more rapidly. When these functions are smoothed in time, the initial approximation is that the first terms will be unaltered and the second eliminated, in each case for ω near β.

The results of forming $X(t)\cos\omega t$, and $X(t)\sin\omega t$ and then smoothing separately are the real and imaginary parts of the complex demodulate of the series X at frequency ω. The procedure may be seen as one of shifting the component of frequency β, in the series X where β is near ω, down to frequency $\beta-\omega$. Low-pass filtering now allows one to look effectively at a restricted range of frequency around ω. Further, if the phenomenon under study is evolving in time, study of complex demodulates may allow the detection of that evolution. The spectrum looks merely at the average behavior across the time period of study.

On many occasions it is convenient to display the running amplitude and phase, rather than the rectangular coordinates, of the demodulate. Displaying the instantaneous frequency also often proves useful.

Complex demodulation leads directly to spectrum estimates. The time average of the square of the running amplitude is proportional to the power spectrum at frequency ω.

7b. REMODULATION

Let $U(t)$, $V(t)$ denote the series resulting from smoothing the $X(t)\cos\omega t$, $X(t)\sin\omega t$ respectively, where ω is the angular frequency used in the complex demodulation. The series

$$U(t)\cos\omega t - V(t)\sin\omega t$$

$$U(t)\sin\omega t + V(t)\cos\omega t$$

tend to fluctuate with frequency ω. The (composed) transformations from X to these series provided the smoothing is by exactly equivalent linear filters, are themselves linear filters. The first formula corresponds to narrow band-pass filtering the series near frequency ω, the second to a combination of narrow band-pass filtering and Hilbert transforming.

7c. SSB

If one started with a series $X(t) = \cos(\beta t + \gamma)$, complex demodulated at frequency ω, near β, and then remodulated at frequency ω, as the above discussion shows, one would be led back to the original series. Suppose however one redemodulated at frequency ω_1, then one is led to $\cos([\beta-\Delta]t+\gamma)$ where $\Delta = \omega - \omega_1$. If ω_1 is small compared to ω, then what has happened is that the frequency β has been pushed down by Δ.

This is the down-modulation version of SSB transmission which is now so popular in CB-radio, (with SSB up-conversion in the transmitter and SSB down-conversion in the receiver).

One early reference is Weaver (1956).

7d. COMPLEX-DEMODULATION AS AN FT PRECURSOR

The appearance of fast Fourier transform programs (e.g. Digital Signal Processing Committee (1979)) allowed a direct approach to the construction of estimates of frequency domain parameters when lengthy data stretches were available. For example the power spectrum could be estimated by i) forming the FT, $d^T(\omega)$, ii) forming the periodogram $|d^T(\omega)|^2$, iii) smoothing the periodogram. (The cross-spectrum (see Section 8c) of two series X and Y could be estimated by i) forming the FT's $d_X^T(\omega)$, $d_Y^T(\omega)$, ii) forming the cross-periodogram $d_X^T(\omega)d_Y^T(-\omega)$, iii) smoothing the cross-periodogram.)

The first direct approaches (earlier than the FFT) to become computationally effective used complex demodulation. The power spectrum at frequency ω was estimated by smoothing $U(t)^2 + V(t)^2$ where $U(t)$, $V(t)$ were the demodulates at frequency ω. (The cross-spectrum was estimated by smoothing $(U_X(t)+iV_X(t))(U_Y(t)-iV_Y(t))$.)

Complex demodulation remains a powerful technique for spectral estimation in real time and for the estimation of higher-order spectra of lengthy series. The demodulates may be computed by an FFT when appropriate. (See Bingham *et al* (1967).) They may be sampled at a crude (time) spacing if there are storage or computing limitations.

7e. AN EXAMPLE; FREE OSCILLATIONS

The response of many dynamical systems to an impulse is a linear combination of decaying cosines,

$$\sum_{k=1}^{K} \alpha_k \exp\{-\beta_k t\} \cos(\gamma_k t + \delta_k) \; . \qquad (7.1)$$

The associated movements of the system are called its free oscillations. For many physical systems, e.g. the Earth's response to major earthquakes, it is of crucial interest to estimate the eigenfrequencies γ_k and the associated quality factors $Q_k = \gamma_k/(2\beta_k)$.

In the aftermath of the great Chilean earthquake of May 1960, many spectrum estimates were computed from seismograms, and eigenfrequencies estimated by the locations of peaks in those spectra (see Tukey (1966).)

Because of the time varying character of the signal (7.1), however, complex demodulation is a substantially more useful technique for this situation. Bolt and Brillinger (1979) present the results of demodulating a Chilean record for a number of frequencies. The logarithm of the running amplitude is seen to fall-off in a linear fashion confirming the appropriateness of expression (7.1). It is further seen that the decay rate β_k depends on the frequency γ_k in a direct fashion.

7f. AN EXAMPLE; EDGE WAVES

Edge waves are (water) waves moving sideways to the shore rather than rolling onto the shore. They are usually caused by the superposition of incident and reflected wave trains. Their formation is a nonlinear phenomenon, indeed if the frequency of the incident wave is γ, and conditions are favorable to the formation of edge waves, then the edge waves will have frequency $\gamma/2$, (a subharmonic).

Lin (1981) makes extensive use of complex demodulation to study the growth and decay of edge waves generated in a wave tank. He fits theoretical models to the growth and decay rates under various experimental conditions. The edge waves are allowed to reach a stationary state. Their frequency in this state is estimated by complex demodulation, then that frequency is used for demodulation of the records in the growth and decay periods (corresponding to the turning on and turning off of the wave generator, respectively). The running amplitude of the resulting complex demodulate provides the growth and decay rates of interest.

Lin finds that the reflection coefficient of the beach is a crucial parameter in the situation. He discovers that it is the (amplitude of the) reflected wave that excites the edge wave and drives its growth. By putting demodulates of the edge wave and reflected wave side by side he is able to see the nonlinear transfer of energy from the latter to the former.

Complex demodulation provided an appropriate tool here, because the phenomenon under study was limited to narrow bands (albeit essentially nonlinear) and changing with time (for the studies carried out). Through complex demodulation, experimental verification of a theoretical mechanism was provided, and the effects of departures from assumptions were determined.

C. VECTOR SPECTRA

8. VECTOR CASES

Once we have two or more simultaneous series, or, if you prefer, a vector valued series, we can do more than consider the spectra of the individual series, which described how self-energy is distributed over frequency. We can also consider how the energy shared between two series is distributed over frequency. Since sharing can be in phase, or in quadrature, or as a combination of these, the resulting cross-spectrum has to be complex-valued. Sometimes we want to consider its values in polar coordinates (magnitude and phase); sometimes in rectangular coordinates (cospectrum and quadspectrum).

Just as, in non-time-series statistics, least-squares regression (based on covariances as well as variances) is a more powerful and searching tool than variance components, so too cross-spectra help us unlock many doors that spectra alone do not -- and cannot.

8a. *COSPECTRA*

Tables of quarter-squares were once widely available to allow multiplication via the quarter-square identity

$$uv = \frac{1}{4}(u+v)^2 - \frac{1}{4}(u-v)^2$$

This expression can serve to define and interpret the covariance of two random variables. We can define, similarly, the cospectrum, either for a process or for an estimate from a realization, by

$$cospec(x_i, x_j) = \frac{1}{4} spec(x_i + x_j) - \frac{1}{4} spec(x_i - x_j)$$

Notice that this implies

$$cospec(x_i, -x_j) = -cospec(x_i, x_j)$$

so that negative values, alongside positive ones, are common.

One analog of least-square fitting of y in terms of x would be to pass X through a linear time-invariant filter with transfer function $B(\omega)$ and then try to minimize the spectrum of

$$Y - B[X]$$

If we restrict the filter to be time-symmetric around zero, thus zero-phase, and with $B(\omega)$ real-valued the spectrum of this difference is, with an obvious notation for the spectra and cross-spectrum

$$f_{YY}(\omega) - 2B(\omega)f_{XY}(\omega) + B^2(\omega)f_{XX}(\omega)$$

where under our restrictions, f_{XY} is real, thus reducing to the cospectrum. This is minimized when

$$-2f_{XY}(\omega) + 2B(\omega)f_{XX}(\omega) = 0$$

that is, when

$$B(\omega) = \frac{f_{XY}(\omega)}{f_{XX}(\omega)}$$

in complete analogy to the covariance/variance expression for an ordinary least-squares regression coefficient.

In meteorology for example, the cospectrum of vertical and horizontal wind velocities gives the frequency analysis of the Reynolds stresses, which mediate the vertical transfer of horizontal momentum, etc. (One early reference is Panofsky (1967)).

8b. *QUADSPECTRA*

When a Hilbert transform, H, is sufficiently closely realizable, then

$$cospec(H([X],Y) = -cospec(X, H[Y]) = quadspec(X, Y)$$

(The Hilbert transform is the filter with transfer function $-i \, sgn \, \omega$, see Brillinger (1981), p. 32.) We discuss the calculation of quadspectra below as part of the calculation of cross-spectrum, where

cross-spectrum = cospectrum + i quadspectrum.

8c. *CROSS-SPECTRA*

We have already discussed estimating a spectrum by averaging values

$$|d_Y^T(\omega)|^2 = \overline{d_Y^T(\omega)} \, d_Y^T(\omega)$$

of the periodogram of the tapered series $Y(t) = h(t)X(t)$. If now $Y_j(t) = h(t)X_j(t)$ and

$Y_k(t) = h(t)X_k(t)$ are the tapered forms of two simultaneous series, with Fourier transforms $d^T_{Y_j}(\omega)$ and $d^T_{Y_k}(\omega)$ respectively, we can form the corresponding cross-periodogram

$$d^T_{Y_j}(\omega)\ \overline{d^T_{Y_k}(\omega)}$$

which is complex-valued, and then average (locally in ω) to estimate a cross-spectrum.

8d. REGRESSION COEFFICIENTS

If we return to $Y - B[X]$ and give up constraining B, we find that the spectrum of this residual series is minimized when

$$B(\omega) = \frac{\text{cross spectrum}_{YX}(\omega)}{\text{spectrum}_{XX}(\omega)} = \frac{f_{YX}(\omega)}{f_{XX}(\omega)}$$

Thus the minimizing transfer function, or Wiener filter, which we might have expected to play the role of an array of complex-valued regression coefficients, not only does just that, but is defined in strict analogy to the usual (non-time-series) definition of regression coefficients.

Three things are important to remember:

- each statistical concept definable by first and second moments in a non-time-series situation has an analog definable by means, spectra, and cross-spectra.
- if the scalar concept is non-negative, (variance, multiple correlation), the time-series concept is a (non-negative) real-valued function of frequency (individual spectra, coherences, etc.)
- if the scalar concept can have either sign, (almost everything else) the time-series concept is a complex-valued function of frequency.

The entire armamentarium of low-moment statistics has its analogs, all functions of frequency, nearly all complex valued.

8e. COHERENCE: DEFINITIONS AND ADJUSTMENT

If we ask what fraction of spectrum (Y) remains when spectrum ($Y - B[X]$) is minimized, or what fraction of spectrum (X) remains when spectrum ($X - C[Y]$) is minimized, the answer is a further function of frequency, 1 - the coherence, where the coherence is given by

$$\text{coherence}_{YX}(\omega) = \frac{\overline{\text{cross-spectrum}_{YX}(\omega)}\text{cross-spectrum}_{YX}(\omega)}{\text{spectrum}_{XX}(\omega)\text{spectrum}_{YY}(\omega)} = \frac{|f_{YX}(\omega)|^2}{f_{XX}(\omega)f_{YY}(\omega)}\ .$$

(This is the analog of $1 - R^2$, where R is the multiple correlation coefficient squared as used in multiple regression studies.)

If we try to fit y with an x irrelevant to it, we expect *minimized* $var(y - bx)$ to be smaller than $var\ y$. Similarly, if we estimate the cross-spectra and the two spectra by averaging the cross-periodogram and the periodograms, all in the same way, one corresponding to simple averages of m terms each, we will find some estimated coherence, even when X is wholly independent of to Y. One simple adjustment

$$\text{adjusted coherence} = \text{coherence} - \frac{2}{m}$$

where the bandwidth of the spectrum and cross-spectrum estimates deserves m degrees of freedom, does fairly well in avoiding trouble from this source. No plot of *raw* coherence should ever be made without including the horizontal line at "coherence $= 2/m$"!

The relation

$$\overline{\text{coherency}} \cdot \text{coherency} = \text{coherence}$$

applies when we define a complex-valued quantity

$$\text{coherency}_{YX}(\omega) = \frac{\text{cross-spectrum}_{YX}(\omega)}{(\text{spectrum}_{XX}(\omega)\ \text{spectrum}_{YY}(\omega))^{1/2}}$$

that is essentially a normalized cross-spectrum. Early on, confusion between coherence and coherency was great; some still remains. Beware of remaining confusions. We follow Wiener (1930) in our use of the terms.

8f. *SPIRALLING*

If Y differs from X by delay or advance by a time-interval τ, the

a. the Fourier transforms differ by $e^{\pm i\tau\omega}$,

b. the spectra are the same, and

c. the cross-spectrum is $e^{i\tau\omega}$ times one spectrum and $e^{-i\tau\omega}$ times the other.

As a consequence, (theoretical) coherency is $e^{\pm i\tau\omega}$ but if τ is large enough, the $e^{\pm i\tau\omega}$ factor rotates so fast that averaging over ω, necessary for decently-variable estimates of spectra and cross-spectra, may return 0 or near 0 for the cross-spectrum.

The essential difficulty is in the cross-spectrum, where anything resembling

$$\overline{\left(d_X^T(\omega_1)\right)}d_Y^T(\omega_1) + \overline{\left(d_X^T(\omega_2)\right)}d_Y^T(\omega_2) + \cdots + \overline{\left(d_X^T(\omega_m)\right)}d_Y^T(\omega_m)$$

has a factor

$$e^{i\tau\omega_1} + e^{i\tau\omega_2} + \cdots + e^{i\tau\omega_m}$$

whose absolute value is less than or equal to $m \pm m$, with frequent zeroes, when τ is large.

It is easy for the elements of the cross-periodogram to spiral, and if they do, simple averaging does not produce usable cross-spectra. This problem requires constant vigilance, especially since it can be important in some frequency ranges and absent in others.

Techniques, based on using phases from moving averages of the cross-periodogram to bring individual values of the cross-periodogram close to zero phase, averaging results, and reinserting the phase have been proposed by Cleveland and Parzen (1975, see their Appendix B and Section 4). This scheme should work for all but the most extreme spiralling problems, where fitting of rough phase dependence may be a needed precursor of the use of this technique. For other approaches see that paper and its references. Some automatic scheme to control spiralling should be routinely used.

8g. *MULTIPLE REGRESSION*

If $Y, X_1, X_2, \ldots, X_k$ are simultaneous time series, we can try to minimize the spectrum of

$$Y - B_1[X_1] - B_2[X_2] - \cdots - B_k[X_k]$$

When we have estimated cross-spectra and spectra for each of several frequency bands, we get essentially the same equations, one set for each frequency band, as we would have for multiple (non-time series) regression of y on $x_1, x_2, \cdots, x_k$. The power spectra and cross-spectra replace the variances and covariances of that case.

The only important changes are that:

a. the cross-spectra are complex-valued

b. the $B_i(\omega)$ we are to solve for are also complex-valued.

8h. *AN EXAMPLE: POLARIZED LIGHT*

The classical analysis of polarized light -- into vertically polarized, horizontally polarized, and unpolarized components -- corresponds in XY-coordinates at 45° to the horizontal and vertical, to dividing

$$\text{spectrum}_X(\omega)\ \text{spectrum}_Y(\omega)$$

into

$$\text{cospectrum}^2_{YX}(\omega)$$

$$\text{quadspectrum}^2_{YX}(\omega)$$

for the linearly polarized components and

$$\text{spectrum}_X(\omega)\ \text{spectrum}_Y(\omega) - \text{cospectrum}^2_{YX}(\omega) - \text{quadspectrum}^2_{YX}(\omega)$$

for the unpolarized components. (A common scale factor is often natural.)

With X, Y horizontal and vertical, the same expressions correspond to right and left circularly polarized light for the first two and unpolarized light for the third. Wiener (1930) discusses coherency and polarized light in some detail.

8i. AN EXAMPLE: TOWER METEOROLOGY

Panofsky and McCormick (1954) discovered wind eddies rolling along the ground by carrying out cross-spectral analyses of horizontal and vertical components of wind velocity at a point on an observation tower. They found that at the proper height on the tower, the quadspectrum had one sign in the morning and the other in the afternoon.

Their explanation of what was happening was that eddies increase in size as the day progresses. If the measurement point were in the upper half of a rolling eddy in the morning and the lower half in the afternoon, then the observed switch in sign would occur.

D. OTHER EXTENSIONS OF SCOPE

9. CEPSTRA: THE AIMS

Repetition in time occurs for a variety of reasons, but mainly in one of two ways: (i) single repetitions, as in an echo, and (ii) quasi-equispaced repetitions, as in human speech which is driven by the repeated opening and closing of the vocal chords. Echo detection and identification has a variety of applications in geophysics. Pitch-detection, the identification of the time spacing between repetitions of vocal-chord behavior, is of considerable communication importance. (For the latter application see Rabiner and Schafer (1978), and Schafer (1979).)

To detect echoes and repetition rates effectively, we need to focus the information about them that we have in our data. This is not focusing on a frequency band -- a great variety of frequencies may participate in an echo. (We may have, however, in more difficult cases, to be prepared for differences in the strength, the phase shift, and even the time delay with which an echo appears at different frequencies.) It is not focusing in time, unless the original signal is itself narrowly focused in time, when the problem may be simple -- usually, however, the original signal extends over considerable time and overlaps its echo very substantially. (Time dependence of echo behavior is much less common -- radar and sonar doppler effects aside -- than frequency dependence.)

Echo identification is different from the problems we have so far dealt with, and requires a new approach. Because frequency dependence may matter, it is natural to begin by going to the frequency side, as we shall shortly do.

Like the other extensions of simple spectrum analysis (vector spectra, whether applied to ordinary or point processes), cepstra -- when their use is appropriate -- are much more effective than simple spectrum analysis.

The original reference is Bogert, Healy and Tukey (1963); a recent review can be found in Childers et al. (1977). See also Bogert and Ossana (1966).

9a. THE CALCULATIONS

If $X(t)$ has a reasonably simply behaved spectrum $f_{XX}(\omega)$ -- a condition met by dilute impulses, by moderately colored Gaussian noise, and by many intermediate processes -- the result of echoing it with a real echo is to convolve the signal with a function having two spikes, and hence to multiply its spectrum by the mod-square of the Fourier transform of the two-spike function. This last is of the form

$$a + b \cos \omega\tau$$

where τ is the time delay, and b/a the strength of the echo.

The data's spectrum is of the form

$$(\text{reasonably smooth function of } \omega) \cdot (a + b \cos\omega\tau)$$

so the natural next steps in isolating τ are

a. taking logs

b. looking for the ripple in $\log(a + b \cos\omega\tau)$

If b/a is large, finding the ripple (possibly somewhat concealed by "rahmonics", see below) should be easy. If b/a is small, then

$$\log(a + b \cos\omega\tau) = \log a + \frac{b}{a} \cos \omega\tau - \frac{1}{2} (\frac{b}{a})^2 \cos^2\omega\tau - \frac{1}{3} (\frac{b}{a})^3 \cos^3\omega\tau$$

$$= \log a - \frac{1}{4} (\frac{b}{a})^2 + \left\{\frac{b}{a} - \frac{1}{4} (\frac{b}{a})^3\right\} \cos \omega\tau + \frac{1}{4} (\frac{b}{a})^2 \cos 2\omega\tau - \frac{1}{12} (\frac{b}{a})^3 \cos 3\omega\tau + \cdots$$

so that $\cos \omega\tau$ is the dominant term of the ripple. ($\cos 2\omega\tau$, $\cos 3\omega\tau$, etc. are rahmonics).

We know how to find ripples in a series, we just have to look at its spectrum! In the present case, the series is a frequency-series (not a time-series) and its values are logs of estimated spectra (for moderately narrow intervals), but we should still look at its spectrum. It is convenient to maintain the frequency-time distinction, however, by saying "cepstrum" instead of "spectrum" however.

So the natural pattern is one of

a). calculating a fine-grained spectrum (but usually not an unsmoothed periodogram),

b). taking logs

c). often, getting rid of mammoth slow changes in the result by "liftering" (the analog of "filtering")

d). calculating the cepstrum from the result, just as we would calculate a spectrum from a time series.

To get information about the size and phase of echoes, we may do well, having found some specific τ, to fit

$$a + b \cos\omega\tau + c \sin\omega\tau$$

either alone or in combination with other terms, to our filtered log spectrum.

If echoing is frequency dependent, we may need to divide our final calculation step, (d) above, into separate calculations for different frequency regions.

9b. OTHER QUESTIONS, OTHER STATISTICS, OTHER APPROACHES

The approach in the last subsection is that of the initial paper (Bogert et al. 1963) and is focused on determining the time delay (of a single echo, or between successive repetitions). That paper also demonstrated the usefulness -- better here, worse there -- of the pseudo-autocorrelation, (which differs from the cepstrum in undoing the logarithm before the final Fourier transform.)

When phases are important, it is natural, from such an approach, to find the time delay or delays (quefrency or quefrencies) from either cepstrum or pseudo-autocorrelation and then to fit a general cosinusoid of the indicated quefrency

$$a \cos(\omega\tau_0+\phi)$$

to the (possibly liftered) log spectrum, when the fitted ϕ estimates the phase of the echo.

Just as cross-spectrum analysis has proved more effective than one-series-at-a-time spectrum analysis in making peaks clearly detectable (e.g. in unpublished work on monetary series by Milton Friedman and one of the authors), it is to be expected that cross-cepstrum analysis, in which we combine FT's of logs of spectra of two series (multiplying one by the complex conjugate of the other, rather than taking the squared absolute values of either) may well be effective in making time delays (quefrencies) common to both series clearly detectable. The necessary empirical evidence may not, however, be yet available.

Fortunately the whole area of desired answers, techniques, and approaches have been reviewed by Childers *et al.* (1977), who emphasize diversity in all three. This paper has 86 references of use to the interested researcher. It further stresses the relation between cepstrum analysis and the whole area of deconvolution.

9c. APPLICATIONS

Childers *et al.* (1977) stress three areas of application for this class of techniques: speech, where the first useful automatic pitch detection of human speech came from such techniques: seismic measurements, and hydroacoustic measurements, where, in particular, Mitchell and Bedford (1975) report measuring depths of distant (250 to 700 nautical miles) explosions in the ocean to standard deviations of $< 2\%$ of the actual depth.

Other interesting examples can be found in the references given by Childers et al. (1977). We content ourselves here with very brief mentions of a few applications in other areas, some of which raise questions or propose generalizations.

Miles (1975) and Syed et al. (1980), among others, report the use of cepstral techniques to correct for ground reflection in narrow-band sound spectra measurement outdoors.

Pearson et al. (1978) discuss the use of cepstral analysis in connection with ESR (Electron Spin Resonance) analysis of substituted triarylaminium cation radicals. They believe that, while "the cepstrum does not provide more information than the ESR spectrum from which it is derived", it "is a useful aid in the analysis of" their ESR spectra.

