Methods for estimating directional wave spectra from multi-element arrays

by Russ E. Davis[1] and Lloyd A. Regier[2]

1. Introduction

The Fourier wavenumber-frequency spectrum of a scalar variable η, with zero mean, is defined

$$S(\mathbf{k},\omega) = (2\pi)^{-D-1} \int_{-\infty}^{\infty} d\boldsymbol{\zeta} \int_{-\infty}^{\infty} d\tau \exp(-i\mathbf{k}\cdot\boldsymbol{\zeta} + i\omega\tau) C(\boldsymbol{\zeta},\tau) , \tag{1}$$

where D is the dimensionality of the $\boldsymbol{\zeta}$ space and C is the covariance, taken to be homogeneous and stationary,

$$C(\boldsymbol{\zeta},\tau) = \langle \eta(\mathbf{x},t)\, \eta(\mathbf{x}+\boldsymbol{\zeta},t+\tau) \rangle , \tag{2}$$

where $\langle\rangle$ represents the average value. The wavenumber-frequency spectrum is of fundamental importance in describing the distribution of variance according to space and time scales and finds widespread use in geophysics both as a description of variability and in using observations to test dynamical theories.

Estimation of wavenumber-frequency spectra from real data involves overcoming the same two fundamental problems encountered in estimating frequency spectra from time series (see Jenkins and Watts, 1968), namely, statistical reliability and resolution. In essence, statistical reliability is limited by the quantity of data available from which the average in (2) is estimated. Resolution, on the other hand, is limited by the range of $\boldsymbol{\zeta}$ and τ over which $C(\boldsymbol{\zeta},\tau)$ in (1) is known. For example, in frequency spectrum analysis the fundamental maximum frequency resolution is the order of the inverse of the record length while knowing the covariance only at discrete values of τ leads to the loss of resolution known as "aliasing."

Measurement of wavenumber-frequency spectra is often accomplished through the use of arrays of instruments which observe η at a finite number of locations $\mathbf{x}_n$. While the statistical reliability and frequency resolution of such spectral estimates

1. University of California, San Diego, Scripps Institution of Oceanography, La Jolla, California, 92093, U.S.A.

2. Meteorology Department, Massachusetts Institute of Technology, Cambridge, Massachusetts, 02139, U.S.A.

are often limited by the time over which observations are made, this limitation is usually less serious than the poor wavenumber resolution resulting from the small number of elements in the arrays and the associated limited numbers of locations in lag space (the ζ space in (2)) at which the covariance is known. The purpose of this paper is to examine ways of designing processing methods which maximize the resolution obtained from arrays with few elements without seriously degrading statistical reliability.

In discussing measures of performance, emphasis will be placed on two-dimensional fields and on variables which obey a single mode dispersion relation such as that describing linear surface waves. Generalization to fields without known dispersion relations and to three-dimensional Fourier spectra ($D = 3$ in (1)) are straightforward.

There are three fundamentally different kinds of spectral analysis methods which find application in processing array data. "Model fitting" involves prescribing in advance a form for the spectrum in terms of an analytic relation (the model) involving adjustable parameters. The other two classes of methods do not involve prescription of the form of the spectrum and are, therefore, model independent. In this paper only model independent methods will be discussed. Model fitting can be used to good advantage in spectral analysis (e.g. super-resolution, Munk and Hasselman, 1964, and the directional spectral estimates by Munk *et al.*, 1963, and Longuett-Higgins *et al.*, 1963) but the efficacy of such efforts depends strongly on how well the trial forms of the model match the true signal. Since there is no *a posteriori* way of testing this, model fitting must be used with caution.

The remaining two methods, both model independent, may be categorized as *a priori* and data adaptive. The former, more familiar in time series analysis, is designed to perform well according to criteria which are specified without reference to the data to be used. The latter, on the other hand, are tailored specifically to the data actually observed and cannot be fully specified until the data are taken.

In what follows the general structure of estimates of wavenumber-frequency spectra and some measures of performance will be discussed. Then, attention will be turned to the design of efficient arrays and to the design of *a priori* analysis schemes. Then data-adaptive techniques will be discussed and a new data-adaptive spectral estimation scheme, designed for analysis of continuous spectra rather than detection of signals in the presence of noise, will be presented. Finally, comparisons of the various methods will be presented and recommendations given.

A more complete and detailed discussion of these topics may be found in Regier (1975).

2. Spectral analysis

This paper is primarily concerned with estimation of wavenumber-frequency (*WF*) spectra using data from observational arrays which are well sampled in time.

An observational array which is well sampled in time is taken to be a collection of N_e elements at locations $\mathbf{x}_n$ which provide records of $\eta(\mathbf{x}_n,t)$ sufficiently long and well sampled that the cross-spectra of all element pairs can be computed with essentially perfect frequency resolution, including the effects of aliasing. This is an idealized case but, in general, it is possible to achieve frequency resolution which is so much better than the achievable wavenumber resolution as to be essentially perfect. It is, however, important to account for the fact that there will inevitably be errors in the estimates of these cross-spectra arising from, if nothing else, the limited statistical reliability of any real observation. Thus, if the spatially lagged frequency cross-spectrum is defined by

$$Q(\boldsymbol{\zeta},\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} d\tau \,<\eta(\mathbf{x},t)\,\eta(\mathbf{x}+\boldsymbol{\zeta},t+\tau)>\exp(i\omega\tau) \tag{3}$$

the observer can compute, for each ω, N estimates $\hat{Q}$ of $Q(\boldsymbol{\zeta}_k,\omega)$ at the N (distinct) lags $\boldsymbol{\zeta}_k$ separating all possible pairs of elements. The data are then N values of

$$\hat{Q}_{nm}(\omega) = \hat{Q}(\boldsymbol{\zeta}_{nm},\omega) = Q(\mathbf{x}_n - \mathbf{x}_m,\,\omega) + q_{nm} \tag{4}$$

where $\mathbf{x}_n$ and $\mathbf{x}_m$ are the locations of the array elements, $\boldsymbol{\zeta}_{nm} = \mathbf{x}_n - \mathbf{x}_m$, and q is the error or noise.

According to (1) the *WF* spectrum is related to Q by

$$S(\mathbf{k},\omega) = (2\pi)^{-D}\int d\boldsymbol{\zeta}\, Q(\boldsymbol{\zeta},\omega)\exp(-i\mathbf{k}\cdot\boldsymbol{\zeta}) \tag{5a}$$

or the inverse relation

$$Q(\boldsymbol{\zeta},\omega) = \int d\mathbf{k}\, S(\mathbf{k},\omega)\exp(i\mathbf{k}\cdot\boldsymbol{\zeta})\,. \tag{5b}$$

Since, according to (5a), the *WF* spectrum is a linear operation on Q it is natural to seek an estimate, $\hat{S}$, of S which is a linear combination of the available data, i.e.

$$\hat{S}(\mathbf{k},\omega) = \sum_{n,m} \alpha_{nm}\,(\mathbf{k},\omega)\,\hat{Q}_{nm}(\omega)\,, \tag{6}$$

where, to make $\hat{S}$ real valued, α_{nm} and α_{mn} are complex conjugates. All estimators discussed in this paper are of the form (6) although the methods of choosing the weights, α, are different; the authors are unaware of any truly nonlinear methods of estimating spectra. Since each frequency band is treated separately the notation will be simplified by dropping ω except when it is required for clarity. In practice the estimation procedure is carried out separately for each frequency band for which $\hat{Q}$ is computed.

The utility of the estimate (6) should be measured by two interrelated criteria which might be called noise-rejection (minimal effect of q) and wavenumber resolution. The nature of the noise, q, depends on its origin and the method of computing the cross-spectral estimates. If the array elements are well calibrated it will usually be the case that the principal noise will be due to statistical uncertainty resulting

from the finite quantity of data. When this is the case and when conventional computational methods (see Jenkins and Watts, 1968) are used, the cross-spectral estimates are unbiased so that

$$<q_{nm}> = 0\,, \qquad <\hat{Q}_{nm}> = Q_{nm}\,.$$

The variance of the spectral estimate about its mean value,

$$\sigma^2 = <[\hat{S}(\mathbf{k}) - <\hat{S}(\mathbf{k})>]^2> = \sum_{n,m}\sum_{i,j} \alpha_{ij}(\mathbf{k})\alpha_{nm}(\mathbf{k}) <q_{ij}\, q_{nm}>\,, \tag{7}$$

depends both on processing method and the noise. The spectral estimate is biased by limited wavenumber resolution. Thus from (5)

$$\begin{aligned} <\hat{S}(\mathbf{k})> &= \sum_{n,m} \alpha_{nm}(\mathbf{k})\, Q_{nm} \\ &= \sum_{n,m} \alpha_{nm}(\mathbf{k}) \int d\mathbf{k}'\, S(\mathbf{k}') \exp\,(i\mathbf{k}' \cdot [\mathbf{x}_n - \mathbf{x}_m]) \\ &= \int d\mathbf{k}'\, S(\mathbf{k}')\, W(\mathbf{k},\mathbf{k}')\,. \end{aligned} \tag{8}$$

The quantity

$$W(\mathbf{k},\mathbf{k}') = \sum_{n,m} \alpha_{nm}(\mathbf{k}) \exp\,(i\mathbf{k}' \cdot [\mathbf{x}_n - \mathbf{x}_m]) \tag{9}$$

will be referred to as the wavenumber window and measures how much an estimate of the spectrum at wavenumber $\mathbf{k}$ is affected by the variance density at wavenumber $\mathbf{k}'$ in exactly the way that the sinc function frequency window encountered in conventional time series analysis measures variance spreading in frequency space.