Roemer and Chen (1980), applying cepstrum technique to cable damage studies, offer evidence (which is hard to evaluate in the absence of information about the adequacy of tapering in their FFT method) that maximum-entropy spectrum analysis is helpful at the last stage of cepstrum analysis. Clearly a combination of empirical fine-tuning and theoretical insight is needed to clarify this question. Rather substantial effort is likely to be necessary, but may well be rewarding.

Rom (1975) discusses the 2-dimensional cepstrum briefly, and suggests its utility with respect to image-blurring problems (reporting success by another in this connection). Image improvement is difficult enough for this suggestion to be both encouraging and daunting.

10. HIGHER-ORDER SPECTRA

Power spectra and cross-spectra, being based on second-order statistics (and thus naturally called "second-order spectra") can only be expected to reveal so much about the situation. It is known that Gaussian processes are characterized completely by their first and second-order moments, but that is a very particular situation. The introduction of higher-order spectra follows naturally from either a desire to analyze higher-order moments (often as cumulants) of a process or a need to handle the effects of nonlinear operations on the process.

10a. THE GENERAL PROBLEM

The power spectrum, $f(\omega)$, provides the representation

$$cov\{X(t+u),X(t)\} = \int e^{iu\omega} f(\omega)d\omega$$

$u = 0, \pm 1,..$ of the covariance function of the stationary time series X. The bispectrum, a third-order spectrum, $f(\omega,\nu)$, provides the representation

$$cum\{X(t+u),X(t+v),X(t)\} = \int\int e^{i(u\omega+v\nu)} f(\omega,\nu)d\omega d\nu \tag{10.1}$$

of the third-order cumulant (or central moment) of the process. The cumulant spectrum of order k, $f(\omega_1, \ldots, \omega_{k-1})$, provides the representation

$$cum\{X(t+u_1),...,X(t+u_{k-1}),X(t)\} \tag{10.2}$$

$$= \int \cdots \int \exp\{i(u_1\omega_1 + \cdots + u_{k-1}\omega_{k-1})\}f(\omega_1, \ldots, \omega_{k-1})d\omega_1 \cdots d\omega_{k-1} \, . \tag{10.2}$$

The joint cumulant, $cum\{X_1, \ldots, X_k\}$, of a group of random variables provides a measure of that part of their joint statistical dependence which is not exhibited by the joint behavior of any $k-1$ of them. It is a polynomial in the individual elementary moments (averages of monomials) that is homogeneous of degree k on the underlying X_j. It is the simplest function of the elementary moments that vanishes if any subgroup of the random variables is statistically independent of the remaining members of the group. The cumulant spectrum of order k, provides a $(k-1)$ frequency breakdown of how collections $\{X(t_1), \ldots, X(t_k)\}$ of k values of the process co-vary in a way not exhibited by the behavior of any $k-1$ of them. It has the further characteristic of measuring an important aspect of how near to, or far from, nearly Gaussian the group behaves, since, for Gaussian variates, all cumulants of order greater than 2 vanish.

Applying a nonlinear operation to a series with only second-order dependencies leads to dependencies of order higher than 2. Suppose one has the series

$$X(t) = \alpha_1\cos(\beta_1 t+\phi_1) + \alpha_2\cos(\beta_2 t+\phi_2)$$

with α_1, α_2, β_1, β_2 constants. If ϕ_1 and ϕ_2 are independent and distributed uniformly on $(-\pi,\pi)$ then X will be stationary, have zero mean and have spectral mass at -- and only at -- the frequencies β_1 and β_2. Suppose that the new series

$$Y(t) = X(t) + \gamma X(t)^2 \tag{10.3}$$

is formed. $Y(t)$ will contain the cosinusoid terms of X. It will further include terms in $\cos(2\beta_1 t+2\phi_1)$, $\cos([\beta_1-\beta_2]t + \phi_1-\phi_2)$, $\cos([\beta_1+\beta_2]t + \phi_1+\phi_2)$, $\cos(2\beta_2 t+2\phi_2)$. Notice that certain third-order moments (also cumulants) of such components, such as

$$ave\{\cos(\beta_1 t+\phi_1)\cos(\beta_2 t+\phi_2)\cos([\beta_1+\beta_2]t + \phi_1+\phi_2)\}$$

will not vanish, because of the way in which ϕ_1 and ϕ_2 appear. Thus a (double) Fourier analysis of

$$cum\{Y(t+u), Y(t+v), Y(t)\}$$

will provide an indication that a nonlinearity, such as (10.3), may have occurred.

Spectra of order k, involving cumulants of order k, are particularly pertinent to nonlinearities that are polynomial of order $k-1$. But, just as spectra (of order 2) are very helpful in dealing with non-Gaussian processes, even those not generated linearly, spectra of order k can be helpful in dealing with the consequence of other kinds of nonlinearities than polynomial ones.

If the series under consideration is vector-valued, joint cumulant spectra provide representations for joint cumulant functions. For example the cross-bispectrum, $f_{XYZ}(\omega,\nu)$, satisfies

$$cum\{X(t+u),Y(t+v),Z(t)\}$$

$$=\int\int e^{i(u\omega+v\nu)}f_{XYZ}(\omega,\nu)d\omega d\nu$$

10b. BISPECTRA (INCLUDING CALCULATION)

The simplest higher-order (cumulant) spectrum is the bispectrum. It is concerned with the dependence among triples of frequencies λ,μ,ν summing to zero. Suppose the stationary series X has the spectral representation

$$X(t)=\int e^{it\omega}dZ(\omega)$$

for Z a stochastic measure. (See for example Brillinger (1981).) Then the representation (10.1) indicates that

$$cum\ \{dZ(\omega),dZ(\nu),dZ(\gamma)\}$$

$$=\ \delta(\omega+\nu+\gamma)f(\omega,\nu)d\omega d\nu d\gamma$$

with $\delta(\cdot)$ the Dirac delta function. (For ω,ν,γ all $\neq 0$, the cumulant here may be replaced by *ave*.) This result suggests how one might form an estimate of the bispectrum and a further interpretation for it.

The series $\exp\{it\lambda\}dZ(\lambda)$ may be viewed as the result of narrow bandpass filtering the series X at frequency ω. This fact suggests a means of estimating the bispectrum (perhaps the first estimation technique to be employed historically, see Hasselman *et al.* (1963)). One narrow-band pass filters the series at various frequencies ω, obtaining a collection of series $X(t,\omega)$. Then one averages in time the triple products $X(t,\omega)X(t,\nu)X(t,\gamma)$ and $X(t,\omega)X(t,\nu)X^H(t,\gamma)$ with $\omega+\nu+\gamma=0$, where X^H denotes the Hilbert transform of the series X. Here a negative value of ω refers to filtering at $|\omega|$, since real filters treat $+\omega$ and $-\omega$ alike. The sign of the Hilbert transform must be reversed for $\omega<0$. (If the time average is not to estimate zero, the three narrow-band series must be behaving in a coherent fashion.) The results are estimates of the real and imaginary parts of the bispectrum respectively. It is clear that the bispectrum may be estimated equivalently by time averaging triple products of complex-demodulates.

A substantially different means of estimating the bispectrum is to proceed by averaging a third order periodogram. The third order periodogram of the stretch $X(t),t=0,\ldots,T-1$ is defined as

$$I^T(\omega,\nu)=(2\pi)^{-2}T^{-1}d^T(\omega)d^T(\nu)\overline{d^T(\omega+\nu)}$$

(In the case that the series X does not have zero mean, it is better that the sample mean -- or some more detailed trend -- be subtracted prior to computing d^T.) The expected value of this third order periodogram is given by

$$ave\ I^T(\omega,\nu)$$

$$=\ (2\pi)^{-2}T^{-1}\int\int\Delta^T(\alpha)\Delta^T(\beta)\overline{\Delta^T(\alpha+\beta)}\ S(\omega-\alpha,\nu-\beta)d\alpha\, d\beta$$

and will be near $S(\omega,\nu)$ for T large and S smooth near (ω,ν). Further, periodogram values evaluated at different frequencies are asymptotically independent of each other. This suggests forming the estimate

$$S^T(\omega,\nu) = \sum_{p,q} W^T\left(\omega - \frac{2\pi p}{T}, \nu - \frac{2\pi q}{T}\right) I^T\left(\frac{2\pi p}{T}, \frac{2\pi q}{T}\right)$$

where the weights W^T sum to 1 and are such that the estimate S^T has the periodicity and symmetry properties of s.

Alternately one might proceed by noting that periodogram values, say $I^T(\omega,\nu;\ell)$, $\ell = 1,...,L$ based on disjoint stretches of series are asymptotically independent. This remark leads to the estimate

$$S^T(\omega,\nu) = \sum_{\ell=1}^{L} I^T(\omega,\nu;\ell)/L \quad .$$

It will have the necessary symmetry and periodicity properties directly.

10c. AN EXAMPLE: EQUIPMENT VIBRATION NOISE

The bispectrum is useful in monitoring mechanical drives. When the teeth of two cogwheels are intermeshing cleanly, a plausible model for part of the noise signal generated is

$$X(t) = \alpha \cos(\beta t + \gamma) + \delta \cos(2\beta t + \epsilon)$$
$$+ \phi \cos(3\beta t + \gamma) + \ldots$$

with β corresponding to the speed with which the driving wheel is turning and all parameters constant. If a running estimate of the bispectrum $f(\beta,\beta)$ is computed, the estimate may be expected to fluctuate about the level

$$\alpha^2 \delta e^{i(2\gamma-\epsilon)} T^2/(2\pi)^2$$

with T the length of the time stretch over which an individual estimate is computed.

As the cogs wear irregularly, the phase angles γ,ϵ may be expected to become random and the bispectrum to become 0. That this indeed happens was shown in Sato *et al.* (1977).

A power spectrum (of order 2) is unable to be used as a diagnostic tool here, because the amount of power present at frequency β remains reasonably constant as time passes on. The power spectrum does not provide information on the phase angles of the frequency components of a time series.

10d. HIGHER ANALOGS

The cumulant spectrum of order k is given by expression (10.2). It may be estimated in several fashions. For example, its real part may be estimated by averaging the product

$$X(t,\lambda_1)...X(t,\lambda_k)$$

in time, where $X(t,\lambda)$ is X narrow-band filtered near frequency λ, and $\lambda_1 + \ldots + \lambda_k = 0$, (but no proper subset of the λ's sums to 0). The imaginary part may be estimated by including a Hilbert transform.

Alternatively it may be estimated by averaging the k-th order periodogram in either time or frequency. This periodogram is defined as

$$I^T(\lambda_1,...,\lambda_{k-1}) = (2\pi)^{-k+1} T^{-1} d^T(\lambda_1)...d^T(\lambda_k)$$

where $-\lambda_k = \lambda_1 + \ldots + \lambda_{k-1}$. Details are given in Brillinger and Rosenblatt (1967).

In the case of a vector-valued (multi-component) series, the Fourier values d^T may be based on different components, depending on the question of interest, thus providing estimates of cross-polyspectra.

11. POINT PROCESSES, ETC.

A stochastic point process is a random entity whose hypothetical realizations are doubly infinite sequences $\{\tau_j\}_{j=-\infty}^{\infty}$ of points along the line with $\tau_j \le \tau_{j+1}$. (Actual realizations involve finite sequences.) Examples include: i) the times of earthquakes within a given region, ii) the times at which a neuron fires, iii) the times at which births occur in some population of interest, iv) the times at which customers arrive at a service facility and v) the times of lightning flashes in a thunder storm. It is convenient to set $N(t)$ = the number of τ_j in the interval $[0,t)$, then $N(t)$ is a step function increasing by 1 whenever an event occurs. One has the symbolic representation

$$\frac{dN(t)}{dt} = \sum_j \delta(t-\tau_j)$$

with $\delta(\cdot)$ the Dirac delta function, showing that a point process may be considered a generalized time series.

A marked point process is a random entity whose hypothetical realizations are doubly infinite sequences $\{(\tau_j, M_j)\}_{j=-\infty}^{\infty}$ of pairs (τ_j, M_j) with $\{\tau_j\}_{j=-\infty}$ a point process and with M_j a mark or value attached to the j-th point. It may take on discrete, continuous or symbolic values. Examples include: i) times and magnitudes of earthquakes, ii) arrival times and waiting times experienced by individuals in a queue, iii) birth times and number born (twins, triplets, etc.) and iv) locations of extrema of a random function $X(t)$ and associated extreme values. It is sometimes convenient to employ the symbolic representation

$$\sum_j M_j \delta(t-\tau_j)$$

for a marked point process.

Point processes and marked point processes may be used to provide descriptions for the important class of time series containing bursts of activity every so often, with the principal stochastic variation corresponding to the location and size of the activity. One can model such processes as

$$\sum_j M_j a(t-\tau_j) + \epsilon(t)$$

with $\epsilon(\cdot)$ a noise series and $a(\cdot)$ a fixed (response) function. Physical examples include: shot noise, river run-off (with τ_j the time of a rainstorm, M_j the storm's size and $a(\cdot)$ depending on the area's geography and location of the measuring gauge), and impulse noise interference, as lengthened by receiver characteristics.

As was the case with ordinary time series, situations involving point and marked point processes may be stationary and may involve linear time invariant operations. One is led to problems of spectrum estimation and system identification.

11a. THE SPECTRUM

For a stretch $(N(t), 0 \le t < T)$ of point process data one may compute the Fourier transform

$$d^T(\omega) = \sum_{0 \le \tau_j < T} \exp\{-i\omega\tau_j\}$$

for a frequency ω. The power spectrum of the process may now be defined as

$$S(\omega) = \lim_{T\to\infty} \frac{1}{(2\pi T)}\, ave\, |d^T(\omega)|^2 \quad , \quad \omega \neq 0$$

and by continuity at $\omega = 0$. Were the realization $\{\tau_j\}$ periodic with period τ, so $\tau_{j+1} = \tau_j + \tau$, then $S(\omega)$ would have infinite peaks at frequencies $2\pi k/\tau$, $k = \pm 1, \pm 2, \ldots$.

In the stationary case one often has, (using differential notation),

$$cov\{dN(t+u), dN(t)\} = [p\delta(u) + q(u)]du\ dt$$

with $\delta(\cdot)$ the Dirac delta function, and one can set

$$f(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\omega u}\, cov\{dN(t+u), dN(t)\}/du$$

$$= \frac{p}{2\pi} + \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\omega u} q(u) du \ . \tag{11.1}$$

Here p is the rate, $(Prob\{dN(t)=1\}/dt)$ of the process and $q(u) + p^2$ has the interpretation $Prob\{dN(t+u)=1$ and $dN(t)=1\}/dt/du$ for $\mu \neq 0$.

If one passes the process N through a linear filter with transfer function A to obtain a series X, then the power spectrum of X is given by $|A(\omega)|^2 f(\omega)$, the same relationship as one has for ordinary time series.

In the marked case one may use the definition

$$\lim_{T\to\infty} \frac{1}{2\pi T}\, ave\, |\Sigma M_j \exp\{-i\omega\tau_j\}|^2 \quad , \quad \omega \neq 0 \tag{11.2}$$

for the power at frequency ω.

11b. CROSS-SPECTRA AND COHERENCE

In the case that one has point processes M and N, i.e. $\{\sigma_j\}_{j=-\infty}^{\infty}$ and $\{\tau_k\}_{k=-\infty}^{\infty}$, one may define their cross-spectrum as

$$f_{NM}(\omega) = \lim_{T\to\infty} \frac{1}{2\pi T}\, ave(\Sigma \exp\{-i\omega\tau_j\})(\Sigma \exp\{i\omega\sigma_k\}) \quad , \quad \omega \neq 0$$

where the sums are over j with $0 \le \sigma_j < T$ and k with $0 \le \tau_k < T$. When M and N are statistically independent, $f_{NM}(\omega) = 0$. If N is a lagged version of M, $\tau_j = \sigma_j + \tau$, $f_{NM}(\omega) = \exp\{-i\omega\tau\} S_{MM}(\omega)$. This last suggests an estimation procedure for such a τ.

The cross-spectrum may also be defined as follows: suppose M and N have rates p_M, p_N respectively. Suppose, further,

$$Prob\{N \text{ point in } (t+u, t+u+a) \text{ and } M \text{ point in } (t, t+b)\} \approx p_{NM}(u)ab$$

for small a, b. Then one could set

$$f_{NM}(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} [p_{NM}(u) - p_N p_M] \exp\{-i\omega u\} du \ .$$

These last two expressions suggest that $f_{NM}(\omega)$ is measuring the association of the processes M and N at frequency ω.

A standardized measure of association is provided by the coherence

$$|R(\omega)|^2 = |f_{NM}(\omega)|^2 / f_{NN}(\omega) f_{MM}(\omega) \ .$$

This satisfies $0 \le |R(\omega)|^2 \le 1$ with 0 in the case of statistical independence and 1 in the case of linear time invariant dependence.

11c. AN EXAMPLE: ARE EARTHQUAKES PERIODIC?

Kawasumi (1970) suggested that large earthquakes in the Kamakura region of Japan displayed a period of 69 years. The dates of 33 events from 818 AD to the present are known. Vere-Jones and Ozaki (1982) present the periodogram

$$I^T(\omega) = \frac{1}{2\pi T} |d^T(\omega)|^2$$

of this data. It displays a sharp peak at $\omega = 0.91$ radians/year corresponding to a period of 69 years. However, the periodogram is subject to substantial statistical instability. Vere-Jones and Ozaki develop a formal test of significance for the presence of a periodic effect. Because of the substantial clustering of the data, sampling uncertainty turns out to be large and the cyclic effect turns out not to be significant.

11d. AN EXAMPLE: PARTIAL COHERENCES OF NEURON PROCESSES

Nerve cell spike trains are conveniently dealt with by the techniques of point processes. Various analyses are described in Brillinger, Bryant and Segundo (1976). One of these concerns three neurons $L2$, $L3$ and $L10$ of the seahare *(Aplysia californica)*. The three neurons were clearly related (substantial coherences between all pairs of output spike trains.) It was known that $L10$ was the driving neuron, however it was not known if the neurons were in series $L10 \to L3 \to L2$, or $L10 \to L2 \to L3$ or if $L3$ and $L2$ had no direct connection, but $L10 \to L3$ and $L10 \to L2$ only.

Partial coherence analysis is a useful tool for examining such questions. Denote the spike trains by A, B, C respectively. The partial coherence between trains A and B, is defined to be the coherence, near frequency ω, between the trains A and B with the linear time-invariant effects of C removed. It is given by the modulus-squared of the partial coherency

$$\frac{R_{AB} - R_{AC}R_{CB}}{\sqrt{(1 - R_{BC}R_{CB})(1 - R_{AC}R_{CA})}}$$

where, for convenience, dependence on ω has been suppressed. In the case cited, the partial coherence of $L3$ and $L2$ with the effects of $L10$ removed was not significant and one is able to essentially infer that there is no direct connection from $L2$ to $L3$.

11e. HIGHER-ORDER SPECTRA

If anything, higher-order spectra may be expected to prove more useful with point process data than with ordinary time series, for point processes are far removed from Gaussianity.

The bispectrum of the point process $\{\tau_j\}_{j=-\infty}^{\infty}$ may be defined as

$$S(\omega,\nu) = \lim_{T\to\infty} \frac{1}{(2\pi)^2 T}\, ave\{d^T(\omega)d^T(\nu)\overline{d^T(\omega+\nu)}\},$$

for ω, $\nu \ne 0$ and by continuity otherwise. From this definition it is apparent that $f(\omega,\nu)$ is measuring the co-relationship of the components of frequencies ω, and ν with that of frequency $-\omega-\nu$ in the process (again the co-relationship of three frequencies summing to zero).

Alternatively it may be defined via the Fourier transform of the product density
$p(u,v) = \lim_{a,b,c\downarrow 0} Prob\{\text{point in } (t+u,t+u+a) \text{ and in } (t+v,t+v+b) \text{ and in } (t,t+c)\}/abc$
As such it is seen to relate to the fashion in which triples of points fall in specified patterns.

Suppose for example that the series X results from input of the point process N to a linear time invariant filter with transfer function A,

$$X(t) = \int a(t-u)dN(u)$$

$a(\cdot)$ being the impulse response of the filter. Then one has

$$f_{XXX}(\omega,\nu) = A(\omega)A(\nu)\overline{A(\omega+\nu)}f_{NNN}(\omega,\nu) .$$

If for example N is Poisson of rate p, then $f_{NNN}(\omega) = p/(2\pi)^2$ and so $f_{XXX}(\omega,\nu) = A(\omega)A(\nu)\overline{A(\omega+\nu)}p/(2\pi)^2$. This bispectrum will generally be nonzero.

There are direct extensions of these ideas to spectra of order k and to vector-valued processes. Extensions to marked point processes are also immediate as in expression (11.2).

12. HIGHER THAN UNIDIMENSIONAL "TIME"

These days many data sets involve functions of several variables, $X(t_1,t_2,...,t_K)$. One might have a picture with t_1 and t_2 the length and width coordinates and $X(t_1,t_2)$ the intensity of light at the location (t_1,t_2). One might have $X(t_1,t_2,t)$ with t_1, t_2 the coordinates of the location of leads on a subject's skull, t time and $X(t_1,t_2,t)$ the EEG voltage level recorded at time t at location (t_1,t_2). With the advent of array processors and clever optical processors (see Turpin (1981)) it turns out that one can deal with data of this sort quite effectively on many occasions.

Some new things arise in these circumstances. These include: alternate forms of stationarity, alternate forms of large sample theory, marginalization, and irregular domains of observation.

12a. FORMALISM

Suppose the data $X(t_1,...,t_K)$, $t_k = 0,...,T_k-1$, $k = 1,...,K$ is available. It is often meaningful to form the Fourier transform

$$d^T(\boldsymbol{\omega}) = \sum_{t_1=0}^{T_1-1} \cdots \sum_{t_K=0}^{T_K-1} X(t_1,...,t_K)\exp\{-i(\omega_1 t_1+...+\omega_K t_K)\}$$

with $\boldsymbol{\omega} = (\omega_1,...,\omega_K)$. The amplitude of d^T is seen to be large for $\boldsymbol{\omega}$ near $(\beta_1,...,\beta_k)$ when $X(t_1,...,t_K)$ contains components of the form $\alpha\cos(\beta_1 t_1 + \cdots + \beta_K t_K+\gamma)$. The transform itself is seen to be equivalent to a repeated transform applied once for each of the arguments of X.

The following limit, when it exists, may be defined to be the power spectrum at wave vector $\boldsymbol{\omega}$,

$$\lim_{T_1,...,T_K\to\infty} \left(\prod_{k=1}^{K}\left(\frac{1}{2\pi T_k}\right)\right) ave\,|d^T(\boldsymbol{\omega})|^2$$

One may write

$$|d^T(\boldsymbol{\omega})|^2 = \sum_{\mathbf{u}} e^{-i(\boldsymbol{\omega},\mathbf{u})}\left(\sum_{t_1=0}^{T_1-|u_1|} \cdots \sum_{t_K=0}^{T_K-|u_K|} X(\mathbf{t}+\mathbf{u})X(\mathbf{t})\right)$$

with $\mathbf{t} = (t_1,...,t_K)$, $\mathbf{u} = (u_1,...,u_K)$, $(\boldsymbol{\omega},\mathbf{u}) = \omega_1 u_1 + ... + \omega_K u_K$ suggesting that this spectrum will be appropriate when one has the stationarity condition: $c_{XX}(\mathbf{u}) = cov\{X(\mathbf{t}+\mathbf{u}), X(\mathbf{t})\} = cov\{X(\mathbf{u}), X(\mathbf{0})\}$ for all $\mathbf{t}$. (One reference is Priestley (1981), Chapter 9.)

In some circumstances, the covariance depends on $|\mathbf{u}|^2 = u_1^2 + ... + u_K^2$ alone, when the process is said to be *isotropic*. The coordinates t_k must be on similar scales, before isotropy can possibly be a physically reasonable assumption.

For stationary biresponse data $(X(\mathbf{t}),Y(\mathbf{t}))$ one can define the cross-covariance function

$$c_{XY}(\mathbf{u}) = cov\{X(\mathbf{t}+\mathbf{u}),Y(\mathbf{t})\} , \qquad (12.1)$$

the cross-spectrum

$$f_{XY}(\omega) = (2\pi)^{-K} \sum_{u} \exp\{-i(\omega,u)\}c_{XY}(u) \ , \tag{12.2}$$

and the coherence $|R(\omega)|^2 = |f_{XY}(\omega)|^2/f_{XX}(\omega)f_{YY}(\omega)$. These provide measures of the degree of linear time invariant relationship between the two data processes. In the case of (12.1), (12.2) one has the representation

$$c_{XY}(u) = \int_{-\pi}^{\pi} \cdots \int_{-\pi}^{\pi} \exp\{i(u,\omega)\}f_{XY}(\omega)d\omega_1 \dots d\omega_K \tag{12.3}$$

The generic form of a linear time-invariant filter in multidimensional time has the form

$$Y(t) = \sum_{u} a(t-u)X(u)$$

with $a(\cdot)$ the impulse response and

$$A(\omega) = \sum_{u} a(u)\exp\{-i(\omega,u)\}$$

the transfer function. If the series X has power spectrum $f_{XX}(\omega)$, then that of Y will be $|A(\omega)|^2 f_{XX}(\omega)$, reducing to the usual relationship in the case $K = 1$.

Suppose that some of the "time" coordinates are fixed in a study of the bivariate series $(X(t), Y(t))$. It is of interest to understand the relationship between the spectra of the full series and those for such a time slice. The relationship is apparent from expression (12.3). The fixed coordinates correspond to $u_k = 0$ in (12.3). In consequence the cross-spectrum of the reduced series is seen to have the form

$$\int \cdots \int f_{XY}(\omega)d\omega_j \dots d\omega_K$$

if $\omega_j, \dots, \omega_K$ correspond to the time coordinates held fixed.

12b. SPATIAL PROBLEMS

The case of a spatial array $X(x,y)$, with x,y indicating geographic coordinates, provides an important, yet direct, extension of the unidimensional case. Standing (or frozen) waves $\alpha \cos(\beta x + \gamma y + \delta)$ provide the leading example of arrays demanding the Fourier approach. As this leading example, consider X made up of a finite number of interfering standing (or frozen) waves

$$X(x,y) = \sum_{j} \alpha_j \cos(\beta_j x + \gamma_j y + \delta_j) + \text{noise}$$

We are naturally interested in estimating the wave numbers β_j, γ_j of the individual waves and the corresponding powers $\alpha_j^2/2$. The Fourier transform is a basic tool for doing this. Here the direct transform is given by say

$$\sum_{x=0}^{T-1} \sum_{y=0}^{T-1} X(x,y)\exp\{-i(\lambda x + \mu y)\}$$

$$= \frac{1}{2} \sum_{j} \alpha_j \exp\{i\delta_j\}\Delta^T(\lambda-\beta_j)\Delta^T(\mu-\gamma_j)$$

$$+ \frac{1}{2} \sum_{j} \alpha_j \exp\{-i\delta_j\}\Delta^T(\lambda+\beta_j)\Delta^T(\mu+\gamma_j)$$

with $\Delta^T(\lambda) = \Sigma \exp\{-i\lambda t\}$ as earlier. The amplitude of this transform will peak near the wave vectors (β_j, γ_j) as desired. It is clear however that because of the rippling character of Δ^T there will be leakage across λ and μ. As in the unidimensional case one will need to taper, forming for example

$$d^T(\lambda,\mu) = \sum_{x\,y} h^T(x\,y)X(x\,y)\exp\{-i(\lambda x + \mu y)\}$$

with $h^T(x\,y)$ a smooth function vanishing everywhere outside the domain of observation of the array. One will want its Fourier transform

$$H^T(\lambda,\mu) = \sum_{x\,y} h^T(x\,y)\exp\{-i(\lambda x + \mu y)\}$$

to be concentrated near $(\lambda,\mu) = (0,0)$ and to die off rapidly as $|\lambda|$, $|\mu|$ increase.

In the case that the array X is stationary with zero mean and power spectrum $f(\lambda,\mu)$ one has, for d^T above,

$$\text{ave}|d^T(\lambda,\mu)|^2 = \int\int |H^T(\lambda-\alpha,\mu-\beta)|^2 f(\alpha,\beta)d\alpha d\beta$$

suggesting that one might base an estimate on $|d^T|^2$. It further suggests the essential need for tapering in the case $S(\cdot)$ is not nearly constant.

Suppose next that the continuous array X is isotropic, $c(u,v) = cov\{X(x+u,y+v),X(x\,y)\}$ depending on u^2+v^2 alone. The surprising thing is that the power spectrum of X also depends on $\lambda^2+\mu^2$ alone. Specifically, if $c(u,v) = g(\sqrt{u^2+v^2})$ then

$$f(\lambda,\mu) = \frac{1}{2\pi}\int_0^\infty J_0(r\sqrt{\lambda^2+\mu^2})g(r)dr$$

with J_0 a Bessel function. The transformation here is a Hankel transformation. This result leads to an improvement over the usual 2-dimensional case, in our ability to estimate such an isotropic power spectrum. (See Brillinger (1970)).

12c. MIXED (SPATIAL-TEMPORAL) PROBLEMS

There are a variety of dynamic spatial problems, in which a process of concern may be represented as $X(x\,y\,t)$, where $x\,y$ are spatial coordinates and t is time. For example

$$\alpha\cos(\beta x + \gamma y + \delta t + \epsilon)$$

may be viewed as a wave moving in direction $-\gamma/\beta$ with velocity $|\delta|/\sqrt{\beta^2+\gamma^2}$. The first two coordinates, $x\,y$, here have quite different character from the remaining one, t. (In some situations it may be reasonable to assume the process isotropic in x and y, but with t and $(x^2+y^2|^{1/2}$ arbitrarily involved together.)

If one has but a single time slice, say $t=0$, of the process, then the situation takes the form of the previous section. On the other hand, in many situations one will have much longer records in terms of the t coordinate than for the $x\,y$ coordinates. In terms of asymptotics it may prove reasonable to envisage the largest observed t-value tending to ∞, but this may not be sensible for x and y. (Cases with t and a single x where this is relevant have been considered by Bretherton and McWilliams (1980).)

The Fourier transform that one might compute from such a data set would be

$$d^T(\lambda,\mu,\omega) = \sum_{x\,y\,t} h^T(x\,y\,t)X(x\,y\,t)\exp\{-i(\lambda x + \mu y + \omega t)\}$$

with h^T a tapering function vanishing off the domain of observation. The power spectrum $f(\lambda,\mu,\omega)$ might be estimated by averaging $|d^T|^2$. Real difficulties arise concerning the display of the resulting estimate $f^T(\lambda,\mu,\omega)$ as a function of 3 variables.

The spectrum of a single time slice, say $X(x,y,0)$ is given by

$$\int f(\lambda,\mu,\omega)d\omega$$

i.e. if only a single slice is available, it should be remembered that the spectrum one has estimated is a marginal (not a conditional) version of the complete power spectrum. Important detail in the full spectrum may not be available in such a marginal spectrum.

12d. AN EXAMPLE: SWOP (STEREO WAVE OBSERVATION PROJECT)

At 1700 GMT on October 25, 1954 stereophoto techniques were employed to measure the state of the seas in the Atlantic at about 39°N, 63.5°W. Basically, two pictures of the sea surface were taken simultaneously at known altitudes and distance apart. From these pictures the water heights at selected points were measured. All told 5400 measurements were made, with x and y spacing of 30' over a 2700' by 1800' rectangle. The data was corrected for trends by least squares fitting of a plane. The autocovariance was estimated at each combination of 90 lags in the X-direction and 60 lags in the Y-direction. This was Fourier transformed and the transform smoothed by an extension of the Hamming window. The estimated spectrum showed a main peak corresponding to the locally generated sea and also a small peak corresponding to some swell.

By physical reasoning (that wave energy generated is driven off leeward) the authors (Cote *et al* (1960)) were able to distinguish the direction corresponding to a peak from its antipode (which would have generated the same spectrum).

12e. AN EXAMPLE: MOVING EARTHQUAKE SOURCES

Bolt *et al* (1982) present the results of spatial spectrum analyses of several nearby earthquakes recorded by an array of 27 seismometers arranged on three concentric circles (and at their common center). Spatial spectra are estimated for various time segments of the seismograms and several (temporal) frequencies. Peaks, corresponding to the direction of the earthquake source and the velocity of the waves are found in the spectra.

The location of the peak for one of the events is found to shift with the time segment analyzed. This shift may provide the first experimental measurement of how a seismic dislocation moves along a rupturing fault.

12f. AN EXAMPLE: EVIDENCE FOR SCATTERING OF SEISMIC ENERGY

Spatial spectra have been used to provide strong evidence for backscattering of energy during the passage of seismic waves. Aki and Chonet (1975) present an estimate of the spatial spectrum for the initial group of shear waves (in the frequency band 1.0-2.0 Hertz) arriving after an explosion. The data are the seismograms recorded at the Large Aperature Seismic Array (LASA) in Montana. The estimate shows a substantial peak of energy in the direction of the blast, moving at an appropriate velocity, and little else. Aki and Chonet next present an estimate of the spatial spectrum of the later arriving shear waves (the coda.) This estimate shows energy arriving from all directions with shear and surface wave velocities. Once again spectrum analysis of separate time segments has displayed the presence of an important scientific phenomenon.

E. SYSTEM IDENTIFICATION

We discuss linear system identification in this part.

13. INPUTS AND OUTPUTS

On many occasions time series are subjected, naturally or artificially, to operations. These operations may be physical or computational. Often the result of the transformation is a time series itself. In this case one speaks of the initial series as the input and the transformed series as the output. The collection of {input, operation, output} is referred to as a system. The problem of system identification is that of developing a useful description for the operation given stretches of input and corresponding output.

13a. ONE OF EACH, NO NOISE

In the simplest case a linear time-invariant system has a single input and a single unique output. Examples include:

$$Y(t) = [X(t+1) + X(t) + X(t-1)]/3$$

$$Y(t) = X(t+1) - X(t)$$

$$Y(t) = \sum_{u=0}^{\infty} \rho^u X(t-u) \text{ , with } 0 < \rho < 1 \text{ .}$$

As mentioned in Section 3, such systems are typically characterized by a single function of frequency, the transfer function. If the input series is $X(t) = \exp\{i\omega t\}$, then the corresponding output series is $Y(t) = B(\omega)\exp\{i\omega t\}$, with B the transfer function. For the examples above $B(\omega)$ is

$$[1+2\cos\omega]/3$$

$$\exp\{i\omega\} - 1$$

$$1/[1-\rho\exp\{-i\omega\}]$$

respectively.

The value $B(\omega)$ provides the change in amplitude and phase that the operation effects on the input series at frequency ω. The amplitude $|B(\omega)|$, is called the gain. The angle, $arg\{B(\omega)\}$, is called the phase.

In practice, linear time-invariant systems are often described by providing the form of $B(\omega)$; for example an ideal (unrealizable) band-pass filter at frequency λ with bandwidth Δ is specified by

$$B(\omega) = K \quad \text{for } |\omega-\lambda| < \Delta/2$$

$$= 0 \quad \text{otherwise}$$

for K constant, and $\omega \geq 0$.

13b. ONE OF EACH, NOISE; IMPORTANCE OF COHERENCE

Many of the systems encountered in practice do not have an output completely determined by the input, since noise (including, for example, round-off) enters at some stage. In a simple case, one has a system, B, a noise series, ϵ, and the system output is given by

$$Y(t) = B[X](t) + \epsilon(t) \tag{13.1}$$

or $Y = BX + \epsilon$. The similarity of this last relationship to the traditional one of regression is clear. The transfer function, B, corresponding to the system, is, as noted in 8d, a natural analog of the ordinary linear regression coefficient, b of elementary statistics. It is distinguished by being complex-valued and a function of frequency.

If $Y(t,\omega)$, $X(t,\omega)$, $\epsilon(t,\omega)$ denote the result of narrow band-pass filtering the series Y, X, ϵ at frequency λ respectively and if the series X and ϵ are statistically independent, then the relationship (13.1) leads to

$$E\{Y(t,\omega)|X\} \sim B(\omega)X(t,\omega)$$

and the regression analog is clear. (The narrower the filters employed, the better the approximation.)

The determination of the linear regression coefficient of a variate y on a variate x by the expression $b = cov\{y,x\}/var\ x$ has the natural analog

$$B(\omega) = \frac{cov\{Y(t,\omega),X(t,\omega)\}}{var\ X(t,\omega)} = \frac{f_{YX}(\omega)}{f_{XX}(\omega)} \tag{13.2}$$

in the present circumstances. The measurement of the proportion of the variation of y explained through the (linear) variation of x by the coefficient of determination $[cov\{y,x\}]^2/[var\ x\ var\ y]$ has the natural analog

$$|R(\omega)|^2 = \frac{|f_{YX}(\omega)|^2}{f_{XX}(\omega)f_{YY}(\omega)}$$

with $|R(\omega)|^2$ the coherence of the series Y with the series X at frequency ω. $|R(\omega)|^2$ is here a measure of the degree of linear time invariant relationship between the series Y and the series X at frequency ω. Specifically, the power spectrum of the series that is the difference between the series Y and the best (mean squared error) fit to Y by the output of linear time invariant system with input X, is $[1-|R(\omega)|^2]f_{YY}(\omega)$.

$|R(\omega)|^2$ also plays an important role in the formulas for the sampling fluctuations of the results of a system identification via a cross-spectral analysis based on the identity (13.2). If

$$B^T(\omega) = f^T_{YX}(\omega)/f^T_{XX}(\omega)$$

and the spectral estimates are each the average of m periodogram values, then

$$var\ \hat{B}(\omega) \sim \frac{1}{m}[1-|R(\omega)|^2]f_{YY}(\omega)/f_{XX}(\omega)\ . \tag{13.3}$$

For small variance one wants both $|R|^2$ near 1 and m large. With the data at hand one cannot control $|R|^2$, but one can control m.

13c. MORE THAN ONE

In many practical situations, a system under study has more than one input and more than one output. The input series, **X**, and the output series, **Y**, are vector-valued. In this case the transfer function; **B**(ω), of the system is matrix-valued. The time series analog of the traditional multiple regression model

$$y = b_1x_1 + \cdots + b_px_p + \epsilon$$

is

$$Y(t) = B_1[X_1](t) + \cdots + B_p[X_p](t) + \epsilon(t)$$

with the B_p single-input single-output systems and with ϵ a noise series. The vector **B**(ω) may be estimated by

$$\mathbf{B}^T(\omega) = \mathbf{f}^T_{YX}(\omega)\mathbf{f}^T_{XX}(\omega)^{-1}$$

once estimates of the needed power and cross-spectra have been formed.

The multiple coherence may be defined as $|R(\omega)|^2 = 1 - f_{\epsilon\epsilon}(\omega)/f_{YY}(\omega)$ in this case. It proves a useful parameter in practice.

14. CHOICE OF INPUTS

Expressions (13.2) and (13.3) are most helpful in addressing questions concerning the choice of an input series. From expression (13.2) one sees that one wants $s_{xx}(\omega) \neq 0$ at all frequencies ω at which an estimate of $B(\omega)$ is desired. When an estimate of the impulse response function $b(t)$, corresponding to B, is wanted, the input must be rich in all frequencies. Expression (13.3) gives indications of how one should choose the input, if this is in fact a possibility, in order to have a handle on the sampling fluctuations of the estimate $B^T(\omega)$. It makes it apparent that one wants an input with the input power at frequency ω, $f_{xx}^T(\omega)$, high.

14a. QUASI-GAUSSIAN INPUTS

By a quasi-Gaussian process is meant one that shares the notable Gaussian properties of different realizations not looking at all identical, of not being repetitive and of recognizable events not dominating realizations. Natural inputs have these characteristics on many occasions. In one important case the series X, is a series of independent identically distributed random variables. Then one has $f_{XX} = \sigma^2/2\pi$, where σ^2 is the variance of $X(t)$. From expression (13.3), one sees that it is desirable to have σ^2 large. (In practice σ^2 can not be taken to be arbitrarily large for most systems.) In this situation the impulse response, b, may be estimated directly via

$$cov\{Y(t+u), X(t)\} = b(u)\ var\ X(t) \tag{14.1}$$

for all t, u.

There are results (Levin (1960), Mehra (1974), Brillinger (1981), page 220) suggesting that, if a single input is to be used, one wants to arrange for $f_{XX}(\omega)$ to be nearly constant, if possible, in order to have efficient estimates. If the input must be bounded, $|X(t)| \le C$, for some finite C, then it will be most effective, in terms of large-sample variance, to take $X(t) = \pm C$, randomly. If an instantaneous nonlinearity is present, or if quadratic and higher-order terms appear in the output, it will be useful to take X to be Gaussian (Wiener (1958), Brillinger (1977)).

14b. AN EXAMPLE: SPEED SPECTRUM RADAR

The notion of generating a noise like input signal to identify a system of interest occurred amazingly early on. In 1938 G. Guanella filed a patent for a noise-correlation radar in which the radiated signal is a noise-modulated carrier wave. The returning signal is demodulated and, following (14.1), cross-correlated with the transmitted noise to estimate the impulse response.

J. Wiesner had a similar idea in the early fifties. Some of the history is given in Scholz (1982).

14c. SINUSOIDAL INPUTS

A direct method of determining $B(\omega)$, the transfer function at frequency ω, for the system B is to take $X(t) = \cos \omega t$ as input. The output is then

$$\frac{1}{2}B(\omega)e^{i\omega t} + \frac{1}{2}B(-\omega)e^{-i\omega t} = |B(\omega)| \cos(\omega t + \phi(\omega))$$

where $\phi(\omega) = arg\{B(\omega)\}$ and $|B(\omega)|$ and $\phi(\omega)$ may be determined directly. In the case that the system output is perturbed by noise, one will regress the output on the input to estimate $B(\omega)$.

This procedure has the substantial disadvantage of leading to an estimate of B only at a single frequency ω at a time. It will be a more precise estimate than that determined by cross-spectral analysis (having variance inversely proportional to the record length rather than the number of periodograms averaged); however it gives no information on the whole function B.

An improvement results from taking

$$X(t) = \rho_1 \cos(\omega_1 t + \phi_1) + \cdots + \rho_Q \cos(\omega_Q t + \phi_Q),$$

however the positive integer Q must be large before the course of the function $B(\omega)$ can be well determined. If the ϕ_q here are random, then X will be approximately Gaussian for large Q and one returns to the input of the previous section.

14d. PULSE PROBING

If $X(t) = K$ for $t = 0$ and $X(t) = 0$ otherwise, then the output of the linear time-invariant system B is proportional to its impulse response function $b(t)$. If the system output is corrupted by noise one will have to input a sequence of pulses, waiting for the effect of each pulse to die off. By stacking and averaging the various transient responses one may estimate the impulse response $b(t)$. The transfer function $B(\omega)$ is the Fourier transform of $b(t)$. It may be estimated by Fourier transforming $\hat{b}(t)$, employing convergence factors in the process.

In practice one will take $|K|$ to be as large as possible.

One difficulty with this approach is in knowing when the transient has died off. Another is that this pulse form of input may in fact be far removed from the type of input the box usually experiences, so that the box may be thrown into an abnormal region of operation.

14e. COMPARISON AND COMBINATION

The inputs described above have quite different characters and are useful in different types of situations. If little is known about the system, then Gaussian input has substantial advantages. If a few frequencies are of particular importance (or if one is looking for possible nonlinearities) then an input that is the sum of a few sinusoids is useful. If the parameter of greatest interest is the impulse response, and if the system naturally operates, at least sometimes, on pulses, then pulse probing may be effective. If the acceptable size of the input is limited, $|X(t)| \le C$, then pseudo-random-binary-noise may be a good input to employ.

If the system is in fact superposable (linear) then one might use an input series that is a hybrid of the three types described above.

For situations in which one wishes to have a large output, concentrated in time, as opposed to spectrum analysis, it is effective to use an input resembling (in time) the time-reversed impulse response. This occurs in radar and seismic exploration, for example.

14f. DIFFICULTIES WITH NATURAL INPUTS

The best circumstance to be in is that of being able to input any chosen series (design an experiment). There are however many important situations wherein the input is not under the researcher's control. As happens in the case of multiple regression, this leads to real difficulties. In particular, near collinearity in the input leads to estimates with substantial variance.

There are difficulties of interpretation of the individual values of the impulse response function, and of the various individual transfer functions in the multiple input case. If $f_{XX}(\omega) \sim 0$ for ω near λ, then one will not be able to estimate $B(\lambda)$ or $|R(\lambda)|^2$ in any useful sense.

Further, as the analysis is correlational in character, one will not be able to make inferences concerning directions of causation and the like. If one is trying to understand a phenomenon, this can prove a substantial difficulty.

Briefly, all the woes of regression (see Mosteller and Tukey (1977), Chapter 13) may be expected to arise and some new ones as well.

15. IMPORTANT DIFFICULTIES

The previous section mentioned some difficulties in the interpretation of the results of a cross-spectral analysis. There are important technical difficulties as well.

15a. FEEDBACK

Suppose that the system under study is described by:

$$Y = B[X + U]$$

$$U = C[Y]$$

where U is fed back to enter on a par with X. The transfer function from X to Y is

$$\frac{B}{1-BC} \sim -\frac{1}{C} \text{ (when } B \text{ is large)}$$

cross-spectral analysis can estimate this transfer function, but what we learn may refer almost entirely to the feedback path.

In practice, such a system must be dealt with by inserting some observable noise into the feedback loop or, perhaps, by modeling the system in greater detail (e.g. the filters are to be realizable.) Cross-spectral analysis alone is not sufficient to observe the inner workings of a system. Two references are Akaike (1967) and Priestley (1969).