High resolution *WF* spectra are obtained when the window W approximates a delta-function centered at $\mathbf{k} = \mathbf{k}'$. The *a priori* analysis methods involve selecting α_{nm} which produce a window which best approximates such a delta-function according to some prescribed criteria. Data adaptive methods, on the other hand, do not strive directly to approximate a delta-function but rather attempt to make $W(\mathbf{k},\mathbf{k}')$ small at those values of $\mathbf{k}'$ (other than $\mathbf{k}$) where $S(\mathbf{k}')$ is significant.

Both methods are influenced by the noise q. In the *a priori* methods it is frequently the case that the "optimal" window is achieved by using large values of α, particularly when there are nearly redundant lags. But, since the noise susceptibility σ^2 of (7) increases as the square of these weights, this can seriously degrade the resulting estimate. Although somewhat less obvious from their development, data adaptive methods can suffer from very much the same problem. Unfortunately, accurate estimates of the noise matrix $<q_{ij}\, q_{nm}>$ are difficult to make. Perhaps the most widely applicable case is when the noise is of purely statistical origin. In this case some estimate of the noise matrix can be made by assuming the field η to be Gaussian. Making such estimates of the noise statistics is discussed in the Appendix but the complexity of this procedure is not warranted by the gain in reliability of the estimate. It is found in practice that when the statistics are reasonably stable, good rejection of statistical noise is obtained from processing schemes which were de-

signed to reject noise which is uncorrelated between the different cross-spectral estimates and has random phase, that is

$$<q_{nm}\, q_{ij}> \begin{cases} = 0 \text{ if } n \neq j \text{ or } m \neq i \\ = \sigma_0^2 \text{ if } n=j \text{ and } m=i. \end{cases}$$

As will be seen ((10b) below) this leads to a measure of σ^2 which is influenced by the magnitude of the weights, α. The success of this gross simplification is, no doubt, due to the fact that, regardless of the precise nature of the noise, the cancellation associated with large weights of opposite sign produces noise amplification.

3. Array design

In addition to the method of processing, the fidelity of *WF* spectral estimates is strongly dependent on the geometry and number of elements in the sampling array. Examination of (9) shows that the crucial feature is not so much the location of array elements as the location of the lags $\boldsymbol{\zeta}_{nm} = \mathbf{x}_n - \mathbf{x}_m$ at which the cross-spectrum $Q(\boldsymbol{\zeta},\omega)$ can be estimated. This has led to the "coarray" as a description of an array. As the array is the collection of locations in physical space, $\mathbf{x}_n$, where η is sampled, the coarray is the collection of locations, $\boldsymbol{\zeta}_{nm}$, in lag space where the cross-spectrum can be estimated. If there are N_e array elements there are N_e^2 lags in the coarray but N_e of these must be redundant samplings of zero lag. One half of the nonzero lags are reflections of each other, i.e. $\boldsymbol{\zeta}_{nm} = -\boldsymbol{\zeta}_{mn}$, so there are at most $N_e(N_e - 1)/2 + 1$ truly distinct lags. In practice it may be difficult to achieve this theoretical maximum in large arrays.

Some simplification in visualizing the wavenumber window $W(\mathbf{k},\mathbf{k}')$ can be achieved by using the weights β defined by

$$\alpha_{nm}(\mathbf{k}) = \beta_{nm}(\mathbf{k}) \exp\left(-i\mathbf{k} \cdot [\mathbf{x}_n - \mathbf{x}_m]\right).$$

In order to insure that the wavenumber window is real valued, β_{nm} must be the complex conjugate of β_{mn}. Then the window of (9) and the noise measure of (7) using the simplified noise statistics become

$$W(\mathbf{k},\mathbf{k}') = \sum_{n,m} \beta_{nm}(\mathbf{k}) \exp\left(i\boldsymbol{\zeta}_{nm} \cdot [\mathbf{k}' - \mathbf{k}]\right) \tag{10a}$$

$$\sigma^2 = \sum_{n,m} |\beta_{nm}(\mathbf{k})|^2\, \sigma_0^2 \tag{10b}$$

Truly optimal design of array geometries depends on establishing a criteria of merit for a given wavenumber window and deciding on the method of processing which will be employed. The process of array design then involves a trial-and-error evaluation of given array geometries in terms of their figure of merit. Short of this exhaustive procedure, some reasonable assessment of array performance can be made from an examination of the coarray. The characteristics to be considered are (a) the maximum extent of the coarray, which most strongly affects the sharpness

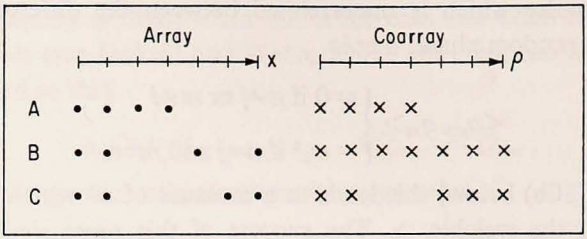

Figure 1. Examples of three linear arrays and the associated coarrays. Only the coarray elements for zero or positive ρ are shown. Array B, which has no redundant separations, has the largest coarray and consequently the highest resolving power.

of the wavenumber window around the central peak at $\mathbf{k} = \mathbf{k}'$, (b) the minimum separations in the coarray, which affect the minimum wavenumber difference $|\mathbf{k} - \mathbf{k}'|$ at which the window will have a secondary peak resulting in aliasing, and (c) the uniformity of the coarray, which will influence the size of the major "side-lobes" of the window at wavenumber separations smaller than the aliasing wavenumber.

Illustration of these general observations, which hold equally well in one or many dimensions, is given by considering the three linear arrays of four elements shown, together with their coarrays, in Figure 1. Array A has many redundant lags and, consequently, has a small coarray which spans only $-3 \leqslant \rho \leqslant 3$. Array B achieves the theoretical maximum size coarray (seven distinct lags) spanning $-6 \leqslant \rho \leqslant 6$. Coarray C also spans $-6 \leqslant \rho \leqslant 6$ but the coarray is quite nonuniform. On the basis of (a) above the widths of the central peaks of the windows for B and C should be comparable and narrower than that for A. All have minimum separations of unity and hence, according to (b), their windows should have large aliasing peaks at $k - k' = 0(2\pi)$; in fact, because all lags are multiples of unity the windows are exactly periodic with period 2π. Array C differs from B only in uniformity of the coarray; from (c) it is expected that B should have the smaller side lobes. These conclusions are demonstrated in Figure 2 whcih portrays the wavenumber windows for the three arrays. These windows were chosen to minimize

$$\int_{l=-\pi}^{\pi} W^2(k, k+l)dl$$

while maintaining $W(k,k) = 1$; only $0 < l < \pi$ is shown since the windows are symmetric about the end points.

Haubrich (1968) has carried out extensive experimentation to discover rules for design of good two-dimensional arrays. For various array geometries he found the window $W(k, k + l)$ which best approximated the function $F(1) = 1$ for $|l| < r$ and $F(l) = 0$ for $r < |l| < R$. He found that the best arrays were those with minimum separations of $0(2\pi/R)$ and the largest uniform coarrays; increasing the maximum lags at the expense of uniform coverage in lag space did not improve resolution.

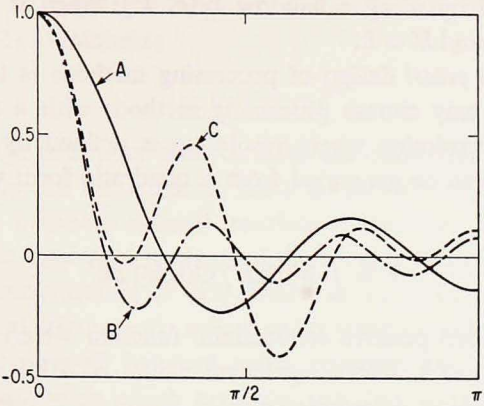

Figure 2. Examples of the resolving windows associated with the arrays shown in Figure 1. See the text for the criterion of optimization.