15b. ERRORS INCLUDED IN INPUTS

Suppose, next, that one has the system

$$V = B[U]$$

$$Y = V + \epsilon$$

$$X = U + \eta$$

with ϵ, η noises, and records of X, Y alone available. (These equations may be seen to have the form of the Kalman state-space system. B is not necessarily realizable here though.) One now has the problem of measurement error in the input series. Assuming ϵ and η orthogonal to each other and to U, V, one has

$$f_{YX}(\omega) = f_{VU}(\omega) = B(\omega) s_{UU}(\omega)$$

$$f_{XX}(\omega) = f_{UU}(\omega) + f_{NN}(\omega) .$$

or

$$\frac{f_{YX}(\omega)}{f_{XX}(\omega)} = B(\omega) \cdot \frac{f_{UU}(\omega)}{f_{UU}(\omega) + f_{NN}(\omega)}$$

Determining an estimate of $f_{YX}(\omega)/f_{XX}(\omega)$ will not provide a reasonable estimate of $B(\omega)$, since the measurement error causes attenuation.

One elegant means of handling this situation, which sometimes works, is via an instrumental series. This is an observable series, W, that is orthogonal to the noises ϵ, η but not to U and V. One then has

$$f_{YW}(\omega) = f_{VW}(\omega) = B(\omega) f_{UW}(\omega)$$

$$f_{XW}(\omega) = f_{UW}(\omega)$$

leading to a consideration of $f_{YW}^T(\omega)/f_{XW}^T(\omega)$ as an estimate of $B(\omega)$. A physical example of this procedure will be presented in Section 16c below.

15c. ALIASING

Assume that the series $X(t)$, $Y(t)$ are defined for continuous time, $-\infty < t < \infty$, and have second-order spectra $\bar{f}_{XX}(\omega)$, $\bar{f}_{YY}(\omega)$, $\bar{f}_{YX}(\omega)$ respectively. Suppose that the sampled series are available with $t = 0, \Delta, 2\Delta, \ldots$ Then one has relationships such as

$$f_{YX}(\omega) = \sum_k \bar{f}_{YX}\left(\omega - \frac{2\pi k}{\Delta}\right)$$

for $-\pi/\Delta < \omega \le \pi/\Delta$. It may be that $\bar{f}_{YX}(\omega)/\bar{f}_{XX}(\omega)$ is the desired transfer function; however, this will not be given by $B(\omega) = f_{YX}(\omega)/f_{XX}(\omega)$ generally. Nor will the coherence $|\bar{f}_{YX}(\omega)|^2/[\bar{f}_{XX}(\omega)\bar{f}_{YY}(\omega)]$ generally by given by $|R(\omega)|^2 = |f_{YX}(\omega)|^2/[f_{XX}(\omega)f_{YY}(\omega)]$.

In the case of the coherence, aliasing causes the power spectra to be biased upwards; however the cross-spectra fluctuate in sign and so, on many occasions, the coherence of the sampled series will tend to be less than that of the original series.

In practice one will want to filter the series prior to sampling and to select a small enough value of Δ to reduce the effects of aliasing.

15d. THE EFFECTS OF BIAS

Tapering and prefiltering are essential if one is to be able to deal with power spectra that are rapidly falling (in ω) or have line components. This is likewise the situation in the system-identification case. The expected value of $f^T_{YX}(\omega)$ commonly has the approximate form,

$$\int W_1(\omega-\alpha)f_{YX}(\alpha)d\alpha = \int W_1(\omega-\alpha)B(\alpha)f_{XX}(\alpha)d\alpha$$

for some weight function W_1. Likewise

$$ave(f^T_{XX}(\omega)) \sim \int W_0(\omega-\alpha)f_{XX}(\alpha)d\alpha \ .$$

For $f^T_{YX}(\omega)/f^T_{XX}(\omega)$ to be near $B(\omega)$ one will want the windows W_0 and W_1 to be concentrated near 0. This desire may be addressed by tapering the individual records and by choice of the window applied to the quadratic in the observations. It is further clear from the above expressions, the more nearly constant are the functions f_{YX}, f_{XX}, the less bias may be expected. In consequence, when possible one should prefilter the series X and Y to make the second-order spectra more nearly constant. If this can be done with the same filter for X and Y, one obtains the transfer function directly, no compensation for prefiltering is required. (In particular, it helps to make the phase of the cross-spectrum more nearly constant.)

If, for example, $B(\omega)$ dies off rapidly with ω, one might detect this occurrence through having employed a powerful tapering operation, e.g. using a prolate spheroidal function (see Thomson (1977), for example.) Suspicions having been aroused, the higher frequencies in the series Y might be emphasized by suitably prefiltering it, prior to the computation of a further spectral estimate. The effect of this filtering may then be "cancelled out" by dividing its transfer function into the transfer function estimated by cross-spectral analysis.

16. EXAMPLES

Cross-spectral analysis has, by now, been used in many many system identification studies. The concern of the studies have sometimes been the determination of a precise estimate of the transfer function, sometimes simply to see if the series X does in fact affect the series Y, sometimes in a search for interesting phenomena, and sometimes to design control systems.

16a. AN EXAMPLE: FEEDBACK IN A NEUROSENSORY SYSTEM

Marmarelis and Marmarelis (1978), page 128, describe an experiment on the retinal system of a catfish with input a light stimulus and output intracellular response of a cell. There is evidence that for this system feedback occurs at high mean light levels, while the feedback loop is inactive at low mean levels. Hence the system can be studied both open and closed loop.

The system equations, with feedback, are, as we saw in Section 15a

$$Y = B[X+U]$$

$$U = C[Y]$$

with U denoting the feedback component. With the feedback loop active, the transfer function, from X to Y, is, again, $B/(1-BC)$. This may be estimated by cross-spectral analysis. Further, B itself may be estimated by running the experiment in open-loop conditions. Hence the feedback transfer function C may be estimated by combining the results of the two experiments.

16b. AN EXAMPLE: INSTRUMENTAL SERIES IN MAGNETOTELLURICS

In magnetotellurics, simultaneous measurements are made of magnetic and electric fields at locations of interest. Of interest is the complex impedance (a transfer function) relating these. Unfortunately the time series involved are all subject to substantial measurement error. Workers in the field found that transfer function estimates determined by direct cross-spectral analysis were unstable, varying substantially with repetitions of experiments.

Gamble et al. (1979) took measurements made at a second location as instrumental series and were able to obtain an acceptable measurement of the impedance function of interest (see Section 15b).

16c. AN EXAMPLE: ON-LINE SPECTRAL CONTROL

Rotating circular discs are widely used basic elements in many different machines such as circular saws and computer memory discs. Large vibration amplitude of a rotating disc, due to transverse instability, can cause inaccurate cutting in a circular saw and headtracking error in a computer disc drive. Rahimi (1982) considers the case of a circular saw in particular. He notes that such a saw in a given environment has certain natural frequencies, that the blade stability depends on the relation of those frequencies to certain critical frequencies and that the natural frequencies may be shifted by applying heat (using say an infrared heat lamp.) Work load variation can have a saw operating with natural frequencies near the critical frequencies, leading to instability. Control is affected as follows: blade transverse displacement is measured, current blade natural frequencies are estimated, these are compared to the critical values, and shifted by heating (or cooling) as necessary. This is all done in real-time making use of a digital computer. Rahimi compares several methods of natural frequency spectrum estimation, including direct Fourier estimates, estimates based on finite parameter models and estimates obtained by inputting observable noise to the system in the manner of 15a. The control algorithm continually attempts to reduce the difference between an optimum frequency and a frequency based on the natural-frequencies. In an example, Rahimi finds that control reduces vibration amplitude to 10% of the uncontrolled value.

F. LEADING CASES IN DATA ANALYSIS

17. LEADING-CASE PHILOSOPHY

Formal statistical theory, like classical mechanics, works with leading cases or leading situations (where these words are used as they are in the law). No one complains about using the mechanics of point masses to describe the motion of planets around the sun, although few of us would think either the Earth or Jupiter was only a point. Yet similar oversimplifications in statistical theory tend to be discussed as "failure of assumptions".

It is not easy to be sure why this difference arose and was maintained. Perhaps because statistics must deal with uncertainty -- often with uncomfortably large uncertainties -- its theorists and teachers felt a stronger urge for certainty. Perhaps because stochastic (probabilistic) assumptions are always harder to check -- since their consequences are more subtle than functional (deterministic) ones. Perhaps because statistical theory deals with methods of analysis -- and it may be very difficult to honestly adjudicate among methods of analysis that would lead to appreciably different results in controversial instances. Perhaps because of a felt need for a coherent body of deductive theory that could be lectured about -- by voice or written word.

Whatever the reason, we have grown far too used to the dangerous ideas that:

1. methods should have -- more realistically should *appear* to have -- "derivations"

2. what happens when the conventional assumptions do not hold does not deserve much attention,
3. a method without an optimality theorem is a poor relation.

Yet none of these really apply in practice. If statistical ideas are to carry out their real function, which is to guide, support, and broaden the analysis of data, we must look at them quite differently.

17a. LEADING SITUATIONS, YES: HYPOTHESES, NO

Statistical techniques come into being in diverse ways, -- sometimes, even, by "derivation". They earn most of their legitimacy from some limited knowledge of how they perform in special cases. Sometimes this knowledge is quite weak; we may understand only that they give appropriate answers "on the average". Occasionally, it is quite strong; we may know that within a reasonably restricted family of procedures and according to an apparently reasonable criterion, this procedure performs "best", at least in very narrow circumstances. (The last is more the exception than the norm.)

In reality, then, the best we can do is to start with a leading situation, or even with three leading situations, and learn what we can about a procedure's performance -- or the comparative performance of several procedures -- in each of these. We do much better in understanding and comparing procedures, not by learning more details about performance in the old leading situations, but rather by learning something about their performance in new leading situations.

As George Kimball (Kimball 1958) once put it in an operations research context "There is a further difficulty with the finding of "best" solutions. All too frequently when a 'best' solution to a problem has been found, someone comes along and finds a still better solution simply by pointing out the existence of a hitherto unsuspected variable. In my experience when a moderately good solution to a problem has been found, it is seldom worthwhile to spend much time trying to convert this into the 'best' solution. The time is much better spent in real research in trying to find the variables which have been overlooked" (or "in trying to find a still better solution simply by considering some more general leading case" (cp. Tukey 1961)). As one of us put it recently (Tukey, 1980b) "In practice, *methodologies* have no assumptions and deliver no *certainties*." They have no assumptions, only situations where they perform better and others where they perform less well.

17b. WHAT THEN?

If our techniques have no hypotheses, what then do they have? How is our understanding of their behavior to be described?

As a generalization of an umbra within a penumbra. Here there are at least 3 successively larger regions, namely:

1. An inner core of proven quality (usually quite unrealistically narrow) where we can prove -- and have proved -- certain properties either by formula-manipulation mathematics, or by numerical and experimental sampling mathematics -- or by both together.
2. A middle-sized region of understanding, where we have a reasonable grasp on our technique's performance, where we know generally what is going on, but where we may not (yet?) be able to be sure of details.
3. A third region, often very much larger than the other two, in which the technique will be used. (Sometimes this region depends on the individual concerned, sometimes there is reasonable agreement among a community of intercommunicating users with experience.)

We have to anticipate three distinct regions, often of very different size, for any technique of data analysis, be it statistical or not.

Since the techniques discussed in this account are directed toward an essentially statistical problem, we must try to understand them in terms of such regions, rather than in terms of narrow hypotheses and associated derivations.

17c. SPECTRUM DEFINITION AND ESTIMATION; IN KNOWLEDGE, IN UNDERSTANDING AND IN USE

Let us apply these ideas to the notion of a spectrum.

What are the innermost leading situations, where knowledge is clear? Surely

- individual realizations that are exactly sums of a few cosines, where all would be clear that contributions to the variance (power, energy) are confined to these few (angular) frequencies,
- processes all of whose realizations are of such a form, involving, in total, only a few (angular) frequencies for all the realizations of a single process.

Here no reasonable (trained) person can be uncertain what a spectrum means.

And no one can give a real-world example that really fits into such a picture. The tides come close, but storms perturb the cosine series of simple theory (and tidal forces slowly change the moon's distance). The frequencies that appear in the changes in the eccentricity, etc. of the earth's orbit, (Milankovitch 1930, 1941) are a better example, but there are irregular perturbations here, too.) In the characterization of laboratory standard oscillations (cesium, etc.), emphasis has simply changed from the study of the gross spectrum to the study of the spectrum of the frequency fluctuations.

This narrow region of complete and inevitable understanding of a spectrum is embedded in a larger leading situation, consisting of processes for which there is a clear definition of what a process spectrum is -- and thus, it would seem, -- of at what a spectrum for a realization should be pointing. (We return in a moment to some questions about this.)

This larger leading situation will, no doubt, be one made up of "stationary" processes with finite variance. This will not be because the region of use of the concept of a spectrum is confined to stationary processes, but only because stationary processes are easier to think about.

For many purposes, indeed, we think about a reduced leading situation where we have not only finite variance and stationarity, but also a continuous spectrum. (In the real world, noise and perturbations always seem to make any spectrum continuous). So we often work with a leading situation which does not include those instances with which we started (only a few frequencies) and for which we best understand the notion of a spectrum. This perhaps seems paradoxical, but is realistic and common in dealing with other concepts and other procedures of data analysis.

By going to this intermediate leading case, we have included instances which deserve more thought. A process, each of whose realizations is of the form $X(t) = \alpha \cos(\omega t + \phi)$, for some α, ω and ϕ depending on the realization, with ϕ both independent of α and ω and uniform on $(0, 2\pi)$ is surely stationary, and can have a smooth continuous spectrum. If we have only one realization, we only know about one ω, and -- unless we have a lot of side information -- a reasonable spectrum estimate will put all of the estimated spectrum at -- or very close to -- a single ω_0. In this case, the process can be thought of as a mixture of subprocesses, one for each ω_0, and estimation from one realization leads only to an estimate for the subprocess. Without side information we cannot estimate things to whose consequences we have not been exposed!

If we leave aside such issues -- always important in principle, and sometimes not negligible in practice -- we can say that we understand a reasonable amount about the concept of a spectrum and about spectrum estimates for stationary processes with continuous spectra. As a result, it is only to be expected, that we will *use* the concept, and *use* our understanding of spectrum estimates far outside this region of understanding, often helpfully. We will use them, for example, (a) in non-stationary situations, both those where non-stationarity is a mild imitation and those where non-stationarity is reason why we can learn anything helpful (e.g. Munk and Snodgrass 1957, see Section 5b) and (b) even in analyzing time series which we find hard (or perhaps impossible) to consider as realizations of clearly conceptualized processes. (In the simpler cases, having a process would require an ensemble of parallel worlds.)

Growth from a seedling leading case, where all is clear, to the domain of understanding and the domain of use has here, as so often elsewhere, involved both:

1. growth by a wider definition -- or a reformulation -- of what is estimated, and
2. growth by using both a concept and a procedure in wider areas than those in which they have been precisely or formally defined or justified.

17d. AND ELSEWHERE?

The discussion of the last section was naturally focused on the concept of spectrum and on procedures of spectrum estimation. They are here for two reasons: (a) because this concept and these procedures are vital to this entire account, and (b) as an example of how, in a statistical context, we have to think, to a greater or lesser degree, about *all* the concepts and *all* the procedures that arise. The sorts of spectrum estimation we are discussing are all oriented toward essentially noise-like data, and have to be thought about statistically. Similar pictures involving:

1. inner situations where all is clear,
2. middle-sized situations where we have reasonable understanding (and which may *not* include the inner situations), and
3. still larger situations where we use the procedures and concepts in hope of help (usually successfully)

are to be expected for all the other concepts and procedures we discuss.

17e. THE SPECTRUM AS A VAGUE CONCEPT

Most important statistical ideas, as well as those in many other fields, are centered in vague concepts (e.g. Mosteller and Tukey 1977, pages 17ff). Their expression in precise terms is always to be tested against their vague expression, with any discrepancy to be resolved by *re*defining the *precise* version.

It follows that we are not likely to find a single precise formulation that applies in complete generality, though we may have a sequence (or tree) of different precise forms which work satisfactorily over broader and broader classes of situations.

We all know, as noted above, what a spectrum is for an ensemble or process, each of whose realizations is a finite sum of cosinusoids of a few frequencies, common to all. The special case where 50% of the realization are pure 60 Hertz, 49% are pure 400 Hertz, and 1% are pure 23×60 hertz may, however, bother some of us.

If our process, now relatively general, can -- conceptually at least -- be extended to $-\infty < t < \infty$ and its second moments are "stationary," then there is no conceptual difficulty in defining its spectrum, although no one realization may even try to tell us about the whole of the spectrum (as just illustrated for a very simplified case). Moreover, since actual realizations will be finite, there will in practice always be limited frequency resolution, with sufficiently close frequencies being effectively indistinguishable. (This is clear for "noise-like" data; would be false for perfectly measured "signal-like" data, something which will never be obtained; and is true again, possibly only for much finer resolution, for "signal-like" data measured with error, as all real data is.)

In terms of data, then, we cannot sensibly define "the spectrum" in arbitrary detail. We adhere to the controlling vague concept when we think of -- and work with -- averages of an ideal, untouchable spectrum over finite frequency ranges.

17f. STATIONARITY.

To go further along these lines, we have to think more deeply about "stationarity" than is customary. It long ago became clear that any solid definition of stationarity had to apply to processes, not realizations. But it is still too common to point to a trace and say that it "is obviously not stationary."

All we know of the world is consistent with the idea that all events are periodic with period 10^{20} years. And a process made up of all displacements of a periodic phenomenon -- with uniform probability -- is a stationary process. Thus anything we find in the world could, with this definition, come from a stationary process.

When spectrum ideas are used, we often have only a few realizations, perhaps only one. Both tightness of inference and the instincts of experimenters and theorists alike lead us to prefer working with a more specialized process, from which our realization(s) might have come, rather than a more generalized one. Now, the more specialized one may fail to be stationary, and we ought to ask both (a) is it too narrow?, and (b) even if it is not stationary, may it not be worthwhile to extend the idea of a spectrum -- as always, averaged over suitably narrow frequency regions -- to it? The example of Munk and Snodgrass (1957), quoted above (section 5b) shows that the answer to the second question can be "definitely yes!". So stationarity is not a necessity for useful spectrum analysis.

To many, "it's not stationary" is likely to mean "it behaves as though the level (as hinted at by local means, perhaps) is not constant." A systematic trend could be a property of the relevant ensemble, or merely the casual result of a relatively large amount of energy at very low frequencies. For the practice of analysis, these cases are likely to be indistinguishable. It is likely to be helpful, if any other band of frequencies is to be examined, to eliminate the "trend" -- either by filtering the data to largely remove only low frequencies, or subtracting a suitable slowly changing function from it -- before proceeding to the main analysis.

If the variation about a time-varying process mean is stationary, there is no formal difficulty in defining a spectrum, and no new practical difficulty in estimating the spectrum -- except for very low frequencies where interpretation might well be at best difficult. And, as we saw above (Munk and Snodgrass, (1957)) even this much of stationarity may not be needed for useful results.

Apparent lack of stationarity in the data is NOT a reason to give up either the idea of a spectrum, or the hope of useful results in estimating (averages over) that spectrum. It is a reason to be careful about computational practice (else leakage may bury the information that might have helped you).

17g. CONFIRMATORY OR EXPLORATORY?

Today, statisticians recognize a clear distinction between exploratory and confirmatory data analysis. In the former analysis, our first and principal aim is to see what the data is saying -- though we may occasionally want to look at a rough standard error (is there any other kind) for general guidance. In the latter we are really trying to confirm -- or disconfirm -- a previously identified indication, hopefully doing this on fresh data. Formal statistics books emphasize the latter. Intelligent statistical practice is heavy on the former.

In our experience, almost all spectrum analysis is exploratory. There will be occasions where some formal calculations appeared to be needed to deal with "doubting Thomases" (of one sign or the other), but these are really the exceptions that prove the rule. In this spectrum analysis is like the analysis of variance.

G. SOME STATISTICAL TECHNIQUES

18. FOLLOW-UP WITH NON-LINEAR LEAST SQUARES

There are a number of interesting scientific situations in which physical theory leads to a model in which a few specific parameters play an essential role. As an example, consider the Chandler wobble. Let $X(t)$, $Y(t)$ denote the location at time t of the Earth's axis of rotation as it cut's the Earth's surface near the north pole. Let $Z(t) = X(t) + iY(t)$. Then Munk and MacDonald (1960), using classical mechanics, deduce the equation of motion

$$\frac{dZ(t)}{dt} = \alpha Z(t) + e(t)$$

with $e(t)$ the excitation process and $\alpha = -\beta + i\gamma$. Here β and γ are of special interest. β determines the rate of damping of an excitation and γ the frequency of oscillation. In practice one would like estimates of β and γ, together with estimated standard errors to go with those estimates.

18a. COSINUSOIDS IN NOISE

Physical theory, complex demodulation or a periodogram analysis may suggest that a pure cosinusoid component is present in a time series of interest. A formal model that could be considered then is

$$X(t) = \alpha \cos(\beta t + \gamma) + e(t) \tag{18.1}$$

where $e(t)$ is a stationary series with power spectrum $f_{ee}(\omega)$. It is of interest to develop specific estimates of the parameters α, β, γ.

As has been mentioned earlier, the periodogram of the series X may be expected to have a peak in the neighborhood of the frequency $\omega = \beta$. This suggests the consideration of the value, $\hat{\beta}$, that maximizes the periodogram $I^T(\omega)$ as an estimate of β. Indeed Schuster (1894) proposed this estimate many years ago. Whittle (1952) developed the large sample distribution of $\hat{\beta}$. He found that $T^{3/2}(\hat{\beta}-\beta)$ was asymptotically normal with mean 0 and variance $48\,\pi f_{ee}(\beta)/\alpha^2$. A variance decreasing as T^{-3} is unusual in statistics. That the variance is here proportional to the error spectrum, and inversely proportional to the cosinusoid's amplitude-squared, is not surprising.

Alternatively one might consider an estimate obtained by minimizing the sum of squares

$$\sum_{t=0}^{T-1} |X(t) - \alpha \cos(\beta t + \gamma)|^2 \tag{18.2}$$

with respect to α, β, γ. This estimate turns out to have the same asymptotic distribution as that of Schuster's method.

It is instructive to consider the form that expression (18.1) takes in terms of frequency domain quantities. By Parseval's formula, it may be written

$$T^{-1} \sum_{u=0}^{T-1} \left| d^T\left(\frac{2\pi u}{T}\right) - \alpha e^{i\gamma} \Delta^T \left(\frac{2\pi u}{T} - \beta\right)/2 - \alpha e^{-i\gamma} \Delta^T \left(\frac{2\pi u}{T} + \beta\right)/2 \right|^2 . \tag{18.2}$$

The terms in Δ^T will have their largest magnitude for $2\pi u/T$ near $\pm\,\beta$. Expression (18.2) makes one wonder if improved estimates might not be obtained by weighting the terms in it differentially. Hannan (1971) shows that nothing is gained, for large T, by doing this. It may be worthwhile tapering the data prior to forming the FT in order to improve frequency resolution, however.

A minor extension of the model (18.1) is

$$X(t) = \sum_{k=1}^{K} \alpha_k \cos(\beta_k t + \gamma_k) + e(t) .$$

One is now led to estimate the β_k by the locations of the K greatest relative maxima of the periodogram or by least squares. The estimates of β_k turn out to be asymptotically independent and to have the same asymptotic distribution as in the case $K = 1$. Another reference is Priestley (1962).

18b. DAMPED COSINUSOIDS IN NOISE

In a variety of circumstances, physical considerations lead to the model

$$X(t) = \sum_{k=1}^{K} \alpha_k \cos(\beta_k t + \gamma_k) \exp\{-\delta_k t\} + e(t)$$

$t \ge 0$, i.e. frequencies β_k are present, (after time 0), however their strength decays as time passes. As indicated in Section 7e, complex demodulation is a potent tool for detecting such decaying oscillations and obtaining estimates of the β_k.

One could leap directly to a consideration of least squares estimates; however, it is perhaps more instructive to proceed as follows.

The FT of the noise,

$$d_e^T(\omega) = \sum_{t=0}^{T-1} e(t) \exp\{-i\omega t\}$$

from what has gone before, may be expected to be distributed approximately as a complex Gaussian with mean 0 and variance $2\pi T f_{ee}(\omega)$. Furthermore, estimates at different frequencies may be expected to be approximately independent. It follows that the distribution of a number of FT values, say $d_X^T(\omega_j)$ for ω_j near β_k, may be approximated by independent complex Gaussians with mean

$$\alpha_k \exp\{i\gamma_k\}\Delta^T(\omega_j - \beta_k + i\delta_k)/2$$

and variance $2\pi T\ f_{ee}(\beta_k)$. The negative log-likelihood of these variates is

$$\text{constant} \ + \sum_j |d_X^T(\omega_j) - \alpha_k \exp\{i\gamma_k\}\Delta^T(\omega_j - \beta_k + i\delta_k)/2|^2/2\pi T\ f_{ee}(\beta_k)$$

with the summation over j such that ω_j is near β_k. (One might take $\omega_j = 2\pi j/T$ say.) One is led to estimate the parameters by least squares in selected frequency intervals. The asymptotic distribution of these estimates is given in Brillinger and Bolt (1979).

18c. OTHER FEW-PARAMETER APPROACHES

There are other situations in which a finite dimensional parameter stands out as of special interest. (Box and Jenkins (1970) is one pertinent reference).

Example 1. (ARMA models). Suppose the series X satisfies the relationship

$$X(t) + a(1)X(t-1) + \cdots + a(p)X(t-p)$$

$$= e(t) + b(1)e(t-1) + \cdots + b(q)e(t-q)$$

where p, q are finite and $e(t)$ is a white noise series of variance σ^2. X is then called an autoregressive moving average (ARMA) process. It is of interest to estimate $\theta = (a(1),...,a(p), b(1),...,b(q),\sigma^2)$ for problems of forecasting and description.

Example 2. (ARMAX models). Suppose one has a bivariate series (X,Y) satisfying the relationship

$$Y(t) + a(1)Y(t-1) + \cdots + a(p)Y(t-p)$$

$$= b(0)X(t) + b(1)X(t-1) + \cdots + b(q)X(t-q) + e(t)$$

where $e(t)$ is a noise series that is an ARMA process of some known order. One will be interested in the estimation of the parameters of the noise process and the a's and b's, given the data stretch $(X(t),Y(t)\ ,\ t=0,...,T-1)$.

Example 3. This generalizes the cosinusoid examples of Sections 18a and 18b. Suppose

$$X(t) = s(t,\theta) + e(t)$$

where $s(t,\theta)$ is a signal of known functional form involving the unknown finite dimensional parameter θ and where e is a stationary noise process. One would like to estimate θ given $X(t), t=0,...,T-1$.

Example 4. In structural engineering one may have a finite dimensional model of a building response to an impulse. Suppose this impulse response is $a(t,\theta)$. One might have data on the response Y to an earthquake X. The situation might be modeled as

$$Y(t) = \sum_u a(u,\theta)X(t-u) + e(t)$$

with e a noise process. (One reference is McVerry (1980).) The series Y being a delayed version of X would be another example of this form.

The general properties of FT's of stationary processes indicated earlier lead to a general way of estimating the parameter θ in these examples and to approximate statistical properties of the estimate. The procedure will be indicated for Example 4. The FT is $d_e^T(\omega)$, of the noise may be treated as if they were complex normal with mean 0 and variance $2\pi T\ S_{ee}(\omega)$. FT's at different frequencies may then be treated as if they are statistically independent. Consider a number of frequencies β_k scattered along the interval $[0,\pi]$, say of the form $2\pi k/K$. Consider a number of frequencies ω_{jk}, $j=1,...,J$ near β_k. The negative log-likelihood of the $d_e^T(\omega_{jk})$ is approximately of the form

$$\text{constant} \ + \sum_k \{\sum_j |d_Y^T(\omega_{jk}) - A(\omega_{jk},\theta)d_X^T(\omega_{jk})|^2\}/2\pi T\ S_{ee}(\beta_k)$$

with $A(\omega,\theta)$ the transfer function of the impulse response $a(t,\theta)$. θ and the $S_{ee}(\beta_k)$ may be estimated by minimizing this expression. One might construct estimates without including S_{ee} in the criterion. They will be consistent, but not necessarily (asymptotically) efficient.

In proceeding in this fashion, it is not necessary that the noise process be Gaussian. The estimates may be expected to be more efficient in the Gaussian case, however. (They would fail dismally if the d^T's followed very stretched tail distributions, but this does not seem very likely.)

18d. AN EXAMPLE: VARIANCE COMPONENTS AND TREND ANALYSIS

One of the important purposes of the analysis of variance ?? is to indicate the stability of meaningful quantities extracted from structured data. The quency-side analysis provides a means of doing this, even when the data is autocorrelated in time.

Bloomfield *et al.* (1980) are concerned with whether there is in fact a trend in atmospheric ozone concentration. Monthly mean values are available for stations in several regions over a twenty year period. An examination of the data suggests the presence of a common component, of a regional component and possibly a trend. Substantive knowledge suggested a particular form for the trend function. The problem was to estimate this trend function's (possibly 0) multiplier and the uncertainty of an estimate.

There exists a frequency domain analogy of the random effects across ?? model (with covariates). Details are given in the above report. The approach is based upon the approximate independence and normality of the discrete Fourier transform values of stationary time series. The problem of Bloomfield *et al.* also involved elements of nonlinear estimation.

19. ROBUST/RESISTANT APPROACHES

The last decade or two has seen the development of statistical techniques that are resistant in the sense that a small percentage of exotic values, no matter how extreme, have little influence on the result. Among these resistant techniques some are robust of efficiency, providing highly efficient results (i.e., results nearly as stable as possible) in *each* of a *variety* of more or less realistic situations. Since it is observed that all techniques that are highly robust of efficiency are also resistant, it is natural to call them robust/resistant.

It may help to motivate the discussion if we distinguish three cases: 1) those where the observational noise is intrinsically long-tailed and independent of the phenomenon being studied (for example nearby lightning strikes or power line inductive transients), 2) long-tailed distributions caused by the process under study, but explainable (for example errors caused by clipping of process extremes or "clicks" and "snaps" Rice (1963)), 3) cases where the exotic values provide the information (for example the earthquakes on a 24 hour seismogram.)

If we are summarizing measurements $y_1, y_2, \ldots, y_n$ of the same thing, the mean, $\bar{y}$, is not resistant, since changing any one value will charge $\bar{y}$ any amount desired. The median y', the middle value among the y's reordered by their value, is usually resistant. If you change only one (even sample size) or two (odd sample size) y's in any way, the most you can do is to replace the original median by one or the other of the values of the y_i's immediately adjacent to that original median. The efficiency of the median is only moderately high for Gaussian samples (higher for samples from distributions that provide more frequent exotic values). We can preserve the resistance and improve the efficiency (a) somewhat, by changing to the midmean y^*, the mean of the middle half of the ordered y's, or (b) considerably further, by going to slightly more subtle estimates, such as the one-step biweights (Mosteller and Tukey (1977)) or (c) somewhat further still by going to two (or few) situation analogs of Pitman estimators (Pitman (1939), Pregibon & Tukey (1981).)

Since we will want to mention analogs, one-step or multi-step, of the one-step biweights, we will define the latter here. To combine $y_1, y_2, \ldots, y_n$ into the one-step biweight estimate of location, we take the scale estimate S = half the difference between the hinges (or quartiles) of the ordered observation; $c = a$ constant, often between 6 and 9; T_0 = either y' or (slightly better) y^*. We then define the exoticity by

$$u_i = \frac{y_i - T_0}{cS}$$

the (bisquare) weights by

$$\begin{aligned} w_i &= (1-u_i)^2, \quad \text{for } u_i^2 \le 1 \\ &= 0, \qquad\qquad \text{else} \end{aligned}$$

and the one-step biweight estimate (sometimes specified as the wc-biweight where c is replaced by the numerical value of c used) by

$$y_{bhw} = \frac{\Sigma w_i y_i}{\Sigma w_i} \; .$$

(In generalizations of this problem, it may be well to write T_1 for y_{bhw}, and then repeat the calculation, replacing T_0 in turn by $T_1, T_2, \ldots$.)

19a. THE NEEDS

It is far from infrequent to deal with data in which a small fraction of the observations have been heavily corrupted in some way. We refer to an example of dust in a waveguide shortly. Low duty-cycle impulse noise, e.g. lightning strokes, provides many others. It is easy, either overall, or in some restricted frequency band, for the corruption to involve more energy than the signal (which may well be noise-like) that we want to study. Estimating the spectrum of the corrupted signal helps us not at all in such cases, since it tells us almost exclusively about the corrupting spectrum.

We must look beneath the impulses. We can do this either by using resistant summaries of the data on the way to estimates of the uncorrupted spectrum, or we can modify the observations in such a way that most of the corruption is removed.

19b. ONE APPROACH.

The latter approach has been studied and developed by Kleiner, Martin and Thomson (1979) to the point where, as we have seen in an example (Section 5c), it can handle very extreme conditions. The basic ideas are simple:

(1) We must iterate our modifications and our spectrum estimate together.

(2) Given a spectrum estimate, we can forecast each observation from the (modified) observations before it (linear prediction).

(3) We can treat the difference between observation and prediction in a robust/resistant way, accepting all of it when it is small, but none of it when it is large.

(4) We can then add the acceptable part of the difference to the prediction to obtain an adjusted value.

For details, see Kleiner et al. (1979).

See subsection 5c for an extreme example; estimating characteristics of a WT4 wave guide that would have been wholly unmanageable without robust/resistant techniques.

20. MISSING VALUES

Missing values plague most kinds of data to some degree; though their frequency of occurrence varies widely. What needs to be done about them also varies widely, but forgetting them is rarely desirable. This is particularly true in time series work, where consistency of spacing is almost always important. They are, however, always much less treacherous than exotic or corrupted values, since their presence (i.e. their absence) is ordinarily unmistakable.

The two descriptive parameters of importance are (i) how frequently missing? and (ii) individually or in blocks of what length?

20a. INTERPOLATION

When missing values are scattered, occurring mainly alone, or in pairs, the solution of interpolation, say between adjacent pairs of non-missing values, is often reasonably satisfactory.

Since we would prefer to interpolate in a white-noise (a flat-spectrum series), one way to proceed is as follows:

(1) interpolate in the original series,

(2) find a good prewhitening filter for the result (always equivalent to the error of a good predictive filter)

(3) form the prewhitened series, delete the values interpolated in step (1), and reinterpolate them in the resulting series (possibly basing the interpolation on the two or three values on each side of the gap)

(4) if necessary, apply the inverse of the prewhitening filter.

Notice that, if we are going on to most forms of spectrum analysis, we need not take step (4), since we can usually do better analyzing the prewhitened series, compensating for the prewhitening at a later stage.

20b. APPROACHES USING ROBUST TECHNIQUES

The details of a robust technique for dealing with missing values have been considered by Schwartzchild (1979) in his Princeton Ph.D. thesis, and by others in references that thesis cites. A less satisfactory approach would be to insert rather bad values wherever values were missing and then apply the robust technique discussed in the last section. (This might be attractive when the robust technique is already implemented.)

H. DIVERSITY IN SPECTRUM ANALYSIS

21. THREE MAIN BRANCHES

The present account has concentrated on the main branch where the data are "noise-like" -- where a "repetition" would mean a record with similar underlying characteristics but a quite different appearance -- where, as we have pointed out, thinking about Gaussian processes may be a very useful guide (but often not one to rely on in detail -- where what we calculate can easily be rather more variable than for a truly Gaussian situation).

Many of the other papers in this issue discuss the second main branch, where the data is "signal-like" and a repetition will differ mainly in ancillary and measurement noise. As we have already stressed, this is quite a different problem, so we should not be surprised by quite different solutions.

Finally, there is a third branch, one to which the limitations of the data have forced many, particularly statisticians. Here we usually lack all the powerful tools that help us otherwise. Our records are not long. The appearance of our data is not distinctive. Our models are not narrowly restricted by reliable subject-matter knowledge. And equivalent records do not look alike. All in all, a horrid fate.

With almost nothing to work with, it is very hard to do anything much more than fitting a few constants. So we go the AR, MA, ARMA, etc. (AR = autoregressive, MA = moving average) route. Statisticians who face such difficult situations (i) are usually familiar with the book by Box and Jenkins, and (ii) are most often concerned with adjusting their data before applying statistical procedures whose leading cases call for uncorrelated observations. This approach seems to work rather a lot better than might be feared, since local irregularities in the spectrum -- inevitable when a real, even very modestly complex, spectrum is "whitened" by fitting a few constants -- turn out not to bother many of these statistical techniques seriously.

Sometimes, though not as generally as often supposed, some of the constants fitted in this third branch are meaningful (cp. Sections 5d, 20 and 20c). Usefulness of the third branch, however, most often comes from preparation for further analysis.

35. THREE ASPECTS

Time series analysis, whether or not specifically based on spectrum estimations, can differ in at least three major ways: in aims, in revealed behavior, and in character of data. These refer, respectively, to us, to what can be seen in the data, and to what could be seen if we had enough repetitions of the data. Each of the first two aspects comes in more than half a dozen flavors; we have repeatedly stressed an at least two-fold separation ("noise-like" and "signal-like") for character. Thus there are at least 100 combinations.

We cannot plan to have 100 different kinds of time series analysis at hand at one time, so we must look toward considerable unification of methods. Part D of Tukey (1980a) can be consulted for suggestions of what a unified procedure might be like, especially in its iterative, graphical, and dualized nature. We content ourselves here with two condensed descriptions of diversity: one of different kinds of aims and one of different kinds of revealed behavior.

23. DIVERSITY OF AIMS

We can easily put down a half-dozen kinds of aims, namely:

(1) Discovery of phenomena -- of distinctive things that are rather isolated -- in frequency, in time, or, as we may sometimes see, in frequency-and-time-combined.

(2) "Modeling" -- often prevalent among those who see an abstract good in identifying a few-constant structure that is at least not powerfully contradicted by the data (even though, as frequently happens, this structure is not at all reasonable in detail).

(3) Preparation for further inquiry -- a practically very important aim, for help in doing which Tukey (1980a) lists 8 classes of techniques. (Should be counted as two or more aims.)

(4) Reaching conclusions in statistical terms -- as when we desire statements about statistical significance or about confidence intervals.

(5) Assessment of predictability -- how well can it be done for how far into the future (usually equivalent to finding a procedure that whitens the data, and then using these whitened data values as prediction residuals).

(6) Description of variability, at least roughly (most useful in spectrum terms) -- where we must be keenly aware of the inevitable uncertainties of variability assessment for noise-like data.

All are practised and all are important (except perhaps "modeling"). No one can be replaced by another (unless in those instances where "modeling" is used, but "description of variability" is meant).

24. DIVERSITY OF BEHAVIOR

When we consider behavior, the issue is usually phenomena. It helps to describe kinds of phenomena according to a number of coordinates:

(1) Where? -- time or frequency (or perhaps both together).

(2) How strong? -- conveniently classified as "dull," "interesting," "distinctive" or "unmistakable."

(3) How simple? -- how many constants does it take to give a useful (but incomplete) description?

(4) How dominant? -- to what extent will the results of our analysis, unless we take special steps, refer mainly to this one kind of phenomenon.

Notice that phenomena can be distinctive without being dominant. It may suffice to be sharp in time or frequency. Examples include the "pole tide" (Haubrich and Munk 1959) and isolated peals of thunder.

Together, these coordinates easily describe more than a half-dozen important patterns of behavior.

25. PARSIMONY VS. FULL AND FLEXIBLE INQUIRY

The unified approach discussed elsewhere (Tukey 1980a) might work with 300 relevant data points, is more likely to work with 1000, and has a good chance of working with 3,000 or 10,000 but will not always. What if we have fewer than 300?

As it is put in the reference just cited:

"(a) We have finite data -- and face infinite possibilities."

"(b) How much can we AFFORD TO TRY to learn?"

There is no trace of jest in this statement. We all know what we would think of a surveyor who used 27 measured angles to locate even 28 coordinates of otherwise unmeasured points (to say nothing of 280). If, for another example, we want to estimate the average spectrum of an electroencephalogram between 10 and 11 Hertz using 30 seconds of data, we need to be keenly aware that we have only the equivalent of 30 relevant data points. If the subject's brain waves are nearly Gaussian we can only get the total power to half a significant figure (cp. our introduction), and dare not think of asking narrower questions about 0.1-Hertz bands, for with 30 seconds of data we would only have 3 independent looks at the spectrum for each.

For those who must deal with "noise-like" data, the conflict between

(1). the necessities of parsimony, AND

(2) our desires for detailed answers

can only be managed -- too often only by giving up on detail -- and cannot be removed. (The conflict between frequency resolution and variability-estimate precision is only one example of a wider conflict.)

I. ACKNOWLEDGEMENTS

We thank our friends and colleagues P. Bloomfield and W. S. Cleveland. Our earlier collaboration with them on a work on the practice of spectrum analysis has had substantial influence on the form of this paper. We also thank D. J. Thomson, J. F. Kaiser, J. B. Allen, L. R. Rabiner, R, Beran, P. Bloomfield, S. Morgenthaler, B. Sagalovsky. referees and the editor for comments. The research was supported in part by National Science Foundation CEE-7901642 and AROD DAAG29-79C-0205.

REFERENCES

H. Akaike, *"Some problems in the application of the cross-spectral method,"* Advanced Seminar on Spectral Analysis of Time Series, (Ed. B. Harris), pp. 81-108, Wiley, New York, 1967.

K. Aki and B. Choet, *"Origin of coda waves: source, attenuation, and scattering effects,"* J. Geophysical Res., vol. 80, pp. 3322-3342, 1975.

T. W. Anderson, The Statistical Analysis of Time Series, Wiley, New York, 1971.

E. D. Becker and T. C. Farrar, *"Fourier transform spectroscopy,"* Science, vol. 178, pp. 361-368, 1972.

S. Bevan, R. Kullberg and J. Rice, *"Ananalysis of cell membrane noise,"* Ann. Statist., vol. 7, pp. 237-257, 1979.

C. Bingham, M. D. Godfrey and J. W. Tukey, *"Modern techniques of power spectrum estimation,"* IEEE Trans. Audio Electroacoust., vol. AU-15, pp. 56-66, 1967.

R. B. Blackman and J. W. Tukey, *The Measurement of Power Spectra from the point of view of Communications Engineering*, Dover, New York, 185pp. 1959.

P. Bloomfield, M. L. Thompson, G .S. Watson and S. Zeger, *"Frequency domain estimation of trends in stratosphere ozone,"* Tech. Report No. 182, Statistics Department, Princeton University, 1980.

B. J. Bogert, M. J. Healy, and J. W. Tukey, *"Chapter 15, the frequency analysis of time series for echoes-cepstrum, pseudo-autocovariance, cross-spectrum and saphe-cracking."* Time Series Analysis, (Ed. M. Rosenblatt), pp. 209-243, J. Wiley, New York, 1963.

B. P. Bogert and J. F. Ossanna, *The heuristics of cepstrum analysis of a stationary complex echoed Gaussian signal in stationary Gaussian noise*, IEEE Trans. on Into. Theory, vol. IT-12, pp. 373-380, 1966.

B. A. Bolt and D. R. Brillinger, *"Estimation of uncertainties in eigenspectral estimates from decaying geophysical time series,"* Geophys. J. R. astr. Soc., vol. 59, pp. 593-603, 1979.

B. A. Bolt, Y. B. Tsai, K. Yeh and M. K. Hsu, *"Earthquake strong motions recorded by a large near-source array of digital seismometers,"* Earthquake Eng., Structural Dynamics, vol. 10, pp. 561-573, 1982.

G. E. P. Box and G. M. Jenkins, *Time Series Analysis, Forecasting and Control*, Holden-Day, San Francisco, 1970.

F. P. Bretherton and J. C. McWilliams, *"Estimations from irregular arrays,"* Reviews of Geophysics, vol. 18, pp. 789-812, 1980.

D. R. Brillinger and M. Rosenblatt, "Computation and interpretation of k-th order spectra," Spectral Analysis of Time Series, pp. 189-232, J. Wiley, New York, 1967.

D. R. Brillinger, *"The frequency analysis of relations between stationary spatial series,"* Proc. Twelfth Bien. Seminar Canadian Math. Congress, (R. Pyke, Ed.), pp. 39-81, 1970.

D. R. Brillinger, *"The spectral analysis of stationary interval functions," Proc. Sixth Berkeley Symp. Math. Statist. Prob.*, vol. 1, pp. 483-513, Berkeley, UC Press, 1972.

D. R. Brillinger, H. L. Bryant Jr. and J. P. Segundo, *"Identification of synaptic interactions,"* Biol. Cybernetics, vol. 22, pp. 213-228, 1976.

D. R. Brillinger, Time Series: Data Analysis and Theory, (Expanded Edition), San Francisco, Holden-Day, 1981.

J. J. Bussgang, *"Crosscorrelation functions of amplitude-distorted Gaussian signals,"* Mass. Inst. Tech. Res. Lab. Electronics Report No. 216, 1952.

G. C. Carter and A. H. Nuttall, *"On the weighted overlapped segment averaging method for power spectrum estimation,"* Proc. IEEE., vol. 68, pp. 1352-1354, 1980.

D. G. Childers, D. P. Skinner, and R. C. Kemerait. *"The cepstrum: a guide to processing,"* Proc. IEEE vol. 65, pp. 1428-1443, 1977.

J. Claerbout. Fundamentals of Geophysical Data Processing, McGraw-Hill, New York, 1974.

W. S. Cleveland and E. Parzen, *"The estimation of coherence, frequency response and envelope delay." Technometrics, vol. 17, pp. 167-172, 1975.*

P. Connes, J. Connes, W. S. Benedict, and L. D. Kaplan, "Traces of HCl and HF in the atmosphere of Venus," Astrophys. J. vol. 147, pp. 1280-1297, 1967.

P. Connes, J. Connes, L. D. Kaplan and W. S. Benedict, *"Carbon monoxide in the Venus atmosphere,"* Astrophysical J., vol. 152, 731-743, 1968.

P. Connes and G. Michel, *"High resolution Fourier spectra of stars and planets,"* Astrophysical J., vol. 190, pp. L29-L32, 1974.