Interestingly, Haubrich (personal communication) finds that the practice of designing large arrays using regular geometric patterns is generally inferior to random selection of element locations, presumably because of the many redundant lags found in arrays which appear regular to the eye.

Array design influences noise rejection as well as resolution. Adding redundant lags does not influence resolution but does reduce noise sensitivity because the sample cross-spectra of these lags may be averaged together. This is particularly effective in reducing instrumental noise but is often not very effective in reducing statistical noise from sampling errors because the errors on the redundant lags are frequently correlated and, consequently, not much reduced by averaging (see Appendix).

The design of optimal arrays requires an *a priori* criterion of merit for the array/processing method combination; such criteria are discussed in §4. While the choice of a measure of merit affects the comparison of array geometries, our experience is that the general rules above produce good arrays by most criteria. This suggests that good arrays for use with data-adaptive processing can also be found using these general principles.

4. A priori design of processing methods

Given an array geometry, the objective in designing a processing method is to obtain a "useful" presentation of the observations, say N cross-spectra observed at N spatial separations. There is an infinite family of spectra which could produce these N observations so a measure of "utility" cannot be unambiguous. The resolution criterion employed here, the one which seems most natural in analysis of continuous spectra, is that the estimate $\hat{S}(k)$ should represent the average spectral intensity of the family of possible true spectra in the wavenumber region around k.

This is accomplished by using a window $W(k, k')$ which is concentrated in the region of k' space around $k' = k$.

The objective of *a priori* design of processing methods is to provide a method whereby the analyst may choose processing methods with a desirable balance of noise rejection and resolution where resolution is defined by an appropriate criterion. Such criteria can be generated from a quadratic form which measures lack of resolution.

$$P = \int W^2(\mathbf{k},\mathbf{k}')\, G(\mathbf{k},\mathbf{k}')\, d\mathbf{k}' \tag{11a}$$

where G is a prescribed positive semidefinite function which vanishes at $\mathbf{k}' = \mathbf{k}$ and is large at values of $\mathbf{k}'$ where significant values of $S(\mathbf{k}')$ are anticipated. It is through G that the analyst determines what is meant by "concentrated" around $\mathbf{k}$. From (8) it is apparent that there is no reason to minimize $W(\mathbf{k},\mathbf{k}')$ at values of $\mathbf{k}'$ where there is no significant variance in the spectrum $S(\mathbf{k}')$ and consequently $G(\mathbf{k},\mathbf{k}')$ should be chosen to vanish there. Other characteristics of G depend on what are judged the most important features of desirable windows. In some circumstances the important characteristic is that the central peak of the window be as sharp as possible; this behavior is encouraged when G increases rapidly as $|\mathbf{k} - \mathbf{k}'|$ increases and then becomes reasonably constant. In other cases a somewhat wider central peak but lower side-lobes away from $\mathbf{k}' = \mathbf{k}$ is desirable; this behavior is encouraged when G increases gradually around the point $\mathbf{k}' = \mathbf{k}$ but continues increasing over the entire region. The decision as to what are the more desirable characteristics of resolution depends on the nature of the spectrum to be investigated and the questions to be answered by the investigation. If the spectrum varies smoothly from very large to very small values, a wide central peak and low sidebands will be best; a spectrum with small amplitude, but narrow, peaks will be better described using the narrowest central peak achievable. Similarly, if the important feature of the spectrum is the average energy density over fairly large regions of $\mathbf{k}$ space it is most important to minimize side-lobes, whereas if the important spectral characteristics are the locations of peaks or other wavenumber localized features a narrow central peak is of paramount importance. Examples of the effect of G on window shapes will be given later.

While minimizing P, it is necessary to constrain the window to have nonzero values at the central peak; a useful general form is

$$V = \int W(\mathbf{k},\mathbf{k}')\, H(\mathbf{k},\mathbf{k}')\, d\mathbf{k}' = 1\,, \tag{11b}$$

where H, like G, is a positive semidefinite function chosen to promote desirable concentration of W at the central peak.

Optimal processing schemes are designed by minimizing some combination of the resolution penalty, P of (11a), and noise sensitivity, σ^2 of (10b), while meeting

the constraint $V = 1$ as in (11b). Computationally this is accomplished by choosing the weights β_{nm} of (11) to minimize

$$P + \nu\,\sigma^2 + \mu\,V \tag{12}$$

where μ is a Lagrange multiplier used to meet the constraint $V = 1$. The parameter ν, or equivalently the fundamental noise parameter σ_0^2, is adjusted to strike the desired balance between optimal resolution and maximum noise rejection. When P, V and σ^2 are expressed in terms of the weights β_{nm} using (11a), (11b), and (10b), respectively, minimization of (12) results in a system of linear equations for the optimal weight and this is easily solved using standard computer routines.

The fundamental trade-off between noise rejection and resolution is another necessarily subjective choice which must be specified by the analyst. Failing to account for noise sensitivity, that is setting $\nu = 0$, can lead to very large values of β and processing schemes which, although having nicely shaped windows, produce meaningless results when applied to real data. Although the analyst's choice of where on the resolution vs. noise rejection curve to operate is subjective, the results are not highly dependent on the choice unless the extreme of high resolution is sought. Our experience is that as the noise parameter ν increases from zero the resolution decreases slowly while the noise rejection increases rapidly at first and then fairly abruptly becomes approximately constant. Our subjective criterion for the appropriate balance of resolution and noise is to operate near the point where the rate of change of σ^2 with ν has begun to decrease noticeably; the overall processing performance is not highly sensitive to this choice, but knowledge of the expected noise levels might alter the choice significantly.

Any knowledge of the area of $\mathbf{k}$ space over which $S(\mathbf{k})$ is small can serve to improve the resolution obtainable over the remainder of the space. A particularly significant improvement occurs when there is adequate reason to believe that a linear theory predicting a dispersion relation is applicable. In this case the region of $\mathbf{k}$ space over which W^2 is to be minimized is reduced to a band (or several bands if there is more than one mode) and it is generally possible to minimize W^2 over this band more completely than can be accomplished over a larger region of wavenumber space. In a similar manner, there is sometimes *a priori* knowledge that some propagation directions are not energetic; this information can be used to improve resolution over the energetic directions by limiting nonzero values of $G(\mathbf{k},\mathbf{k}')$ to the energetic region of $\mathbf{k}'$ space.

We turn now to examination of some *a priori* processing schemes using, as an example, estimation of the directional spectrum of oceanic surface waves. The important feature of this example is that it is reasonable to assume that a majority of energy in the spectrum is concentrated on the dispersion curve

$$|\mathbf{k}| = k_0 = \omega^2/g\,.$$

In view of this concentration it is logical to seek estimates of $S(k)$ with k on the dispersion curve and to optimize resolution so that there is minimum of energy spreading from the region of wavenumbers k' also on the curve.

In conjunction with analysis of ocean surface waves, Barber (1963) introduced an *a priori* analysis method which has received much use, often under the name of Beam Former estimate. If cross-spectra are observed at lag $\zeta_0 = 0$ and the $2N$ non-zero lags ζ_n, where $\zeta_n = -\zeta_{-n}$ and $1 \leqslant n \leqslant N$, Barber argued that the estimate

$$\hat{S}(k) = \sum_{n=-N}^{N} \hat{Q}(\zeta_n) \exp(-i\mathbf{k} \cdot \zeta_n)$$

is a good approximation of the exact form (5a). This corresponds to weights

$$\beta_{lm} = 1/M(x_l - x_m), \tag{13a}$$

where M is the number of element pairs associated with the lag $x_l - x_m$, and the window

$$W(k, k') = 1 + 2 \sum_{n=1}^{N} \cos(\zeta_n \cdot [\mathbf{k} - \mathbf{k}']). \tag{13b}$$

Although not usually presented as such, the Beam Former is an "optimized" processing method. For every finite coarray there exist two generating vectors ζ_x and ζ_y such that every lag can be represented as

$$\zeta_n = L\,\zeta_x + M\,\zeta_y$$

where L and M are *integers*. It then follows that the exponential functions in (10a) are periodic (and hence orthogonal) over rectangles of dimensions $2\pi/|\zeta_x|$ and $2\pi/|\zeta_y|$. The Barber processing scheme is the one which minimizes

$$\int W^2(\mathbf{k},\mathbf{k}')d\mathbf{k}'$$

over this rectangle while maintaining

$$W(k,k) = 2N + 1.$$

The nonobjective elements in these design criteria are that the area over which W^2 is minimized is determined by the details of the array geometry, not by any properties of the spectrum, and that the constraint on W at the central peak is not motivated by any objective principle. The results are the ambiguity associated with nearly redundant separations, which may be traced to the sensitivity of the generating lags ζ_x and ζ_y to nearly redundant lags, and unreasonable estimates of the total energy, $\int d\mathbf{k}\,\hat{S}(\mathbf{k})$.

These fundamental weaknesses of Barber's scheme do not imply a lack of utility. In cases where the coarray is a regular one, so that ζ_x and ζ_y are essentially the smallest separations, the resolution and noise rejection of the Barber scheme is

reasonable and the computational simplicity attractive. The remaining difficulty is easily overcome by renormalizing the spectral estimates so that the total energy is that computed directly from the auto-spectra of the individual elements. Consequently this method, although not truly an optimal one by any reasonable criterion, is included in the comparisons of performance.

A second interesting scheme is achieved by optimizing only for noise rejection, that is by setting $\nu = \infty$ in (12). If the constraint (11b) is taken as

$$W(\mathbf{k},\mathbf{k}) = N_e^2 - N_e + 1$$

the optimal weights are

$$\beta_{nm} = 1$$

and the window is

$$W(\mathbf{k},\mathbf{k}') = N_e + 2 \sum_{n=1}^{N_e} \sum_{m=n+1}^{N_e} \cos(\zeta_{nm} \cdot [\mathbf{k} - \mathbf{k}']) .$$

This window bears a strong resemblance to the Barber window (13b), differing in that redundant separations are weighted more heavily. It has the interesting property that W is positive semidefinite, a characteristic which occurs whenever β_{nm} has the factorable form

$$\beta_{nm} = \overline{\beta}_n \beta_m .$$

Although the resolution is inferior to the Barber window and the method suffers from the same difficulty that the estimate must be renormalized to preserve total energy, in cases where the noise is large this may be a useful processing scheme.

Two additional *a priori* optimizing methods, designed to make use of knowledge of the dispersion relation, have been examined. In the first of these, called the Omnidirectional A Priori (OAP) scheme, it is assumed that energy is confined to a circular region surrounding the point at which the spectrum is being estimated. This omnidirectional criteria of optimization has the virtue that the coefficients β_{nm} are independent of the wavenumber at which the estimate is being made and depend only on R, the radius of the circular region over which the spectrum is taken to be confined. In the second method, called the Steered A Priori (SAP) scheme, a more detailed specification of the region over which the spectrum is assumed significant is used and the weights β_{nm} depend on the wavenumber at which the spectrum is being estimated. While this allows some improvement in resolution and noise rejection, the added computational difficulty may not always be warranted. Although the presentation is confined to the two-dimensional problem, the methods are easily generalized to higher dimensions.

The OAP and SAP schemes are motivated by the following considerations. If the dispersion relation $|\mathbf{k}| = k_0$ (where k_0 depends on the frequency band being analyzed) is correct then the estimate

$$<\hat{S}(\mathbf{k})> = \int d\mathbf{k}'\, S(\mathbf{k}')\, W(\mathbf{k},\mathbf{k}') \tag{8}$$

is influenced only by the value of $W(\mathbf{k},\mathbf{k}')$ along the band $|\mathbf{k}'| = k_0$. The most appropriate criterion of resolution would then be obtained by letting the resolution penalty function $G(\mathbf{k},\mathbf{k}')$ in (11a) take zero value except near this curve. This is the motivation for the SAP scheme. Representing $\mathbf{k}'$ in its polar components k' and θ', $G(\mathbf{k},\mathbf{k}')$ is taken as $\delta(|\mathbf{k}'|-k_0)\, \tilde{G}(\theta' - \theta)$ where θ is the angle of the wavenumber $\mathbf{k}$. Choice of the angular function $\tilde{G}$ is subject to the same considerations as pertain to the choice of G itself. The constraint (11b) is similarly chosen to account for the localized nature of the spectrum; in the examples to be given here $H(\mathbf{k},\mathbf{k}')$ is taken to be $\delta(|\mathbf{k}'| - k_0)/k_0$ so that if the spectrum $S(\mathbf{k}')$ is isotropic the estimate (8) will be exact.

The primary disadvantage of the SAP scheme is that the weights β_{nm} depend on the $\mathbf{k}$ (or θ) at which S is being estimated. Thus to make directional spectrum estimates at various angles it is necessary to compute and store a large number of coefficients. This disadvantage is overcome, at some expense to performance, by the OAP scheme which is based on the observation that for every energetic $\mathbf{k}$ all the energy is concentrated in wavenumbers $\mathbf{k}'$ such that $|\mathbf{k} - \mathbf{k}'| \leqslant 2k_0$. Thus there is no advantage to making $W(\mathbf{k},\mathbf{k}')$ small at values of $|\mathbf{k} - \mathbf{k}'|$ greater than $2k_0$. This suggests that an optimal processing scheme based on weights β which do not depend on $\mathbf{k}$ itself can be obtained by taking $G(\mathbf{k},\mathbf{k}') = g|\mathbf{k} - \mathbf{k}'|$ where g vanishes for $|\mathbf{k} - \mathbf{k}'| > rk_0$ and r is on the order of two. It would be inappropriate to constrain the integral of W over the entire region $|\mathbf{k} - \mathbf{k}'| \leqslant rk_0$ so the constraint $W(\mathbf{k},\mathbf{k}) = 1$ is employed; this leads to the same difficulties with respect to total energy preservation as are encountered with the Barber and noise rejector schemes, but this is easily overcome by renormalizing the estimated spectrum so that total energy in each frequency band is correct.

While these descriptions of the OAP and SAP schemes demonstrate how the area over which resolution is optimized may be selected, they do not address the question of selecting the weighting functions $\tilde{G}(\theta)$ and $g(k)$. As discussed earlier, the shape of these weights is how the analyst selects the optimizing principle which strikes the desired balance between sharp central peaks and low side-lobes in the resolution window W. An example of the effect of different shapes is given in Figure 3. The example is based on the one-dimensional array A of §3. In Figure 2 of that section the resolving window obtained by minimizing

$$P = \int_{-\pi}^{\pi} W^2(k,k+l)dl$$

subject to $W(k,k) = 1$ was shown. This corresponds to minimizing (12) with $\nu = 0$ and

$$G(k,k') = 1, \qquad H(k,k') = \delta(k - k'),$$

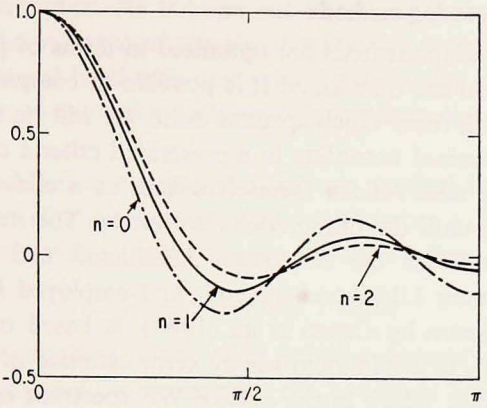

Figure 3. Examples of the resolving window associated with Array A of Figure 1. The three examples are the windows obtained by minimizing the resolution penalty $\int_{-\pi}^{\pi} W^2(k,k+l)\,|l|^n\,dl$ for n=0,1,2. Note that n=2, which penalizes most strongly for nonzero W^2 at large l, produces the smallest side-lobes whereas n=0 leads to the sharpest central peak.

for $|k - k'| < \pi$ and $G = H = 0$ elsewhere. In Figure 3 this window is depicted along with those obtained with the same ν and H but with

$$G(k,k') = |k - k'| \text{ and } G(k,k') = (k - k')^2$$

for $|k - k'| < \pi$ and zero elsewhere. The general pattern is clear. The more rapidly G increases with $|k - k'|$ the lower the side-lobes away from the central peak and the wider the central peak itself.

The discussion above describes the various considerations employed in design of *all* spectral estimation techniques, namely, rejection of noise and maximization of resolution which may be considered as a balance between sharp central peaks and low side-lobes in the resolving window. An attempt has been made to emphasize the subjective nature of optimization of these characteristics. In the following section data-adaptive estimators will be discussed. These methods have some real virtues relative to *a priori* estimators but their primary disadvantage is that the effects of different choices of optimization criteria are not so easily seen. It is important to remember that these schemes are associated wtih the same subjective elements that are found in *a priori* schemes. They have susceptibility to noise and wavenumber windows and these windows differ in the relative rejection of energy at wavenumbers near and far from the estimation wavenumber. One must not allow the simplicity of their optimization principles to hide the fact that they are still basically subjective. In the final analysis it is the analyst who must select the method which performs best according to his own criteria. It is for this reason that examples of various processing schemes are included in a later section.