L. J. Cote, J. O. Davis, W. Marks, R. J. McGough, E. Mehr, W. J. Pierson Jr., J. F. Ropek, G. Stephenson and R. C. Vetter, "The directional spectrum of a wind generated sea as determined from data obtained by the Stereo Wave Observation Project," Meteorol. Papers, College of Engineering, NYU, 1960.

Digital Signal Processing Committee, Programs for Digital Signal Processing, IEEE Press, NY, 1979.

A. S. French and A. V. Holden, *"Alias-free sampling of neuronal spike trains,"* Kybernetik, vol. 8, pp. 165-171, 1971.

T. D. Gramble, W. M. Goubau and J. Clarke, *"Magnetotellurics with a remote magnetic reference,"* Geophysics, vol. 44, pp. 53-68 and 959-968, 1979.

E. J. Hannan, *"Non-linear time series regression,"* J. Appl. Prob., vol. 8, pp. 767-780, 1971.

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

K. Hasselman, W. Munk and G. MacDonald, *"Bispectra of ocean waves,"* Time Series Analysis (Ed. M. Rosenblatt). J. Wiley, New York, pp. 125-139, 1963.

R. Haubrich, and W. H. Munk, *"The Poletide,"* J. Geophys. Res., vol. 53, pp. 2373-2388, 1959.

K. N. Hellard, K. S. Lii and M. Rosenblatt, *"Bispectra of atmospheric and wind tunnel turbulence,"* Applications of Statistics (Ed. P. R. Krishnarah), North-Holland, Amsterdam. pp. 223-248, 1977.

H. James, N. Nichols and R. Phillips, Theory of Servomechanisms, Radiation Laboratory Series, vol. 25, McGraw-Hill, New York, 1947.

J. F. Kaiser, Chapter 7, Digital Filters, in System Analysis by Digital Computer, e.d. S. F. Kuo and J. F. Kaiser, New York-Wiley, 1966.

H. Kawasumi, *"Proofs of 69 years periodicity and imminence of destructive earthquakes in Southern Kwanto district and problems in the countermeasures thereof,"* Chigaku Zasshi, vol. 79, pp. 115-138, 1970.

G. E. Kimball, *"A critique of operations research,"* J. Wash. Acad. Sci. vol. 48, pp. 33-37 (especially p. 35). 1958.

B. Kleiner, R. D. Martin and D. J. Thomson, *"Robust estimation of power spectra,"* J. Roy. Statist. Soc. B., vol. 41, pp. 313-351, 1979.

M. J. Levin, *"Optimum estimation of impulse response in the presence of noise,"* IRE Trans. Circuit Theory, vol. CT-7, pp. 50-56, 1960.

G. C. Levy and D. J. Craik, *"Recent developments in nuclear magnetic resonance spectroscopy,"* Science, vol. 214, pp. 291-299, 1981.

N. K. Lin, Excitation of Edge Waves, Ph.D. thesis, Civil Engineering Department, University of California, Berkeley, 1981.

P. Z. Marmarelis and V. Z. Marmarelis, Analysis of Physiological Systems, New York, Plenum, 1978.

G. H. McVerry, "Structural identification in the frequency domain from earthquake records," Earthquake Eng. Structural Dynamics, vol. 8, pp. 161-180, 1980.

R. K. Mehra, *"Optimal input signals for parameter estimation in dynamic systems — a survey and new results,"* IEEE Trans. Auto Control vol. AC-19, pp. 753-768, 1974.

M. M. Milankovitch, *"Mathematische Klimalehre und Astronomische Theorie der Klimaschwankungen,"* in Handbook der Klimatologie, ed. W. Köppen and R. Geiger, Borntrāger, Berlin Band 1, Tel A, 1-176, 1930.

M. M. Milankovitch, Kanon des Erdbestrahlung und seine Anmendung auf das Eiszeitproblem, R. Serb Ac. Sci., Special Publication 133, (Kónglich Serbische Akadémie, Beograd), 1941.

M. M. Milankovitch, Canon of Insolation and the Ice-Age problem (English Translation of above by the Israel Program for Scientific Translations, Jerusalem) published for the U. S. Dept. of Commerce and the National Science Foundation, Washington, 1969.

J. H. Miles, *"Analysis and correction of ground reflection effects in measured narrow-band sound spectra using cepstral techniques,"* J. Acoust. Soc. Am., vol. 58, p. S109, 1975.

K. S. Miller, Complex Stochastic Processes, Addison-Wesley, Reading, 1974.

S. K. Mitchell and N. R. Bedford, "*Long-range sensing of explosive source depths using cepstrum*," J. Acoust. Soc. Am., vol. 68, p. S20, 1975.

F. Mosteller and J. W. Tukey, Data Analysis and Regression, Reading, Addison-Wesley, 1977.

W. H. Munk and J. F. Gordon MacDonald, The Rotation of the Earth, Cambridge, University Press, pp. 323, 1960.

W. H. Munk and F. E. Snodgrass, "*Measurements of southern swell at Guadelupe island*," Deep-Sea Research, vol. 4, pp. 272-286, 1957.

W. Munk, F. Snodgrass, and F. Gilbert, "*Long waves on the continental shelf — an experiment to separate trapped and leaky modes*," J. Fluid Mechanics, vol. 20, pp. 529-554, 1964.

A. H. Nuttall, "*Some windows with very good sidelobe behavior*," IEEE Trans. Acoustics, Speech, Signal Processing, vol. ASSP-29, 84-91, 1981.

H. Nyquist, Certain topics in telegraph transmission theory. Trans. A.I.E.E., pp. 617-649, April 1928.

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

H. A. Panofsky, "*Meterological applications of cross-spectrum analysis*," Spectral Analysis of Time Series, (Ed. B. Harris), J. Wiley, New York, 109-132, 1967.

H. A. Panofsky and R. A. McCormick, "*Properties spectra of atmospheric turbulence at 100 meters*," Quart. J. Roy. Meteorol Soc., vol. 80, pp. 546-564, and vol. 81, pp. 603-608, 1954.

G. A. Pearson, M. Rocek, and R. I. Walter, "*Cepstral and direct analysis of electron spin resonance spectra of substituted triarylaminium cation radicals. Correlation of spin distribution with substituent constants.*" *J. Phys. Chem., vol. 82, pp. 1181-1592, 1978.*

E. J. G. Pitman, "The estimation of the location and scale parameters of a continuous population of any given form," Biometrika, vol. 30, pp. 391-94, 1939.

D. Pregibon, and J. W. Tukey, "*Assessing the behavior of robust estimates of location in small samples: Introduction to configured polysampling*," Technical Report No. 185 (Statistical Techniques Research Group), Department of Statistics, Princeton University, 1981.

M. B. Priestley, "*The analysis of stationary processes with mixed spectra I*," J. R. Statist. Soc. B, vol. 24, pp. 511-529, 1962.

M. B. Priestley, "*Estimation of transfer functions in closed-loop stochastic systems*," Automatica, vol. 5, pp. 623-632, 1969.

M. B. Priestley, Spectral analysis and Time Series, vols. I, II, Academic Press, London, 1981.

L. R. Rabiner, et al., "*Terminology in digital signal processing*," IEEE Trans. Audio Electroacoust., vol. AU-20, pp. 322-337, 1972.

L. R. Rabiner and R. W. Schafer, Digital Processing of Speech Signals, Prentice-Hall, Englewood

Cliffs, 1978.

A. Rahimi. On-line spectral control of rotating circular discs using thermal membrane stress, Ph.D. Thesis (Mechanical Engineering), University of California, Berkeley, 1982.

S. O. Rice, "*Noise in FM receivers*," Time Series Analysis (Ed. M. Rosenblatt), pp. 395-424, J. Wiley, New York, 1963.

L. E. Roemer and C. S. Chen, "*Maximum entropy method spectral estimation applied to power cable diagnostics via cepstrum processing*," J. Franklin Institute, vol. 310, pp. 145-153, 1980.

R. Rom, "*On the cepstrum of two-dimensional functions*," IEEE Trans. Info. Theory, vol. IT-21, pp. 214-217, 1975.

T. Sato, K. Sasuki and Y. Nakamura, "*Real-time bispectral analysis of gear noise and its application to contactless diagnosis*," J. Acoust. Soc. Am., vol. 62, pp. 382-387, 1977.

R. W. Schafer, (Ed.), Speech Analysis IEEE Press, NY, 1979.

R. A. Scholz, "The origins of spread-spectrum communications." IEEE Trans. Commun., vol. COM-30, pp. 822-854, 1982.

C. Van Schoonveld and D. J. Frijling, "*Spectral analysis: on the usefulness of linear tapering for leakage suppression*," IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-29, pp. 323-329, 1981.

A. Schuster, "*On interference phenomena*," Phil. Mag., vol. 37, pp. 509-545, 1894.

M. Schwartzchild, New observation-outlier-resistant methods of spectrum estimation. Ph.D. Thesis (Statistics), Princeton University, 1979.

A. A. Syed, J. D. Brown, M. J. Oliver and S. A. Hills, "The Cepstrum: a viable method for the removal of ground reflections," J. Sound Vibr., vol. 71, pp. 299-313, 1980.

D. J. Thomson, "*Spectrum estimation techniques for characterization and development of WT4 waveguide — I, II*," Bell System Tech. J., vol. 56, pp. 1769-1815 and 1983-2005, 1977.

J. W. Tukey, "*Statistical and qualitative methodology" pp. 84-136, (especially pp. 99-100) of Trends in Social Science, (ed. D. P. Ray) Philosophical Library, New York, 1961.*

J. W. Tukey, "*Uses of numerical spectrum analysis in geophysics*," Bull. International Statistical Institute, vol. 41, pp. 267-306, 1966.

J. W. Tukey, "*Can we predict where 'time series' should go next?*," Directions in Time Series, (Eds. D. R. Brillinger and G. C. Tiao), pp. 1-31, 1980a.

J. W. Tukey, "*Chapter 16, Methodological Comments Focused on Opportunities*," Multivariate Techniques in Human Communication Research (Eds. P. R. Monge and J. N. Cappella), pp. 490-528, Academic Press, New York, 1980b.

T. M. Turpin, "*Spectrum analysis using optical processing*," Proc. IEEE, vol. 69, pp. 79-92, 1981.

D. Vere-Jones and T. Ozaki, "*Some examples of statistical estimation applied to earthquake data,* *I*," Ann. Inst. Statist. Math., vol. 34, to appear.

D. K. Weaver, "*A third method of generation and detection of SSB signals*," Proc. IRE, vol. 44, pp. 1703-1705, 1956.

J. M. Whittaker, Interpolating Function Theory, Cambridge University Press, (Chapter IV), Cambridge, England, 1935.

P. Whittle, "*The simultaneous estimation of a time series' harmonic components and covariance structure*," Trab. Estad., vol. 3, pp. 43-57, 1952.

N. Wiener, "*Generalized harmonic analysis*," Acta Math., vol. 55, pp. 117-258, 1930.

N. Wiener, Nonlinear Problems in Random Theory, Cambridge, MIT Press, 1958.

A. M. Yaglom, "*Correlation theory of processes with stationary n-th order increments*," Amer. Math. Soc. Transl., vol. 8, pp. 87-142, 1958.

Supplement To:

SPECTRUM ANALYSIS IN THE PRESENCE OF NOISE: SOME ISSUES AND EXAMPLES

David R. Brillinger

Statistics Department
University of California
Berkeley, California

John W. Tukey

Statistics Department
Bell Laboratories and Princeton University
Murray Hill, New Jersey and Princeton, New Jersey

1. DISTRIBUTIONS OF FOURIER TRANSFORMS

1a. FOURIER TRANSFORMS OF NEAR-GAUSSIAN PROCESSES

Directly or indirectly, spectrum estimates of the sort we are discussing are constructed from Fourier transforms. For Gaussian processes, the corresponding transform processes are simply and nicely behaved. For processes that are nearly Gaussian, transform process behavior is nearly simple and nearly nice. We need to understand a little more detail than this, but in the present context not too much.

1b. THE GAUSSIAN CASE

If time is discrete, and we focus on a finite segment of it (as data always must), the resulting finite process consists of finite realizations $\{X(0), \ldots, X(T-1)\}$ corresponding to finite transforms $\{\hat{X}(0), \hat{X}(1), \ldots, \hat{X}(T-1)\}$ related by

$$\hat{X}(p) = \sum_{t=0}^{T-1} X(t) \exp\{-i2\pi pt/T\}$$

$p = 0, \ldots, T-1$ and

$$X(t) = T^{-1} \sum_{p=0}^{T-1} \hat{X}(p) \exp\{i2\pi tp/T\}$$

Thus, since the relations are linear, if either the X-vector or the $\hat{X}$-vector is multivariate Gaussian, so too is the other.

If the variance and covariances of the X-vector are stationary, $(cov\{X(t+u), X(t)\} = c(u))$, then

$$cov\{\hat{X}(p), \hat{X}(q)\}/T$$

$$= \sum_{u=-T+1}^{T-1} \exp\{-i2\pi pu/T\} c(u) \left(\sum_t \exp\{i2\pi(q-p)t/T\}\right)/T$$

with the inner sum over t satisfying $0 \le t, t+u \le T-1$. For large T, the inner sum will be essentially 1 if $p = q$ and essentially 0 if $p \ne q$. Taking note of the definition of the power spectrum we see that the transform values satisfy

$$cov\{\hat{X}(p),\hat{X}(q)\}/T \sim 2\pi S\left(\frac{2\pi p}{T}\right) \text{ if } p=q$$

$$\sim 0 \text{ if } p \neq q$$

Distinct values of the transform are approximately orthogonal, and the variance of a particular value is proportional to the power spectrum at its frequency. The basic requirement here is that the spectrum changes only slowly over the frequency ranges relevant to single transforms (say of width 2, 4 or 6 times $2\pi/T$). The distribution of the values is further (complex) Gaussian.

Such near orthogonality (corresponding to near independence in the Gaussian case) is one of the reasons it is often simpler to proceed in the frequency domain.

One can study the details of this near-orthogonality in two quite different ways: (i) making asymptotic calculations, (ii) working with exact representations and then making approximations suggested by easily established properties. We shall do only the first here, except for one illustrative formula at the close of the subsection.

If the Fourier transform is evaluated at an arbitrary frequency ω,

$$d^T(\omega) = \sum_{t=0}^{T-1} X(t)\exp\{-i\omega t\}$$

then one has

$$cov\{d^T(\omega),d^T(\nu)\} \sim 2\pi\Delta^T(\omega-\nu)S(\omega) \tag{1.1}$$

with

$$\Delta^T(\omega) = \sum_{t=0}^{T-1} \exp\{-i\omega t\}.$$

Being a linear function of the $X(t)$, values of d^T have a Gaussian distribution. For ω and ν not too close, (modulo 2π), approximate orthogonality of $d^T(\omega)$ and $d^T(\nu)$ holds, under the usual conditions.

In the tapered case, where

$$d^T(\omega) = \sum_t X(t)h^T(t)\exp\{-i\omega t\}$$

the values of d^T continue to have a Gaussian distribution and one now has

$$cov\{d^T(\omega),d^T(\nu)\} \sim 2\pi H_2^T(\omega-\nu)S(\omega) \tag{1.2}$$

with

$$H_2^T(\omega) = \sum_t h^T(t)^2\exp\{-i\omega t\}.$$

By choice of h^T, such that H_2^T dies off rapidly, the values $d^T(\omega)$ and $d^T(\nu)$ may be made more strongly orthogonal.

One also has the (non-asymptotic) result, illustrative of the other approach,

$$cov\{d^T(\omega),d^T(\nu)\} = \int H^T(\omega-\alpha)\overline{H^T(\nu-\alpha)}\,S(\alpha)d\alpha \tag{1.3}$$

where

$$H^T(\omega) = \sum_t h^T(t)\exp\{-i\omega t\}$$

(The covariances indicated above are derived in Chapter 4 of Brillinger (1981).

1c. NON-GAUSSIAN CASES: LARGE T

These simplifications, of approximate independence and Gaussianity for the Fourier values, continue to hold, to a satisfactory approximation, for a broad class of non-Gaussian processes as well. This approximate Gaussianity has long been part of engineering knowledge in the guise of a folk theorem -- narrow bandpass filtered noise is Gaussian -- for the series

$$\text{real part } (T^{-1}d^T(\omega)\exp\{i\omega t\})$$

$t = 0,1,...$ may be viewed as the output of a narrowband filter centered at frequency λ. (See Brillinger (1981), page 97).

The natural assumptions under which this folk theorem may be derived mathematically are stationarity and mixing of the process X. Mixing in the sense that well-separated values of the process are at most weakly dependent in a statistical sense. Mixing typically accompanies continuous spectra (of all orders not just of 2nd order). (In the absence of mixing, ergodic components of the process are likely to have Fourier values with approximate independence and Gaussianity.)

The covariance continues to be given, approximately, by expressions (1.1) and (1.2) in the untapered and tapered cases, respectively. The transform value $d^T(\omega)$, $\omega \neq 0$, can be shown to be asymptotically complex Gaussian with variance proportional to the power spectrum $S(\omega)$ and the transform values $d^T(\omega)$, $d^T(\nu)$ can be shown to be asymptotically independent for distinct frequencies ω and ν. It turns out that not only are the values at distinct frequencies approximately independent, but so too are values of the Fourier transform at the same frequency that are based on distinct stretches of data.

The approximate distributions indicated here are useful in suggesting approximations to the distributions of various time-series statistics and for suggesting solutions to estimation problems of interest. For example, in the untapered case the periodogram is given by

$$I^T(\omega) = (2\pi T)^{-1}|d^T(\omega)|^2$$

with $d^T(\lambda)$, $\lambda \neq 0$, approximately complex Gaussian having mean 0 and variance $2\pi T S(\omega)$. The mod-squared of a complex Gaussian has an exponential distribution (a chi-square distribution with 2 degrees of freedom). Hence one is led to approximate the distribution of $I^T(\omega)$ by that of an exponential variate having mean $S(\omega)$ -- a surprisingly useful approximation. (It must be remembered however, that if there is an appreciable dynamic range in the population spectrum, leakage may occur, and the mean may be far from $S(\omega)$.)

The approximate normality of Fourier transform values is a form of Central Limit Theorem result. The approximation may be expected to break down in situations where the Central Limit Theorem is likely to break down, e.g. long range statistical dependence in the process or low order moments either failing or nearly failing to exist.

1d. IDEAL DEGREES OF FREEDOM

The preceding discussion suggests a direct estimate of a power spectrum, and an approximation to the distribution of that estimate. The estimate is

$$\hat{S}(\omega) = L^{-1}\sum_{\ell=1}^{L} I^T(\omega_\ell)$$

where the ω_ℓ are distinct frequencies near ω, say of the form $2\pi(\text{integer})/T$. The mean of the approximating distribution is $S(\omega)$, the variance is $S(\omega)^2/L$. An approximate 100 β% confidence interval is given by

$$2L\ \hat{s}(\omega)/\chi^2_{2L}\left(\frac{1+\beta}{2}\right) < s(\omega) < 2L\ \hat{s}(\omega)/\chi^2_{2L}\left(\frac{1-\beta}{2}\right)$$

suggesting that it will be better to graph $\log \hat{S}(\omega)$, rather than to look at $\hat{S}(\omega)$ itself. (Here $x_f^2(\gamma)$ denotes the 100 γ% point of χ^2 on f degrees of freedom.)

This last expression gives guidance on how to choose the value of L if one has some idea of the desired precision for $\log \hat{S}(\omega)$. It must be remembered that it has been assumed here that the frequencies ω_ℓ used in the construction of the estimate are near ω, in the sense that all $S(\omega_\ell)$ are near $S(\omega)$. In practice this limits the size of L, for, as some ω_ℓ get removed from ω, the estimate gets more biased.

1e. PRACTICAL DEGREES OF FREEDOM

It is seldom that the direct estimate will be formed in such an elementary fashion as in the previous section. In practice the data will have been tapered and the periodogram values will be averaged with unequal weights. If periodogram values based on different stretches of the data are averaged, the stretches may be overlapping (and a *shingled* estimate formed).

In these more complicated cases, approximating the distribution of the estimate by $S(\omega)\chi_f^2/f$ remains useful, however, the degrees of freedom may not be determined by simple counting. If $\hat{S}(\omega)$ denotes the estimate, then one way to determine an f is to equate the large sample variance of the estimate to that of the approximating distribution, i.e. to take

$$f \sim 2\,S(\omega)^2/var\ \hat{S}(\omega)\ .$$

($S(\omega)$ will drop out of the right hand side here for large T.)

Many estimates may be written out as quadratic forms

$$\sum_{t,u} q^T(t,u)X(t)X(u)$$

in the data, with the q^T varying with the tapering and weighting employed. As indicated in Section 6g the variance of such an expression is, in the Gaussian case,

$$2\int\int |Q^T(\alpha,\beta)|^2 S(\alpha)S(\beta)d\alpha\, d\beta$$

where

$$Q^T(\alpha,\beta) = \sum_{t,u} q^T(t,u)\ \exp\{i(t\alpha-u\beta)\}\ .$$

This discussion leads to the expression, if $S(\omega)$ is nearly enough constant,

$$f = 1/\int\int |Q^T(\alpha,\beta)|^2 d\alpha\, d\beta$$

$$= 1/(2\pi)^2 \sum_{t,u} q^T(t,u)^2$$

for the degrees of freedom.

For the simplest estimates, if the halfwidth (frequency resolution) of $Q^T(\alpha,\alpha)$, the window with which the estimate looks at the spectrum (see Section 6g again), is δ and if there are N data points equispaced by Δ, so that N degrees of freedom are equally spaced from 0 to π/Δ (hence at a rate of 1 every $\pi/N\Delta$), then about

$$8/(\pi/N\Delta) = (8/\pi)(N\Delta) \tag{18.4}$$

degrees of freedom ought to fall in each estimate. Expression (18.4) is then the product of the total observation time and frequency resolution.

Actually we are likely to get up to 50% fewer -- perhaps because the underlying spectrum is not constant enough -- or 50% more -- because the tails of our windows contribute some stability. Thus (18.4) with a large grain of salt (±25% on ±50% perhaps) is often a reasonable basis for assessing degrees of freedom.

1f. RESULTS FOR VECTOR CASE

Suppose the series of interest $\mathbf{X}(t)$, $t = 0, \pm 1, \ldots$ is r-dimensionally vector-valued, stationary, with variance-covariance function

$$\mathbf{c}(u) = ave\{[\mathbf{X}(t+u) - \mu][\mathbf{X}(t) - \mu]^\tau\}$$

where "τ" indicates transposing a matrix, and with spectral density matrix

$$S(\omega) = (2\pi)^{-1} \sum_{u=-\infty}^{\infty} \mathbf{c}(u) \exp\{-i\omega u\} .$$

The Fourier transform of the data stretch $\mathbf{X}(t)$, $t = 0,\ldots,T-1$ namely

$$\mathbf{d}^T(\omega) = \sum_{t=0}^{T-1} \mathbf{X}(t) \exp\{-i\omega t\}$$

satisfies central limit results analogous to those of the univariate case. Before indicating these, we define two distributions that are basic to the description of statistics in the vector case.