5. Data-adaptive processing methods

The *a priori* processing methods are optimized in terms of prescribed criteria of merit; once the criteria are established it is possible to compute, once and for all, the weights α_{nm} or β_{nm} from which spectral estimates will be made. Data-adaptive methods are also optimized according to a prescribed criteria of merit but this prescription involves the observations themselves and, as a consequence, the weights α_{nm} cannot be found until the observations are made. The most widely employed data-adaptive methods are the Maximum Likelihood and Maximum Entropy methods. The Maximum Likelihood method, first employed for estimating wavenumber-frequency spectra by Capon *et al.* (1967), is based on a criteria of merit which in essence leads to a minimum square error estimate of a plane wave signal in the presence of noise, which in the case of WF spectrum estimation is the contribution to the observations from components with different wavenumbers. We are unaware of an algorithm for estimating *WF* spectra by applying the maximum entropy method to coarrays of arbitrary form; McDonough (1974) provides a suitable method for an equally-spaced linear coarray. Lacoss (1971) has presented the method and an evaluation of its performance relative to a particular *a priori* method and the Maximum Likelihood method in the analogous case of frequency spectrum estimation. Only the Maximum Likelihood method and an extension of it tailored to estimating continuous spectra, rather than signal detection, will be discussed here.

Data-adaptive spectral estimation, like the *a priori* estimation methods of the previous section, is based on linear weightings of the observed cross-spectral matrix of the form (8). One unique feature is that the weights are of the factorable form $\alpha_{nm} = \gamma_n \gamma_m^*$ where γ_m^* is the complex conjugate of γ_m. This is a natural feature when it is the signal amplitude, rather than the variance, which is to be estimated since the appropriate amplitude estimator is a linear weighting of the signals from the array elements; when the variance is estimated as the square of the estimated amplitude this leads to a variance estimator like (8) but with factorable weights. The result of the factorable property of the weights α_{nm} is that the window W becomes the magnitude squared of a function and is, therefore, nonnegative. While this is not necessarily a desirable property for spectral estimators (the *a priori* criteria employed in the previous section do not usually lead to nonnegative windows) it is fundamental to the data-adaptive methods discussed here.

The optimization principle upon which the Maximum Likelihood Estimator (MLE) is based concerns detecting a signal consisting of a single plane wave (or in other than Fourier spectral analysis, some other prescribed functional form) in the presence of noise. Capon *et al.* (1967) originally presented the method as the minimum expected square error estimator of the complex amplitude (representing amplitude and phase) of the signal or, equivalently, the estimator of the most likely amplitude when the noise at the various array elements has a joint Normal (Gaus-

sian) probability distribution. As it turns out this is equivalent to minimizing the error in estimating the variance of the signal when the processing method is constrained to pass the signal itself with unit gain. Since it is the variance of the signal which is most relevant to spectral estimation, it is this latter line that will be followed here.

Application of the MLE to continuous spectra is clarified by considering the observed field to consist of a plane wave with wavenumber $\mathbf{l}$ and variance (energy) $E(\mathbf{l})$ plus "noise". The wavenumber-frequency spectrum is then

$$S(\mathbf{k}) = \delta(\mathbf{k}-\mathbf{l})\, E(\mathbf{l}) + S_N(\mathbf{k})$$

where S_N is the noise spectrum. Here, as in previous sections, each frequency band is considered separately and the ω denoting the frequency under consideration has been dropped. An estimate of the variance E is

$$\hat{E}(\mathbf{l}) = \sum_{n}^{N_e} \sum_{m}^{N_e} \gamma_n(\mathbf{l})\, \gamma_m{}^*(\mathbf{l})\, \hat{Q}_{nm} \tag{14a}$$

$$= E(\mathbf{l}) \sum_n \sum_m \gamma_n \gamma_m{}^* \exp\left(i\mathbf{l} \cdot [\mathbf{x}_n - \mathbf{x}_m]\right) + \int d\mathbf{k}\, W(\mathbf{l},\mathbf{k}) S_N(\mathbf{k})$$

where W, in analogy with (9), is given by

$$W(\mathbf{l},\mathbf{k}) = \left|\sum_n \gamma_n(\mathbf{l}) \exp i\,\mathbf{k} \cdot \mathbf{x}_n\right|^2 .$$

The crucial feature of the MLE is the imposed constraint that in the absence of noise the signal be passed with unit gain which leads to

$$\sum_n \gamma_n \exp\left(i\,\mathbf{l} \cdot \mathbf{x}_n\right) = 1 \tag{14b}$$

and

$$\hat{E}(\mathbf{l}) = E(\mathbf{l}) + \int dk\, W(\mathbf{l},\mathbf{k})\, S_N(\mathbf{k}) .$$

Once the constraint (14b) is imposed the optimal weights are obviously those which minimize the convolution of W and S_N which, since E, W and S_N are all nonnegative, is equivalent to minimizing the estimate $\hat{E}$, itself.

Minimization of (14a) subject to constraint (14b) results in a linear system from which the weights $\gamma_n(\mathbf{l})$ can be found for each $\mathbf{l}$ at which E is to be estimated. Substitution of these into (14a) leads to the estimate

$$\hat{E}(\mathbf{l}) = \left\{ \sum_{n}^{N_e} \sum_{m}^{N_e} \hat{Q}_{nm}{}^{-1} \exp\left(i\,\mathbf{l} \cdot [\mathbf{x}_n - \mathbf{x}_m]\right) \right\}^{-1} \tag{15}$$

where $\hat{Q}_{nm}{}^{-1}$ is the inverse of the matrix $\hat{Q}_{nm}$. This inverse almost always exists when $\hat{Q}_{nm}$ is obtained from real data since its nonexistence implies that the measured field can be described as made up of less than N_e wavetrains. Even in this pathologic case an estimate $\hat{E}$ could be achieved by minimizing (14a); the solution

for the γ_n would be nonunique but the estimate itself would be the same for all possible solutions.

A fundamental difficulty with applying the MLE to estimation of continuous spectra is disclosed by considering the practical aspect of using $\hat{E}$, ostensibly an estimate of the variance of the wave with wavenumber $\mathbf{l}$, to estimate the variance density $S(\mathbf{l})$. The natural choice would be to divide $\hat{E}$ by some estimate of the area in $\mathbf{l}$ space which contributed to the estimate $\hat{E}$. A reasonable approach, the one usually employed, is to compute $\hat{E}(\mathbf{l})$ at a number of discrete wavenumbers $\mathbf{l}$, connect these by a smooth curve and normalize the curve by the constant required to preserve the total variance estimated directly from $\hat{Q}_{nn}$; this *a posteriori* renormalization is also required in some of the methods described in the previous section. The difficulty exposed here is that the MLE is optimized to estimate the variance at a particular wavenumber whereas in a continuous spectrum this energy is zero. This does not mean that the spectral estimates achieved by the renormalization procedure are not useful ones but it does suggest that a data-adaptive technique designed to extract a signal, with finite variance at a single wavenumber, might not be the optimal one with which to estimate continuous spectra.

It is application of constraint (14b), insuring a unit gain response to the "signal" wave, which poses the greatest question about use of the MLE for estimation of continuous spectra. It is alteration of this constraint which has led us to propose an alternative estimator, which will subsequently be referred to as the Data-Adaptive Spectral Estimator (DASE) to distinguish it from MLE which is actually a signal detector. The method is similar to the MLE in that factorable weights γ_n are used to form an estimate

$$\hat{S}(\mathbf{l}) = \sum_{n}^{Ne} \sum_{m}^{Ne} \gamma_n(\mathbf{l})\, \hat{Q}_{nm}\, \gamma_m{}^*(\mathbf{l}) = \int d\mathbf{k}\, W(\mathbf{l},\mathbf{k})\, S(\mathbf{k}) \tag{16a}$$

where, as for the MLE,

$$W(\mathbf{l},\mathbf{k}) = |\underset{n}{\Sigma}\, \gamma_n(\mathbf{l}) \exp(i\,\mathbf{k} \cdot \mathbf{x}_n)|^2 .$$

Rather than insisting that $W(\mathbf{l},\mathbf{l}) = 1$, as is done for the MLE, the constraint employed in the DASE method is that the integral of the window over some region surrounding the estimation wavenumber $\mathbf{l}$ be unity. In the case of interest here, estimation of the directional spectrum of waves obeying the dispersion relation $|\mathbf{k}| = k_0$, it is the integral along this line which is appropriate so the constraint is taken as

$$k_0 \int_{\theta_0-\Delta/2}^{\theta_0+\Delta/2} W(\mathbf{l},\mathbf{k}) d\theta = 1$$

where $\theta = \arg(\mathbf{k})$, $\theta_0 = \arg(\mathbf{l})$ and the integral is taken along $|\mathbf{k}| = k_0$. When the form of the window is substituted this constraint becomes

$$\sum_{n}^{N_e} \sum_{m}^{N_e} \gamma_n T_{nm} \gamma_m{}^* = 1 \tag{16b}$$

$$T_{nm} = k_0 \int_{\theta_0-\Delta/2}^{\theta_0+\Delta/2} \exp\left(i\,\mathbf{k} \cdot [\mathbf{x}_n - \mathbf{x}_m]\right) d\theta \,.$$

Because the window is nonnegative, minimization of $\hat{S}$ subject to the constraint minimizes the influence of $S(\mathbf{k})$ from $\mathbf{k}$ outside the band $k = [k_0, \theta_0 \pm \Delta/2]$. This minimum is achieved at the extreme of

$$\sum_{n}\sum_{m}\left[\hat{Q}_{nm} - \frac{1}{\lambda} T_{nm}\right] \gamma_n \gamma_m{}^*$$

where λ is the Lagrange multiplier. Minimization of this form may be shown to be equivalent (Regier, 1975) to solving the eigenvalue problem

$$\sum_{m} [\sum_{l} \hat{Q}_{ml}{}^{-1} T_{ln}] \gamma_m = \lambda\, \gamma_n$$

where, as before, $\hat{Q}^{-1}$ is the inverse of $\hat{Q}$. Combining the solution of this equation with the constraint (16b) and the form of the estimate in (16a) demonstrates that the optimal estimate is

$$\hat{S}(\mathbf{l}) = \lambda_0{}^{-1} \tag{16c}$$

where λ_0 is the largest eigenvalue of the Hermitian matrix

$$R_{nm} = \sum_{l} \hat{Q}_{nl}{}^{-1} T_{lm} \,. \tag{16d}$$

The DASE involves the adjustable parameter Δ which determines the size of the area in $\mathbf{k}$ space over which the integral of $W(\mathbf{l},\mathbf{k})$ is constrained to be unity. In the limit $\Delta = 0$ the DASE is equivalent to the MLE. When Δ is small the estimate $\hat{S}(\mathbf{l})$ will be larger than the true value $S(\mathbf{l})$ since in (16a) much of the estimate will be due to the convolution of $W(\mathbf{l},\mathbf{k})$ and $S(\mathbf{k})$ from outside the range of $\mathbf{k}$ over which the integral of W is constrained to unity. After considerable experimentation with model spectra we find that as Δ approaches 2π the integral of the DASE estimate becomes less than the true estimate unless the true spectrum is nearly isotropic. The analyst is presented with two options for choosing an appropriate value of Δ. On the one hand, if some *a priori* knowledge of the width of the true spectrum is available, he may take Δ as some prescribed value of approximately this size. This will result in an estimate $\hat{S}(\mathbf{l})$ which does not correspond to the true variance as determined from the auto-spectra $\hat{Q}_{nn}$ but this deficiency is easily overcome by the renormalization procedure employed with the other estimators discussed above. The second alternative is to search for that value of Δ which leads to the proper correspondence between variance and the integral of the spectral estimate. Although this latter approach involves considerable computational effort, we have found that

the resulting spectral estimates are, for a wide range of true spectra, accurate and the lack of any subjective element in the design the processing method is attractive.

6. Comparison of processing methods

Three *a priori* processing methods were introduced in §3. The Beam Former (BF) estimator, introduced to ocean wave spectral estimation by Barber (1963), is the only one for which the window and coefficients can be computed analytically (see (13)). Both the Omnidirectional A Priori (OAP) and the Steered A Priori (SAP) methods involve optimization of the measure of merit given in (12). They differ in the specific definitions of P and V employed (see equations (11)); as a consequence of these definitions the weights β_{nm} for the SAP method depend on the direction of the wavenumber at which the spectrum is being estimated whereas the weights for the OAP method do not. One of the primary advantages of *a priori* methods is that, once the array geometry is established, a set of weights β_{nm} (and the associated α_{nm}) can be computed once and the process of estimating spectra from different sets of data simply involves summing (6). Regier (1975) gives details of computing the weights which, by accounting for symmetries in the specific array and in the processing scheme itself, reduce the amount of computation involved. Clearly the BF method involves the least computational difficulty while the SAP method involves the most. Even for modest array sizes and waves obeying a known dispersion relation, the SAP computations can be significant; to estimate the directional spectrum of such waves at 10° increments at 40 frequencies from a six element array over 2000 complex valued coefficients are required, even after the natural symmetries involved are accounted for.

The data adaptive Maximum Likelihood Estimator (MLE) and the Data Adaptive Special Estimator (DASE) introduced in §4 require extensive calculation each time a spectral estimate is to be made. In the case of the MLE the inverse of the cross-spectral matrix $\hat{Q}_{nm}$ must be computed for each frequency and the sum (15) computed at each estimation wavenumber, $\mathbf{l}$. The DASE requires even more extensive calculation since the dominant eigenvalue of the matrix R of (16d) must be found for each frequency/estimation-wavenumber pair.

Comparison of the computational difficulties of *a priori* and data adaptive methods depends on the number of different cross-spectral matrices to be analyzed. Since the weighting coefficients for the *a priori* methods need be evaluated only once, these methods are more economical when the number of observations is large; the number of coefficients can become large enough that storing them is a problem if several frequency bands of data from a large array are to be examined.

In order to explore and compare the performance of the various processing methods we have selected an array geometry, generated simulated cross-spectral matrices from a known directional spectrum and portrayed the estimated spectra obtained from the three *a priori* processing methods and the two data adaptive

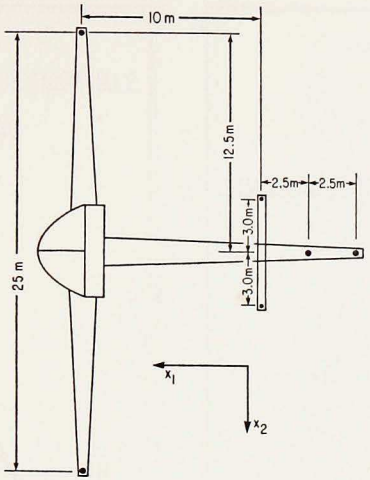

Figure 4. Top view of FLIP and the six element array used for measurements and testing the processing methods. The elements are shown as dots. In operation FLIP was oriented with prevailing wind parallel to the x_1 axis. The angle $\theta = 0$ referred to in simulations corresponds to waves propagating parallel to the x_1 axis.

methods described above. This exercise was carried out in conjunction with an experimental program to observe the directional spectrum of surface waves using an array deployed from the spar buoy FLIP; the results are described in the companion paper by Regier and Davis (1977). The six element array employed in the experiments and the one used in examinations of processing method performance is shown in Figure 4. All test data were generated to obey the linear surface wave dispersion relation so that

$$S(\mathbf{k},\omega) = \frac{1}{|\mathbf{k}|}\,\delta(|\mathbf{k}| - k_0)\,S(\theta,\omega)$$

where k_0 is given by the dispersion relation $k_0 = \omega^2/g$, θ is the direction of the wavenumber vector, and $S(\theta,\omega)$ is the directional frequency spectrum.

All estimated spectra have been normalized to produce the correct total variance. The OAP and SAP estimators were optimized with respect to both resolution and noise rejection according to the methods outlined in §3. Specifically, the simplified measure of noise sensitivity (10b) was used and, as discussed in §3, the noise parameter was increased to the point where noise rejection began to decrease slowly. Regier (1975) provides quantitative measures of noise rejection which show all methods to have approximately the same noise sensitivity. The OAP estimator was optimized using in (11) $H(\mathbf{k},\mathbf{k}') = \delta\,(\mathbf{k}-\mathbf{k}')$ and $G(\mathbf{k},\mathbf{k}') = 1$ for $|\mathbf{k}-\mathbf{k}'| < 2.5\,k_0$ and $G = 0$ outside that region. The SAP estimator was optimized using $H(\mathbf{k},\mathbf{k}') = \delta(|\mathbf{k}'| - k_0)$ and $G(\mathbf{k},\mathbf{k}') = \delta(|\mathbf{k}'| - k_0)\,[\theta - \theta']^2$ where θ and θ' are, respectively,