A random variate of the form $\mathbf{Y} = \mathbf{U} + i\mathbf{V}$ is said to be multivariate complex normal of dimension r with mean $\mathbf{0}$ and and covariance matrix Σ if $[\mathbf{U}^\tau,\mathbf{V}^\tau]^\tau$ is ordinary multivariate normal with mean $\mathbf{0}$ and

$$ave\ \mathbf{U}\mathbf{U}^\tau ,\ ave\ \mathbf{V}\mathbf{V}^\tau = \frac{1}{2}\,\mathrm{Re}\,\Sigma$$

$$ave\ \mathbf{U}\mathbf{V}^\tau = \frac{1}{2}\,\mathrm{Im}\,\Sigma$$

The averages of matrices here mean the matrices of corresponding averages. One has $ave\ \mathbf{Y}\bar{\mathbf{Y}}^\tau = \Sigma$ and $ave\ \mathbf{Y}\mathbf{Y}^\tau = \mathbf{0}$. A random variate of the form

$$\mathbf{W} = \sum_{j=1}^{n} \mathbf{Y}_j \bar{\mathbf{Y}}_j^\tau$$

is said to be complex Wishart of dimension r with degrees of freedom n and parameter Σ if $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$ are independent multivariate complex normals with mean $\mathbf{0}$ and covariance matrix Σ. (Note that these are complex degrees of freedom, each corresponding to the real degrees of freedom.)

These definitions having been made, one can now state one of the central limit results for $\mathbf{d}^T$. If the series $\mathbf{X}$ is stationary and suitably mixing (see 18b), then $\mathbf{d}^T(\omega)$, for $\omega \neq 0$, is asymptotically multivariate complex normal with mean $\mathbf{0}$ and covariance matrix $2\pi T\ S(\omega)$. It follows directly that the matrix of periodograms

$$\mathbf{I}^T(\omega) = (2\pi T)^{-1}\, \mathbf{d}^T(\omega)\, \overline{\mathbf{d}^T(\omega)}^\tau$$

is asymptotically complex Wishart with degrees of freedom 1 and parameter $S(\omega)$.

Fourier transform values d^T at distinct frequencies (e.g. of the form $2\pi(\text{integer})/T$ are asymptotically independent. This last suggests taking

$$\hat{S}(\omega) = L^{-1} \sum_{\ell=0}^{L} I^T(\omega_\ell) ,$$

with the ω_ℓ distinct and near ω, as an estimate of $S(\omega)$. It further suggests approximating the distribution of the estimate by L^{-1} times a complex Wishart with degrees of freedom L and parameter $S(\omega)$.

Details of the large sample distributions of various statistics based on $\hat{S}(\omega)$, e.g. coherences, and of the construction of confidence intervals, as well as references, are given in Brillinger (1981), Chapter 7.

1g. RESULTS FOR CEPSTRA

For the Gaussian case, see Bogert and Ossanna (1966).

1h. RESULTS FOR BISPECTRA

The variance of the third order periodogram is given by

$$var\ I^T(\omega,\nu) \sim \frac{T}{4\pi} S(\omega)S(\nu)S(\omega+\nu)$$

for $0 < \omega < \nu < \pi$. Periodogram values at distinct bifrequencies are asymptotically uncorrelated. Hence if a bispectrum estimate is computed by averaging L distinct periodogram values, then

$$var\ \hat{S}(\omega,\nu) \sim \frac{T}{4\pi L} S(\omega)S(\nu)S(\omega+\nu)$$

This last expression is of importance, both in looking to see if a non-zero bispectrum seems to be demonstrable in a situation at hand and in the setting of confidence intervals around a computed estimate.

The distribution of the estimate may be further shown to be asymptotically normal, see Brillinger and Rosenblatt (1967).

In some circumstances it turns out to be more convenient and appropriate to estimate the variance by the expression

$$\sum_{j=1}^{n} (Y_j - \overline{Y})^2 / n(n-1)$$

with the Y_j estimates based on disjoint collections of periodograms (see Helland *et al.* (1977).)

1i. FOURIER TRANSFORMS OF POINT PROCESSES, ETC.

One can Fourier transform anything, often meaningfully. Like many other representations, the Fourier transform can represent anything in discrete time and almost anything in continuous time. From what we have seen in the case of ordinary time series, the use of a Fourier transform does not imply that there are periodic phenomena (although there may be). Speaking more broadly, the importance of the FT is not a consequence of the nature of the data observed, rather it is based on the presence or usefulness of linear-time invariant filters in many situations of interest.

As suggested above, one can carry through spectrum estimation and system identification by frequency methods for random processes of characters different from those of ordinary time series. Some details will now be presented.

1j. POINT PROCESSES

The finite Fourier transform of the stretch $\tau_1 \le \ldots \le \tau_{N(T)}$ of point process data is given by

$$d^T(\omega) = \sum_{j=1}^{N(T)} \exp\{-i\omega\tau_j\}.$$

Suppose that this transform is computed for a point process, N, that is stationary with rate p and power spectrum $S(\omega)$, and that is mixing (that is, that values of the process far apart in time are not strongly dependent statistically). For the transform $d^T(\omega)$ one then has

$$ave\ d^T(\omega) = p\delta^T(\omega)$$

$$var\ d^T(\omega) = \int_{-\infty}^{\infty} |\delta^T(\omega-\alpha)|^2 S(\alpha) d\alpha$$

where the function

$$\delta^T(\omega) = \int_0^T \exp\{-i\omega t\} dt$$

has its mass concentrated in the neighborhood of $\omega = 0$. The mean and variance of the standardized transform $(2\pi T)^{-1/2} d^T(\omega)$, $\omega \neq 0$, may now be seen to tend to 0 and $S(\omega)$ respectively as $T \to \infty$. This suggests basing an estimate of $S(\omega)$ on the periodogram $I^T(\omega) = (2\pi T)^{-1}|d^T(\omega)|^2$. In the neighborhood of $\omega = 0$, the expected value of d^T is seen to go to ∞ with T. This suggests that the, necessarily present, very-low-frequencies (crudely "d.c.") leakage would be reduced by basing an estimate on the mean-corrected FT $d^T(\omega) - \hat{p}\delta^T(\omega)$ where $\hat{p} = N(T)/T$. Further, the earlier discussion of the key role of tapering suggests computing the FT

$$d^T(\omega) = \sum_{j=1}^{N(T)} h^T(t)\exp\{-i\omega\tau_j\} - \hat{p}H^T(\omega)$$

where

$$H^T(\omega) = \int h^T(t) \exp\{-i\omega t\} dt$$

with the support of h^T corresponding to the domain of observation of the process. For large T one has, for this Fourier transform,

$$ave\ |d^T(\omega)|^2 \approx \int_{-\infty}^{\infty} |H^T(\omega-\alpha)|^2 S(\alpha) d\alpha \tag{2.1}$$

$$\approx S(\omega) \int_{-\infty}^{\infty} |H^T(\alpha)|^2 d\alpha$$

for all ω.

In the case that well-separated values of the process are not strongly dependent, a central limit effect is present. For large T the distribution of the FT is approximately normal (with mean 0 and variance (2.1)). Further, and this is perhaps surprising, for ω and ν moderately far apart, the values $d^T(\omega)$ and $d^T(\nu)$ are asymptotically independent.

This discussion leads to the consideration of

$$\hat{S}(\omega) = \frac{1}{K}\sum_{k=1}^{K} |d^T(\omega_k)|^2 / \int |H^T(\alpha)|^2 d\alpha$$

as an estimate of $S(\omega)$ for $\omega_1,...,\omega_K$ distinct frequencies near ω. The discussion further suggests that the distribution of $\hat{S}(\omega)$ may be approximated by $S(\omega)\chi^2_{2K}/(2K)$ with χ^2 a chi-squared variate. This is the same distribution as was appropriate in the ordinary time series case.

Expression (2.1) indicates that if $S(\cdot)$ is far from constant, prewhitening may be essential. This is possible in the present situation, at the expense of converting the data from point process to ordinary time series. Specifically, suppose one forms the ordinary series

$$X(t) = \sum_j a(t-\tau_j)$$

for a function $a(\cdot)$, i.e. one inputs the point process to the corresponding filter with impulse response $a(t)$. Then the spectrum of X will be $|A(\omega)|^2 S(\omega)$. By judicious choice of the filter one may have $|A(\cdot)|^2 S(\cdot)$ nearly constant and reduce one's concerns over bias in its estimation. The spectrum estimation procedure, with prewhitening, is, then, first pass the point process through a linear filter to obtain an ordinary equi-spaced time series with relatively flat spectrum, then estimate that spectrum, then take this result times $|A(\cdot)|^{-2}$ as the final point-process spectrum estimate. If one proceeds via this route, an FFT may be used in the computations.

As remarked above, FT values at distinct frequencies are asymptotically independent when the process is mixing. It may also be shown, (see Brillinger (1974) for example), that, under these hypotheses FT's evaluated at the same frequency, ω, but for disjoint stretches of data, are asymptotically independent and Gaussian. Thus estimates of spectra may even be formed by averaging the periodograms based on different segments of the data.

1k. MARKED PROCESSES

For the marked point process $\{(\tau_j, M_j)\}_{j=-\infty}^{\infty}$ set

$$X(t) = \sum_{0\le\tau_j<t} M_j$$

i.e. $X(t)$ is the accumulation of the marks from time 0 to time t. Were the τ_j earthquake times and M_j the energy of the j-th quake, then $X(t)$ would represent the total energy released from time 0 to time t. The FT of a stretch of this process is given by

$$d^T(\omega) = \sum_{0\le\tau_j<T} M_j \exp\{-i\omega\tau_j\} = \int_0^T \exp\{-i\omega t\}\, dX(t) .$$

If the process X is stationary, then $ave\{dX(t)\}$ and $cov\{dX(t+u), dX(t)\}$, for example, will not depend on t. The spectrum of the process may be defined as

$$S(\omega) = (2\pi)^{-1} \int_{u=-\infty}^{\infty} \exp\{-i\omega u\} cov\{dX(t+u), dX(t)\}/dt .$$

As in the previous section, this spectrum may be estimated by smoothing $|d^T(\omega)|^2$, or better yet smoothing the mod-squared of a mean-corrected tapered (complex-valued) FT. The FT will be asymptotically normal provided that well-separated segments of the process are only weakly dependent statistically. This once again leads to a chi-squared approximation for the distribution of the spectral estimate.

Prewhitening may be carried out, forming

$$Y(t) = \int a(t-u)dX(u) = \sum_j M_j a(t-\tau_j)$$

prior to spectrum estimation, if desired. Essentially, this reduces the problem to one of the estimation of the spectrum of an ordinary time series, provided there is enough data that the initial transient from the filtering has died away.

The corresponding bispectrum, $S(\omega,\nu)$, appears, for example, as

$$\lim_{T\to\infty} \frac{1}{(2\pi)^2 T} \, ave \, \{d^T(\omega) d^T(\nu) \overline{d^T(\omega+\nu)}\}$$

$\omega, \nu \neq 0$, as it did for the FT of an ordinary series.

II. PROCESSES WITH STATIONARY INCREMENTS

It may be the case that a series, $X(\cdot)$, is not stationary; however that differences $\Delta X(t) = X(t+\Delta) - X(t)$ are stationary. (For example see Lindsey and Chie (1976).) Such a series has a time-side representation

$$X(t) = \alpha + \beta t + \int_0^t \epsilon(u) du$$

with ϵ a stationary series, and a frequency-side representation

$$X(t) = \int_{-\infty}^{\infty} \left[\frac{\exp\{i\omega t\} - 1}{i\omega}\right] Z(d\omega)$$

where $Z(d\omega)$ is a random measure with the property $cov\{Z(d\omega) Z(d\nu)\} = \delta(\omega-\nu) S(\omega) d\omega d\nu$ (see Yaglom (1958).) An estimate of the spectrum $S(\omega)$ may be based on the FT

$$\int_0^T \exp\{-i\omega t\} \, dX(t) \, .$$

In the discrete time case one would evaluate the FT of the differenced series,

$$\sum_{t=0}^{T-2} \exp\{-i\omega t\}[X(t+1) - X(t)] \, .$$

The operation of differencing plays an essential role in these analyses of ordinary time series data developed in Box and Jenkins (1970), for example. Its use can be thought of as seeking out operations to apply to a nonstationary series, or to one containing pure periodicities (seasonal effects) to make it more nearly stationary with smooth spectrum.

2. PLANNING/DESIGN

We plan to review only the most crucial issues in designing data collections (possibly even planning experiments) to collect data for later spectrum analysis. Much more detailed consideration of more diverse issues is likely to be very rewarding.

Careful thought about measurement techniques is always very desirable. In particular, what is wanted is *useful* measurements, for examination over some *limited* frequency range (which may cover many octaves, but may not). Good frequency response outside the desired frequency range is a *dis*advantage, leading to aliasing problems and, often, to unnecessarily small time intervals in digitized series. It may be worthwhile to modify the measuring device to make its frequency response *poorer!* And if we cannot do this, we may want to put a suitable filter as close to the device as is reasonable.

Careful thought about the character of the measuring device's errors may be very helpful. Even simultaneous measurement with two devices may be worthwhile, either because they will have reasonably independent errors (when the co-spectrum of their outputs will be closer to the spectrum of their common input than either of their spectra will be) or because they have different frequency responses (when we can either use their outputs in different frequency regions or, even, do a careful job of combining their outputs into a single, broader-band result).

Besides issues of frequency response and independence of measurement errors, it is often important to face up to questions of dynamic range. How large a ratio, of largest measurable value to least difference that needs to be measured, must we be prepared for? Can we reduce this ratio a lot by carefully distorting (linear filtering) the frequency response of the measuring device, either by changing it or putting a filter close to it? (Differentiation or integration before digitization are just special cases.) Such considerations can also be vital.

It is probably best to think of all measuring devices as including an attached filter, carefully linear, chosen to meet the real needs, *not* to gain a flat response.

In every case, we should at least think about measuring the frequency response -- in amplitude *and* phase -- of our measuring devices; often this will be important, not only to think about, but to do. (This is clearly a matter of system identification, see parts E and K.)

If, as is so often the case, we plan to digitize the output from the measuring device, we need to ask also about the quality of the digitizer, both in dynamic range and in noise background. For example one might take a slow ramp as input.

2a. DESIGN FOR SIMPLE SPECTRA.

Here, besides the general issues just discussed, the main issues are measurement interval and (relevant) record length, both considered in relation to aliasing, resolution, and precision of estimate.

2b. ALIASING.

Reduction of aliasing problems deserves our most earnest and careful consideration: Realism about the highest frequency that matters. Efforts to filter out (reducing to unimportant levels) all frequencies much higher than those needed. Choice of time interval for digitization. All these deserve careful attention. All must fit together.

2c. PRECISION AND RESOLUTION

For noise-like inputs, precision of spectrum-average assessment and narrowness of bands averaged over are antithetical in nature. We have to buy the *product* of precision and resolution by how many data points we collect (see our introduction for the simple formulas). Costs of recording and processing, though still falling rapidly, can still control. Time durations required can often try to grow to unacceptable lengths. Good judgment and good engineering at this point is crucial.

2d. PLANNED UNEQUAL SPACING

When our concern is with time and (1-dimensional) space together -- or, if you like, with spectra and cross-spectra for points along a line -- we often have to occupy each point from which we take data with a separate measuring device. Expense and effort may dramatically limit the number of occupied points. We may be able to compensate somewhat for a shortage of points by taking records for extreme durations. In such a case, uniform spacing is usually far from desirable, and the difficulties of partial aliasing, though formidable, may have to be faced. Bretherton and McWilliams (1980) have treated this problem with considerable care.

Occasionally, in such combined time-space analyses, we can successfully collect data for different spacings at different times. Trying to do this is usually the mark of the eternal optimist, but Munk et al. (1964) have shown one instance where it works.

2e. DESIGN FOR VECTOR SPECTRA AND SYSTEMS

As long as we worked with single inputs, consistency of time was time's only vital aspect. Time-jitter, as in time of digitization, for example, would be crucial, but absolute time would ordinarily only be needed roughly. Once we go to two or more inputs, we must add to all the single input considerations a very real care about time shifts between inputs, both systematic and varying. Being very sure of what is likely here can be vital.

If we know them, time-shifts between simultaneously recorded or processed inputs (a) must be accurately known and (b) can sometimes be helpful. In studying tsunamis originated near East Asia, for example, collecting records for the same dates and hours at Hawaii and La Jolla would be silly. The rough travel time of the waves is well known. We ought to displace our collection intervals accordingly. This both ensures looking at the same wave trains at both sites and reduces technical difficulties with spiraling. This happens naturally for large earthquakes, when strong-motion seismometers trigger on the first arrival of the signal.

We need not use the same filters on all inputs, but we need to know the performance of each measuring device-filter pair, now *both* in amplitude *and* in phase.

If delays -- often group delays involving nearby frequencies rather than overall delay times -- are unknown, we will do well to allow for correspondingly more extended collection intervals, so that compensation for spiralling (recall subsection 8f) is more thoroughly available.

Frequency resolution is likely to retain its importance. Precision of spectra may remain its effective antithesis, but precision of coherence may take over from it. The antithesis -- and the frequent need for more data than seems desirable (or even bearable) remains.

All the other considerations for single inputs retain their own importances.

2f. DESIGN FOR CEPSTRA

If we hope that cepstra will help us, our first thoughts should probably be about aliasing. The ripple associated with an echo goes right on and on, through the folding frequency determined by the digitization interval. Thus, if substantial power occurs above the folding frequency, its ripples will be folded on top of principal alias ripples, and cancellation may well occur. The best cure here seems to be careful bandlimiting of the input data prior to digitization. The procedure is then:

(a) careful analog filtration, with cutting-off concentrated between $\omega = \frac{2}{3}(\frac{\pi}{\Delta})$ and $\omega = (2 - \frac{2}{3})(\frac{\pi}{\Delta})$,

(b) followed by digitization at interval Δ,

(c) followed by fine-grained spectrum calculation,

(d) followed by discarding estimates for $\omega > \frac{2}{3}(\frac{\pi}{\Delta})$

(e) followed by taking logs,

(f) followed by broad-stop liftering,

(g) followed by a spectrum calculation.

We are not now really working in the frequency domain, so the resolution(bandwidth)-precision antithesis is not as clear. We do gain, however, the ability to use a finer-grain (alternatively, more precise) initial spectrum when we collect more data, and this can be vital for a clean, well-interpretable cepstrum. The antithesis may be concealed, but is almost always there.

A further question about frequency ranges arises. While many echoes are sharp, many are dispersed. Once delay is frequency dependent, it may be urgent to separate two or more frequency bands for separate cepstrum analysis. In any event, discarding the very-lowest-frequency part of the initial spectrum, and the part well into or above the actual cutoff, is likely to be important.

2g. DESIGN FOR BISPECTRA

Various practical considerations arise in the estimation of a bispectrum. These include: sampling interval, resolution, prefiltering, tapering and stability among other things. All the old questions and difficulties of the second-order case arise and some new ones as well.

2h. ALIASING

Suppose that the continuous series X has bispectrum $\tilde{S}(\lambda,\mu)$. Suppose that the values $X(t)$, $t = 0,1,...$ will be analyzed. The bispectrum of X as a discrete time series relates to the continuous one via

$$S(\lambda,\mu) = \sum_{j,k} \tilde{S}(\lambda+2\pi j, \mu+2\pi k) \tag{2.1}$$

Further complications arise because $\tilde{S}$ has the symmetry properties

$$\tilde{S}(\lambda,\mu) = \tilde{S}(\mu,\lambda) = \tilde{S}(\lambda,-\lambda-\mu) .$$

The fundamental domain over which s needs to be estimated reduces to the triangle with vertices $(0,0)$, $(\pi,0)$, $(2\pi/3,2\pi/3)$. All other bifrequencies are reflected into the domain.

Expression (2.1) makes it clear that one needs to take the sampling interval sufficiently small that not much bispectral mass lies outside the fundamental domain and that, where possible, one wants to lowpass filter X to remove bispectral mass outside that domain *prior* to sampling the series.

2i. PREWHITENING AND TAPERING

If the series X has zero mean and if

$$d^T(\omega) = \sum_t X(t)h^T(t) \exp\{-i\omega t\}$$

denotes the FT of tapered values, then

$$ave\ \{d^T(\omega)d^T(\nu)d^T(-\omega-\nu)\}$$

$$= \int \int H^T(\alpha)H^T(\beta)H^T(-\alpha-\beta)S(\omega-\alpha,\nu-\beta)d\alpha\, d\beta . \tag{2.2}$$

This last expression makes it clear that tapering may be used to increase the resolution of a bispectral estimate (decrease the interference between neighboring bifrequencies). If the bispectrum has substantial fall-off, tapering may be absolutely essential.

Prewhitening may also be of great importance. If the series X is passed through a linear filter with transfer function A, then the bispectrum of the resulting series is given by

$$A(\omega)A(\nu)A(-\omega-\nu)S(\omega,\nu) .$$

If this last bispectrum is more nearly constant (white) than $S(\omega,\nu)$ itself, then expression (2.2) may be expected to be more nearly proportional to it. After correcting for the effect of prefiltering, i.e. dividing the estimate of the bispectrum of the filtered series by $A(\omega)A(\nu)A(-\omega-\nu)$, a less biased estimate may be expected to have been obtained.

2j. BIOCOHERENCY AND BIOCOHERENCE

In dealing with the bispectrum it is often convenient to consider the bicoherency

$$S(\omega,\nu)/\sqrt{S(\omega)S(\nu)S(\omega+\nu)} \tag{2.3}$$

whose magnitude is not altered by linear filters, and which may be estimated by

$$\hat{S}(\omega,\nu)/\sqrt{\hat{S}(\omega)\hat{S}(\nu)\hat{S}(\omega+\nu)} . \tag{2.4}$$

The large sample variance of the estimate (2.4) is given by $T/4\pi L$ if $\hat{S}(\omega,\nu)$ is constructed in the manner of Section 18g and if the power spectral estimates are formed with the same bandwidth, $2\pi L/T$, as $\hat{S}(\omega,\nu)$. (In consequence their stability is much greater.) The expression $T/4\pi L$ may be used to decide on values for T and L.

The estimate (2.4) is further complex Gaussian with mean (2.3) and the indicated variance. In the case that (2.3) is 0, the mod-squared of (2.4), (the sample bicoherence), will be approximately exponential with mean $T/4\pi L$. This result may be used to examine the significance of bicoherence estimates.

2k. DESIGN FOR POINT PROCESSES

In some circumstances, one has some control over the collection of point process data. One may be able to select the length of the time period, T, over which the data is collected, and thereby have some effect on $N(T)$ the number of points observed. One may be able to increase or decrease the rate of the process, which also affects the number of points observed. In some situations one may even be able to choose the particular process, e.g. Poisson, from which inputs to a system are selected.

2l. THE POWER SPECTRUM CASE

Suppose that the power spectrum, $S(\omega)$, of the point process N is estimated by: (1) tapering the data with $h^T(t) = h(t/T)$, (2) Fourier transforming, (3) forming the periodogram, and (4) smoothing the periodogram with $W^T(\alpha) = B_T^{-1}W(B_T^{-1}\alpha)$. (where $W(\alpha)$ is a weight function and B_T is a bandwidth parameter). For this estimate one has

$$var\ \hat{S}(\omega) \sim (B_T)^{-1}2\pi \int W(\alpha)^2 d\alpha \int h(t)^4 dt \left(\int h(t)^2 dt\right)^{-2} S(\omega)^2 ,$$

(See Brillinger (1972)). This relationship is the same as the one for ordinary time series. It makes it clear what one wants: T large, and both W and h nearly constant on their supports. From bias considerations one knows that one wants the spectrum S near constant, B_T small, and W and the Fourier transform of h dying off rapidly.