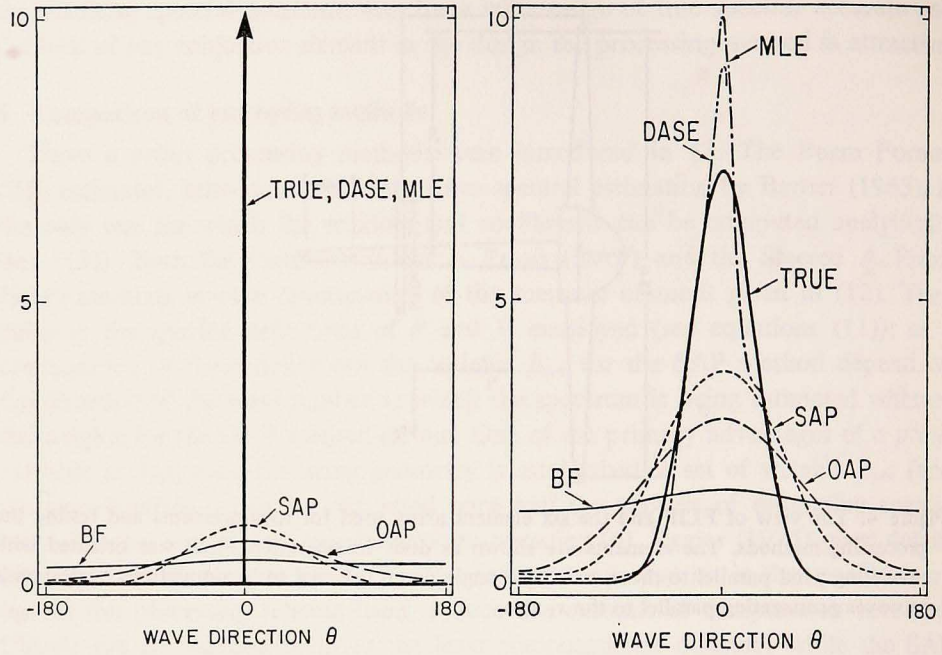

Figure 5. Estimated directional spectra for a true spectrum consisting of a single wave with frequency 0.1 Hz.

Figure 6. Estimated directional spectra for 0.1 Hz waves having a true directional spectrum proportional to $\cos^{16}(\theta/2)$.

the directions of $\mathbf{k}$ and $\mathbf{k}'$; this choice of G was made to provide a seemingly desirable balance of side-lobe suppression and sharpness of the central peak (see §3) and the choice of H insures an exact representation of an isotropic spectrum.

Here the estimated spectra found by the five processing methods are shown for six true spectra, all symmetric about the x_1 axis of the array shown in Figure 4. The estimated spectra were computed along the circle $|\mathbf{k}| = k_0$ and are portrayed as a function of propagation angle θ; values were computed in 12 degree steps. Since the array and the true spectra are both symmetric about the x_1 axis, the estimated spectra are also; consequently, in some figures the estimated spectra are shown only over a 180° range.

Two of the examples were chosen to represent low frequency swell which were expected to have a relatively narrow directional spectrum as compared with higher frequency waves. It is unfortunately true that the directional resolution of long waves is generally lower than for waves whose scale is of the order of the array itself whereas long waves tend to have a narrower directional spectrum (see Regier and Davis, 1977). The two low frequency test spectra consisted of a single plane wave propagating along $\theta = 0$ (parallel to the x_1 axis) and a narrow directional

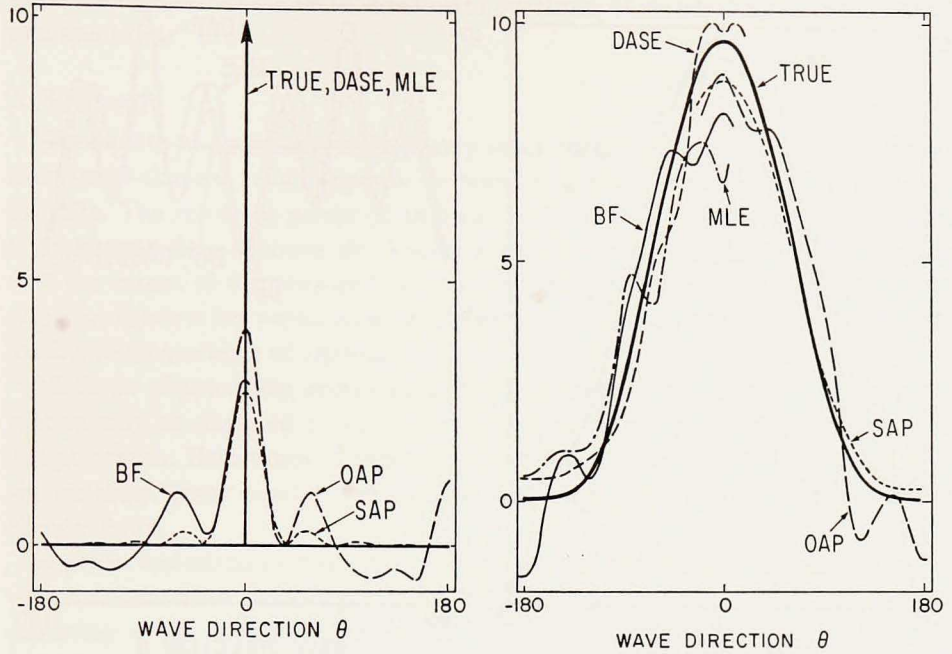

Figure 7. Estimated directional spectra for a true spectrum consisting of a single wave with frequency 0.3 Hz.

Figure 8. Estimated directional spectra for 0.3 Hz waves having a directional spectrum proportional to $\cos^4(\theta/2)$.

spectrum proportional to $\cos^{16}(\theta/2)$ both with frequencies of 0.1 Hz. The 156 m wavelength of these waves is considerably larger than the overall dimensions of the array so rather poor resolution might be anticipated. The estimated spectra are depicted in Figures 5 and 6. In both examples the *a priori* optimized methods fail to resolve these long, highly directional spectra. As might be expected from the relative sophistication of the optimizing criteria, the SAP estimate is best resolved and the BF estimate is the poorest. The two data adaptive methods both resolve exactly the plane wave. In this case the two methods are equivalent because the Δ associated with the DASE approaches zero (see §4). It is a simpler matter to show that data adaptive methods of this general type can almost always perfectly resolve a single wave since it is possible to find weights γ_n which reject completely the single wave while maintaining the constraints (14b) or (16b). The surprising result is that the data adaptive methods both over-focus the finite width spectrum in Figure 6; the MLE appears worse, in this respect, than the DASE.

Four more examples, considered representative of the spectrum at higher frequencies, are shown in Figures 7 through 10. All examples correspond to a frequency of 0.3 Hz and a wavelength of approximately 17 m. Since the spatial extent of

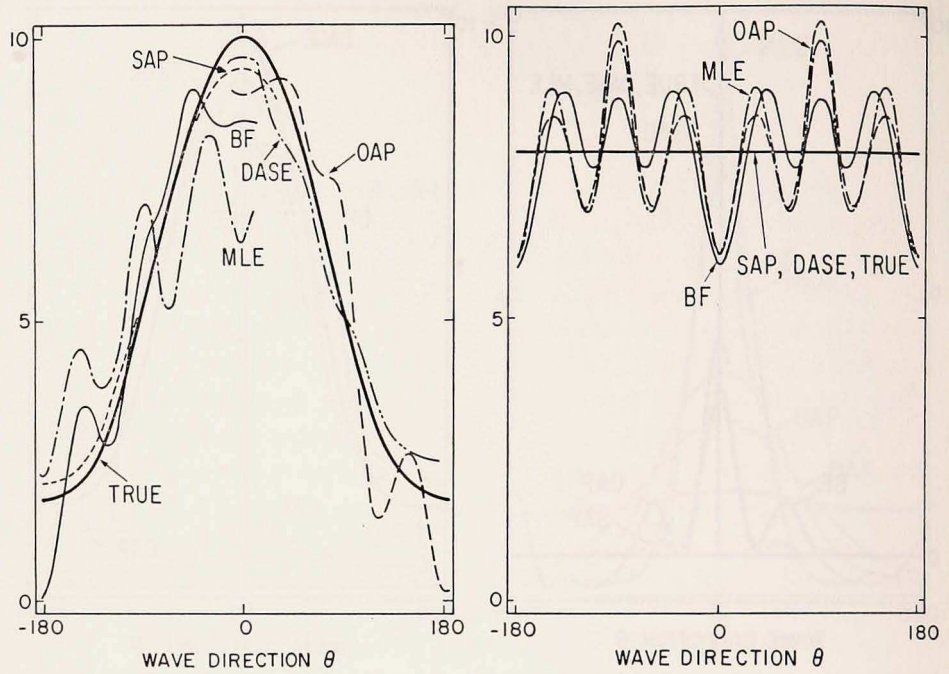

Figure 9. Estimated directional spectra for 0.3 Hz waves having a true directional spectrum proportional to $\cos^{16}(\theta/2) + 0.18 \cos^2(\theta-\pi)/2$.

Figure 10. Estimated directional spectra for an isotropic true directional spectrum of 0.3 Hz waves.

these waves is more nearly matched to the array size than are the 0.1 Hz waves, an improved resolution of all processing methods is anticipated and found. Figure 7 depicts the responses of the various estimation methods to a single wave from $\theta = 0$. Again the data adaptive methods resolve a single wave exactly and SAP method is the best of the *a priori* methods. Figure 8 represents the estimates for a narrow directional spectrum proportional to $\cos^4(\theta/2)$. The most faithful representation appears to be the SAP estimate but the DASE performs nearly as well; the primary failings of the DASE are a slight over-focusing of the peak and some spurious angular structure. The OAP and BF methods are significantly less accurate at large angles, where the estimates are negative, but it is the data adaptive MLE which is most disappointing; the MLE seriously underestimates the peak while introducing large spurious peaks. A similar case, corresponding to a true spectrum proportional to $\cos^8(\theta/2) + .18 \cos^2((\theta - \pi)/2)$ is shown in Figure 9. Again the SAP estimator and the DASE are best and the MLE, which produces quite large spurious peaks, is the least faithful. The final example (Fig. 10) is an isotropic directional spectrum. By design the DASE and SAP estimates are exact. The remaining estimates show significant spurious peaks; the MLE and OAP methods

show a range of variations in directional spectrum of nearly half the value of the true spectrum.

7. Summary

The fidelity of wavenumber-frequency or directional-frequency spectra estimated from multi-element arrays depends on both the array and the method of processing the data. The resolving power of an array is determined by its coarray, the collection of separations between all element pairs. Resolution of waves large compared with the extent of the coarray is limited and there are significant side-lobes in the response window for waves comparable to or smaller than the smallest separations, leading to an analogue of aliasing.

Methods of processing array data to form estimates of wavenumber or directional spectra may be classified as *a priori* optimized or data adaptive depending on their design criteria. Both types of processing methods produce spectral estimates which are weighted linear combinations of the frequency cross spectra of various array element pairs.

A priori optimized methods seek to optimize the weights in this estimate to provide a wavenumber resolving window (9) with characteristics which are desirable according to pre-specified measures of merit which include insensitivity to instrumental or statistical noise and desirable spectral resolution. A quantitative measure of these characteristics can be defined as in (12) and this measure optimized by adjustment of the weights. Three such methods, optimized according to various criteria, were described in §3; these were the conventional Beam Former (BF) method, a relatively simple method optimized without regard to array orientation (Omnidirectional A Priori, or, OAP) and a more sophisticated method (Steered A Priori, or SAP).

Data adaptive techniques are based on optimizing the estimation weights according to criteria which depend on the data itself. The two methods examined here were the Maximum Likelihood Estimator (MLE) and the Data Adaptive Spectral Estimator (DASE). Both methods attempt to create a wavenumber filter which rejects as much of the observed input as possible while passing a certain prescribed input unattenuated. The MLE is designed as a signal estimator and consequently is constrained to pass a single plane wave with unit gain. The DASE, designed specifically to estimate continuous spectra, is constrained to have an average unit gain to waves within a prescribed passband; the passband is determined from the data and in the limit of a single wave input the passband becomes focused and the DASE becomes equivalent to the MLE.

The estimated spectra resulting from the various processing methods have been examined for a number of input spectra consisting of both narrow and wide directional distributions and wavelengths which are large compared with the sampling array and wavelengths comparable to the array size. In summary, the most sophisti-

cated *a priori* method (SAP) generally outperforms the other *a priori* methods which are operationally much simpler to apply. The data alaptive MLE produces very high resolution of narrow spectra but tends to introduce spurious fine structure and artificial peaks when the input spectrum is broad. The DASE, over a wide range of conditions, generally outperforms both the MLE (particularly for broad input spectra) and the *a priori* methods (particularly for narrow input spectra).

Acknowledgment. We wish to acknowledge the patient support of the Office of Naval Research under Contract N00014-75-C-0152.

APPENDIX. The error matrix

As discussed in Section 2, the fidelity of *WF* spectral estimates is degraded by errors in Q_{nm}, the estimated cross-spectrum between elements n and m. In that section it was suggested that adequate reduction of noise sensitivity is achieved by assuming the cross-spectra to have independent errors of equal magnitude. In some cases it is worthwhile to use more sophisticated estimates of the error matrix. Here the error matrices associated with the two most common types of noise are discussed.

Let the frequency Fourier transform of the signal, $\eta(x_n,t)$, from element n be

$$H_n + h_n$$

where h_n is due to instrumental noise. The cross-spectral estimate $\hat{Q}_{nm}$ of (4) is then formed by sample averaging products of Fourier coefficients. If $\{\ \}$ denotes the sample average while $<\ >$ continues to denote the true average then

$$\hat{Q}_{nm} = \{H_n H_m{}^*\} + \{H_n h_m{}^*\} + \{h_n H_m{}^*\} + \{h_n h_m{}^*\}\,.$$

When the errors, h, result from instrumental noise which is independent between elements and the statistical errors vanish ($\{\ \} = <\ >$) then

$$q_{nm} = \delta_{nm} <h^2>, \quad <q_{ij} q_{nm}> = \delta_{ij}\,\delta_{nm} <h^2>^2\,.$$

The noise parameter σ^2 appearing in (12) is then given by

$$\sigma^2 = \sum_{n,i} \alpha_{nn}\alpha_{ii} <h^2>^2\,.$$

A more complicated, but probably more generally useful, case occurs when the instrumental noise is negligible ($h \approx 0$) but the sample average is not equal to the true average. In this case of statistical noise the error matrix is

$$<q_{ij} q_{nm}> = <\{H_i H_j{}^*\}\,\{H_n H_m{}^*\}> - <H_i H_j{}^*> <H_n H_m{}^*>\,.$$

In general this expression cannot be determined from two-point statistics. But if the statistics of H are approximately Gaussian, so the H values have approximately joint normal probability distributions (see Cramer, 1962), and M is the number of (independent) realizations from which the sample average is constructed, then

$$<q_{ij} q_{nm}> = \frac{1}{M} <H_i H_m{}^*> <H_j{}^* H_n> + \frac{1}{M} <H_i H_n> <H_j{}^* H_m{}^*> + 0(M^{-2})\,.$$

If the sample statistics are reasonably accurate the last term is negligible; if the process is stationary the second term vanishes. Then

$$\langle q_{ij} q_{nm} \rangle = \frac{1}{M} Q_{im} Q_{nj} .$$

A plausible way of estimating this error matrix is to substitute the observed sample cross-spectra $\hat{Q}$ in place of Q, committing an additional error of $0(M^{-2})$. This, however, prevents truly *a priori* design prior to acquisition of the observations. For truly *a priori* design which seeks to minimize accurately statistical noise, an estimate of the cross-spectra (or equivalently the *WF* spectrum) must be made in advance of any observations.

REFERENCES

Barber, N. F. 1963. The directional resolving power of an array of wave detectors, *in* Ocean Wave Spectra. Englewood Cliffs, New Jersey, Prentice-Hall, 357 pp.

Capon, J., R. J. Greenfield and R. J. Kolker. 1967. Multidimensional maximum-likelihood processing of a large aperture seismic array. Proc. IEEE, *55*, 192–211.

Cramer, H. 1962. Random Variables and Probability Distributions. Cambridge Univ. Press, 118 pp.

Haubrich, R. A. 1968. Array design. Bull. Seismolog. Soc. of Amer., *58*, 977–991.

Jenkins, G. M. and D. G. Watts. 1968. Spectral Analysis and Its Applications. San Francisco, Holden-Day, 525 pp.

Lacoss, R. T. 1971. Data adaptive spectral analysis methods. Geophysics, *36*, 661–675.

Longuet-Higgins, M. S., D. E. Cartwright and N. D. Smith. 1963. Observation of the directional spectrum of sea waves using the motions of a floating buoy, *in* Ocean Wave Spectra. Englewood Cliffs, New Jersey, Prentice-Hall, 357 pp.

Lumley, J. L. 1970. Stochastic Tools in Turbulence, New York, Academic Press, 194 pp.

McDonough, R. N. 1974. Maximum-entropy spatial processing of array data. Geophysics, *39*, 843–851.

Munk, W. H. and K. Hasselmann. 1964. Super resolution of tides, *in* Studies on Oceanography. K. Yoshida, ed., Univ. of Washington Press,

Munk, W. H., G. R. Miller, F. E. Snodgrass and N. F. Barber. 1963. Directional recording of swell from distant storms. Phil. Trans. *A 255*, 505–584.

Regier, L. 1975. Observations of the power and directional spectrum of oceanic surface waves. Ph.D. Dissertation, Univ. of Calif., San Diego.

Regier, L. A. and R. E. Davis. 1977. Observations of the power and directional spectrum of oceanic surface waves. J. Mar. Res., *35*, this issue.

Received: *3 September, 1976;* revised: *28 February, 1977.*