2m. THE BIVARIATE CASE

Suppose that one has the bivariate point process $\{M,N\}$ and either: N is the output of a linear system with input M or one is interested in predicting the process N from the process M in a linear fashion. For both of these problems one is lead to consider the estimate $\hat{A}(\omega) = \hat{S}_{NM}(\omega)/\hat{S}_{MM}(\omega)$ of the transfer function $A(\omega) = S_{NM}(\omega)/S_{MM}(\omega)$. If estimates are constructed as in the previous section the asymptotic variance of $\hat{A}(\omega)$ is

$$(B_T T)^{-1}2\pi \int W(\alpha)^2 d\alpha \int h(t)^4 dt \left(\int h(t)^2 dt\right)^{-2} (1-|R(\omega)|^2) s_{NN}(\omega)/\hat{s}_{MM}(\omega) .$$

It is clear that the behaviors indicated as desirable for B_T, T, W, h in the previous section remain desirable here. It is further apparent that one wants: the coherence $|R|^2$ to be near 1, $S_{NN}(\omega)$ to be small, and $\hat{S}_{MM}(\omega)$ to be large.

These considerations refer to the estimation of A at a single frequency. If one considers the problem of the estimation of the whole course of the function A, then there are arguments suggesting that one should arrange for the input spectrum to be constant (at as large a value as is possible.) This is the case for a Poisson input.

One final remark relates to bias. If the process N is a delayed version of the process M then, as was the case with ordinary time series, it can be absolutely essential that the two data stretches be realigned. Otherwise the estimate $\hat{S}_{NM}$ can be attenuated to 0 by the spiralling.

3. COMPUTATION

The estimates and techniques discussed mean little if they are not actually evaluated in circumstances of interest. These days exceedingly large and complex data sets are collected routinely. In other circumstances it may be necessary to work in real time. Computational considerations can be very important.

3a. THE FFT's

In much of what has gone before, the FT

$$\sum_{t=0}^{T-1} X(t) \exp\{-i\omega t\} ,$$

has been basic. Typically its values are required at frequencies of the form $\omega = 2\pi u/U$, $u=0,...,U-1$ with U an integer $\geq T$. If $X(t)$ is set equal to 0 for $t = T,...,U-1$, then what is being sought is the discrete Fourier transform

$$\sum_{t=0}^{u-1} X(t) \exp\{-i2\pi ut/U\} \tag{3.1}$$

$u=0,...,U-1$. Ignoring the operations required to evaluate the complex exponential, the direct evaluation of (3.1) requires U^2 multiplications. If U is large, this can require much computing time. Further, round-off error may become substantial. (Gentleman and Sande (1966)).

Luckily, fast Fourier transform algorithms exist to reduce the number of operations required. In the most often quoted case, $U = 2^n$ and the required number of multiplications is $O(nU) = O(U \log_2 U)$. One algorithm for achieving this result is given in Cooley and Tukey (1965). See also Gentleman and Sande (1966). More recently algorithms leading the FFT's symmetrics have been presented that involve only $O(U)$ multiplications, see Winograd (1978), Morgera (1980).

The multidimensional case involves repeated Fourier transforming with respect to the various time arguments. This may be done via repeated uni-dimensional algorithms. Optical systems may sometimes be used to advantage. Turpin (1981) indicates how a large uni-dimensional transform (10^6 or 10^7 points) may be taken with a modest optical system, by stacking successive segments of the series.

Tapering a data set merely involves replacing the value of $X(t)$ by $h^T(t)X(t)$ in the uni-dimensional case. This doesn't add much in the way of computations. The support of h^T corresponds to the domain of observation. If this domain is irregular, as often happens in the spatial case, h^T handles the difficulty quite directly, if its transform can be kept manageable. The FT may then be computed over a regular region, with h^T equal 0 where no datum is available.

3b. FILTERING AND COMPLEX DEMODULATION

Filtered series and, in particular, complex demodulates may often be usefully computed via an FFT. Suppose one has a stretch of a series X. Suppose one wishes the series Y corresponding to passing X through an linear time invariant filter with transfer function A. If $d_X^T(\omega)$ denotes the FT of X, then that of Y will be approximately $A(\omega)d_X^T(\omega)$. This suggests determining $Y(t)$ as

$$U^{-1} \sum_{s=0}^{U-1} \exp\{i2\pi su/U\} A\left(\frac{2\pi s}{S}\right) d_X^T\left(\frac{2\pi s}{S}\right) \tag{3.2}$$

with U taken to be sufficiently large that aliasing difficulties do not arise. The FT values $d_X^T(2\pi u/U)$ may be evaluated by an FFT, having padded the stretch $X(t)$, $t=0,...,T-1$ with $U-T$ zeroes. The inverse transform of (3.2) may be evaluated via an FFT. (See Gentleman and Sande (1966).) Programs may be found in Digital Signal Processing Committee (1979).

Suppose one wishes the complex demodulate of the series X at frequency λ. Let $A(\omega)$ be near 1 for ω near λ and near 0 elsewhere. The demodulate is then given, approximately by expression (3.2) times $\exp\{-i\lambda t\}$. Being slowly varying it will not need to be computed for every value of t, e.g. one might compute it for $t=0,\Delta,2\Delta,(J-1)\Delta$. The inverse transform in (3.2) can therefore be arranged to be of size U/J only.

3c. THE POINT PROCESS CASE

The point process FT is

$$\sum_{j=1}^{N(T)} \exp\{-i\omega\tau_j\}\,. \tag{3.3}$$

In some circumstances it is convenient to approximate this FT by that of an ordinary discrete time series $X(\ell\Delta)$, $\ell = 0,1,\ldots$, where Δ is small and $X(\ell h) = 1$ for ℓ the integer part of τ_j/Δ and $X(\ell\Delta) = 0$ otherwise. This series will be 0 most of the time. The difficulty is that now a transform of length at least T/Δ is required and this may be very large. In essence one is here choosing to analyze the series $X(\ell\Delta) = N((\ell - \frac{1}{2})\Delta, (\ell + \frac{1}{2})\Delta]$.

The point process N may be replaced by ordinary time series in other ways as well. French and Holden (1971) suggest forming

$$X(t) = \sum_j [\sin \pi(t-\tau_j)]/\pi(t-\tau_j)$$

at $t = 0,1,\ldots$. An FFT may then be employed.

Yet another approach would proceed as follows: set

$$\tau_j = m_j h + \delta_j \tag{28.1}$$

with m_j integral and $|\delta_j| \le h/2$. Then may be written

$$\sum_j \exp\{-i\omega m_j h\}(1 + i\omega\delta_j - \omega^2\delta_j^2/2 + \cdots)$$

$$= \sum_m e^{-i\omega h m} k_0(m) + i\omega \sum_m e^{-i\omega h m} k_1(m)$$

$$- \frac{\omega^2}{2} \sum_m e^{-i\omega h m} k_2(m) + \cdots \tag{28.2}$$

where

$$k_0(m) = \sum_{m_j = m} 1$$

$$k_1(m) = \sum_{m_j = m} \delta_j$$

$$k_2(m) = \sum_{m_j = m} \delta_j^2$$

Expression (28.2) may now be calculated by FFT's.

If we want to use ω's up to ω_{MAX} and require to approximate $\exp\{i\omega\delta\}$ to 1% then we can use the first two terms for

$$|\omega\delta| \le \omega_{MAX} h/2 \le .14$$

To meet the same constraint with the first three terms would require

$$\omega_{MAX} h/2 \le .39$$

The ratio of numbers of intervals is $.39/.14 = 2.7$ which is greater than the ratio of numbers of FFT's, namely 1.5, so that going as far as the third term may easily be worthwhile. Similar calculations give

terms = FFT's	max $\lvert\omega\delta\rvert$	ratio ratio
2	.14	14
3	.39	7
4	.7	5.7
5	1	5

Suggesting it may pay to go to 4 or 5 terms.

If we use the approximation

$$e^{i\theta} \approx .9994 + .9567i - .4853\,\theta^2$$

which is good to $|\text{error}| \le .01$ for $-.6 \le \theta \le .6$ instead of the leading terms of the Taylor series, we can use the first two terms over about 4 times the interval. Economizing (see e.g. Anramonitz and Stegûn 1964, p. 791) the higher-order approximations over well chosen intervals will allow corresponding gains.

It will be seen shortly that, in the point process case, it may be more effective to do some computations on the time side, before proceeding to the frequency domain.

3d. THE SECOND-ORDER CASE

Consider the problem of forming estimates of power and cross-spectra, transfer functions and coherence for ordinary time series and point processes.

3e. THE POWER SPECTRUM OF AN ORDINARY TIME SERIES

The steps in one power spectrum estimation direct procedure include:

1. preanalysis (e.g. trend removal, prefiltering, sampling, prewhitening
2. tapering,
3. padding with zeros,
4. Fourier transformation,
5. mod-squaring (to obtain the periodogram),
6. smoothing (the periodogram), (in time or frequency).

The estimate at frequency $\omega = 2\pi u/U$ may be written

$$\hat{S}(\omega) = \sum_{u} W_u\, I^T\left(\omega + \frac{2\pi u}{U}\right) \tag{3.1}$$

where the W_u are weights summing to 1, where the data has been padded by adding $U-T$ zeros, where

$$I^T(\omega) = \left|\sum_t e^{-i\omega t} h^T(t) X'(t)\right|^2 \Big/ 2\pi \sum_t h^T(t)^2 \tag{3.2}$$

is the periodogram, and where X' denotes the series after preanalysis. Padding with zeros plays several roles. It allows an FFT requiring highly composite U to be employed, it avoids the unnecessary circularization of the data with the accompanying aliasing (discussed at the end of Section 6a), and it lets the periodogram values be as finely spaced as desired.

In some circumstances step 6. may be replaced by:

6'. inverse transform

7'. multiply by lag window

8'. transform

for an indirect estimation procedure. This makes the smoothing window choice a time -- rather than frequency -- side problem. (The computations may now need to be done in double precision.)

As one has the relationships

$$\hat{c}(v) = \frac{1}{T}\sum_{t=0}^{T-|v|-1} X'(t+v)X'(t)$$

$$= \frac{2\pi}{U}\sum_{t=0}^{U-1} I^T(2\pi u/U)\,\exp\{i2\pi vu/U\}$$

and

$$I^T(\omega) = \frac{1}{2\pi}\sum_{v}\hat{c}(v)\,\exp\{-i\omega v\}$$

proceeding in the last fashion is largely a matter of convenience.

The estimate (3.1) is based on smoothing the periodogram of the whole data stretch, as a function of frequency. There are circumstances in which it is more useful to proceed by computing the periodogram, $I^T(\omega,\ell)$ of the ℓ-th segment of the data as a function of ℓ, to form an estimate

$$L^{-1}\sum_{\ell=1}^{L} I^T(\omega,\ell)\ .$$

If the segments overlap, this is called a shingled estimate. In some circumstances it may be convenient to weight the segments unequally.

This last estimate has the form of the narrow-band-pass filtered estimate or of the complex-demodulate estimate. If $d^T(\omega,\ell)$ denotes the FT of the ℓ-th segment, then $\exp\{i\omega t\}d^T(\omega,\ell)$ for t in the ℓ-th segment provides a band-pass filtered (at frequency ω) version of the series.

In some circumstances one will smooth periodograms in both the time and frequency domains.

Each of the above estimates are quadratic functions of the observations. Employing an FFT can reduce computation time and round-off error.

3f. THE CROSS-SPECTRUM

Suppose the stretch $\{X(t),Y(t)\}$, $t=0,\ldots,T-1$ of bivariate series is available. The spectrum estimate analogous to (3.1) -- (3.2) is given by

$$\hat{S}_{xy}(\omega) = \sum_{u} W_u I^T_{xy}\left(\omega + \frac{2\pi u}{U}\right)$$

where

$$I^T_{xy}(\omega) = \left[\sum_t e^{-i\omega t}h^T(t)X'(t)\right]\left[\sum_t e^{i\omega t}h^T(t)Y'(t)\right]/2\pi\sum_t h^T(t)^2\ .$$

X' and Y' denote the results of preanalyzing the series (removing trends, prewhitening, re-aligning, etc.) Different tapers may be employed for X' and Y'. (Notice the complex conjugate implied by $-i$ in the final exponent.) It is clear that an FFT may prove useful.

As in the previous section, it may make sense to form an estimate by smoothing the cross-periodograms of segments of the series,

$$L^{-1} \sum_{\ell=1}^{L} I_{XY}^{T}(\omega,\ell) .$$

Once the estimates $\hat{S}_{XX}$, $\hat{S}_{XY}$, $\hat{S}_{YY}$ are available, estimates of the transfer function, coherence and residual spectrum may be formed. The impulse response may be estimated by an expression of the form

$$\hat{a}(u) = P^{-1} \sum_{p=0}^{P} \exp\{i2\pi up/P\}\hat{A}(2\pi p/P)k^{P}(p)$$

with $k^{P}(p)$ a window function and P a positive integer. When the basic series involved are continuous in time, insertion of a suitable window function can be crucial. Once again an FFT generally proves useful.

3g. THE POWER SPECTRUM OF A SPATIAL PROCESS

Let

$$d^{T}(\lambda,\mu) = \sum_{x\,y} \exp\{-i(\lambda x + \mu y)\}h^{T}(x,y)X(x,y)$$

denote the FT of a piece of the spatial array X, with taper inserted. The periodogram of this (tapered) data is given by

$$I^{T}(\lambda,\mu) = |d^{T}(\lambda,\mu)|^2/2\pi \sum_{x\,y} h^{T}(x,y)^2 .$$

To obtain an estimate of the power spectrum of the series X, one simply smooths this periodogram as a function of λ,μ. The data may have been detrended prior to computing the (repeated) FT.

Naturally, in an often used alternate approach, periodograms might be computed for (tapered) segments of the data and these periodograms averaged together. This might be useful, for example, if there was a very large volume of data.

In the isotropic case, the power spectrum is a function of $\lambda^2 + \mu^2$ only. This means that a further smoothing may be carried through, with a consequent increase in stability. Specifically, periodogram values at frequencies λ,μ with similar values of $\lambda^2 + \mu^2$ may be averaged together. One interesting effect that occurs is that estimates with larger $\lambda^2 + \mu^2$ are more stable, relatively, because more periodogram values may be averaged together. Brillinger (1970) provides a proof of this last.

3h. THE POWER SPECTRUM OF A POINT PROCESS

The power spectrum of a point process may be estimated at frequency $\omega \neq 0$ by smoothing the periodogram

$$I^{T}(\omega) = (2\pi T)^{-1}\left|\sum_{j} \exp\{-i\omega\tau_j\}\right|^2 .$$

Alternatively it may be estimated by averaging together periodograms based on different stretches of the process.

In a number of circumstances it turns out to be faster to form the estimate in an indirect fashion, following definition (11.1). Specifically one estimates $p(u) = Prob\{dN(t+u) = 1$ and $dN(t) = 1\}$ by

$$\hat{p}(u) = \sum_{j \neq k} \#\{|\tau_j - \tau_k - u| < \beta/2\} / \beta T \tag{3.3}$$

for some binwidth β. Then one computes

$$\hat{S}(\omega) = \frac{\hat{p}}{2\pi} + \frac{1}{2\pi}\sum_j \exp\{-i\omega j\beta\} w^T(j\beta)\left[\hat{p}(j\beta) - \hat{p}^2\right] \tag{3.4}$$

where $\hat{p} = N(T)/T$ is an estimate of the rate of the process and where w^T is a smoothing window. The estimate (3.3) may be computed rapidly, as it simply involves counting. Expression (3.4) may be computed via an FFT.

3i. THE HIGHER-ORDER CASES

In estimating higher-order spectra, such as the bi- or tri-spectrum, computational considerations can become crucial.

3j. THE BISPECTRUM

Set

$$d^T(\omega) = \sum_t \exp\{-i\lambda t\} h^T(t) X'(t)$$

where $X'(t)$ denotes the preanalyzed values. The third-order periodogram is given by

$$I^T(\omega,\nu) = d^T(\omega) d^T(\nu) d^T(-\omega-\nu)/(2\pi)^2 \sum_t h^T(t)^3$$

The bispectrum may be estimated by averaging this function. In carrying out this averaging, the periodicity and symmetry properties should be preserved. These include:

$$\begin{aligned}
I^T(\omega,\nu) &= I^T(\nu,-\omega-\nu) &&= I^T(-\omega-\nu,\omega) \\
I^T(\omega,\nu) &= I^T(\omega+2\pi,\nu) &&= I^T(\omega,\nu+2\pi) \\
\overline{I^T(\omega,\nu)} &= I^T(-\omega,-\nu) &&.
\end{aligned}$$

Alternatively, the bispectrum may be estimated by averaging third-order periodograms evaluated at the same bifrequency (ω,ν), but based on different stretches of the data or by averaging $X(t,\omega)X(t,\nu)X(t,-\omega-\nu)$ as a function of t where $X(t,\omega)$ denotes the result of narrow-band-pass filtering the series X at frequency ω. [This filtering may be carried through by complex demodulation and an FFT if desired.]

3k. SPECTRA OF ORDER K

Suppose that a stretch of all J vector-valued series $\{X_j(t); j=1,\dots,J\}$ is available for analysis. Set

$$d_j^T(\omega) = \sum_t \exp\{-i\lambda t\} h^T(t) X_j'(t)$$

with X_j' denoting the pre-analyzed j-th series. The periodogram of order K, corresponding to components $a_1, \dots, a_K$ (selected from $1,\dots,J$) is given by

$$I^T_{a_1 \cdots a_K}(\omega_1, \dots, \omega_K)$$

$$= d^T_{a_1}(\omega_1) \cdots d^T_{a_k}(\omega_K)/(2\pi)^{K-1} \sum_t h^T(t)^K$$

It is convenient to make this definition by means of a dummy argument ω_K given by $-\omega_1 - \dots - \omega_{K-1}$. Doing this makes the symmetries and computational approaches stand out more clearly.

The corresponding spectrum is now estimated by smoothing the periodogram, e.g. by

$$\sum_{u_1 \cdots u_K} W_{u_1 \cdots u_K} I^T \left(\omega_1 + \frac{2\pi u_1}{U}, \ldots, \omega_K + \frac{2\pi u_K}{U} \right)$$

where the weights sum to 1, vanish if any $u_k = 0$, and are symmetric in $u_1, \ldots, u_K$. (It is here assumed that the ω_j have the form $2\pi(\text{integer})/U$ and that discrete FT's of length U have been formed.) This gives the estimate appropriate symmetry and periodicity properties.

Estimates involving the averaging of periodograms based on separate stretches and based on narrow-band-pass filtered series are also available. Brillinger (1965) and Brillinger and Rosenblatt (1967) are relevant references.

4. IDENTIFICATION OF NONLINEAR SYSTEMS BY FREQUENCY METHODS

Section D indicated how linear time invariant systems could be examined by spectrum methods. This section considers some nonlinear, but time-invariant, systems.

4a. INSTANTANEOUS NONLINEARITY

Consider a system described by

$$U(t) = \sum_u a(t-u)X(u)$$

$$Y(t) = g[U(t)] + e(t)$$

with e a noise series, $a(\cdot)$ the impulse response of an linear time invariant filter and with $g(\cdot)$ a function of a single variable. A surprising result, first noted by Bussgang (1952) for the case of polynomial g, is that if the input X is Gaussian stationary, then

$$s_{yx}(\omega) = cA(\omega)s_{xx}(\omega)$$

with c a constant. In other words if one identifies the system, by cross-spectral analysis, as if it were linear, then the transfer function obtained is proportional to that of the linear part of the system. Korenberg (1973) goes on to consider the case of several linear filters and several instantaneous nonlinearities. Brillinger (1977) presents some statistical details and considers general $g(\cdot)$.

The essence of this result is that if a system is identified by cross-spectral analysis with Gaussian input, then the resulting transfer function can have a simple interpretation in a much broader class of instances than one might have anticipated.

4b. QUADRATIC SYSTEMS

A natural generalization of an linear time-invariant system is an bilinear time-invariant system, a system with two inputs that obeys superposition in each input separately. Such a system is bilinear, not linear, and can behave in non-linear ways, particularly when we connect the same $X(t)$ to *both* inputs. The resulting quadratic time-invariant system, with one input and one output, can be defined abstractly by the property $C[X+Y] - C[X-Y]$ is a bilinear system with inputs X and Y. If this holds for all $X(t)$ and $Y(t)$, then C is a quadratic time-invariant system, and will have each of the representations for such. Let us turn to this case. Suppose that

$$Y(t) = \sum_{u,v} b(t-u, t-v)X(u)X(v) + e(t) \tag{4.1}$$

with e noise. Let

$$B(\omega,\nu) = \sum_{u,v} \exp\{-i(\omega u + \nu v)\} b(u,v)$$

denote the bitransfer function. Suppose that X is zero mean, Gaussian, stationary with power spectrum $S_{XX}(\omega)$. Then the cross-bispectrum of Y with X satisfies

$$S_{XXY}(\omega,\nu) = 2B(-\omega,-\nu)S_{XX}(\omega)S_{XX}(\nu) \ . \tag{4.2}$$

This result was recognized by Tick (1961).

An estimate of B may be formed once estimates of S_{xx} and S_{xxy} have been constructed. The bi-impulse response function b, may be estimated by Fourier transforming $\hat{B}$, using convergence factors as necessary.

If the system (4.1) is extended to contain a linear term, $\sum_u a(t-u)X(u)$, the relationship (4.2) continues to hold. The transfer function A may be estimated by cross-spectral analysis.

Hung *et al* (1979) determine the first and second degree kernels of the human pupillary system by this technique.

4c. *COSINUSOIDAL INPUT*

An informative technique for the examination of some systems is to take a pure cosinusoid $\alpha\cos(\beta t+\gamma)$ as input. If the system is linear time-invariant, then frequency β alone will appear in the output. If it is quadratic time-invariant, then the frequencies β, 2β will appear. If it is time-invariant polynomial of order L, then frequencies $\beta, 2\beta, \ldots, L\beta$ will appear.

On occasion subharmonics $\beta/2$, $\beta/3$,... may appear, as with the edge waves in Section 7f. This is an indication of a less simple sort of nonlinearity. Further information concerning the system may be found by taking a pair $\alpha_1\cos(\beta_1 t+\gamma_1) + \alpha_2\cos(\beta_2 t+\gamma_2)$ of cosinusoids as input. Subcombination frequencies $\omega = (\epsilon_1\beta_1+\epsilon_2\beta_2)/n$, $n = 2,3,\ldots$; $\epsilon_1,\epsilon_2 = \pm 1, \pm 2,\ldots$ will be induced by some (non-polynomial) non-linear systems. The relationship of β_1 and β_2 to the natural frequencies of the system becomes very important here.