The Total Variance of a Periodogram-Based Spectral Estimate of a Stochastic Process With Spectral Uncertainty and Its Application to Classifier Design

Yanwu Zhang, *Member, IEEE*, Arthur B. Baggeroer, *Fellow, IEEE*, and James G. Bellingham

***Abstract*—The variance of a spectral estimate of a stochastic process is essential to the formulation and performance of a spectral classifier. The overall variance of a spectral estimate originates from two sources: the within-class spectral uncertainty and the variance introduced in the spectral estimation procedure. In this paper, we derive the total variance of a periodogram-based spectral estimate under some assumptions. Using this result, we formulate a linear spectral classifier based on Fisher's separability metric. The classifier is used to classify two oceanographic processes: ocean convection versus internal waves.**

***Index Terms*—Classifier, periodogram, spectral uncertainty.**

I. Introduction

THE Fourier transform retains all information in the original time series since the transform basis is complete and orthogonal. By the linear system theory, the Fourier transform produces spectral outputs that are uncorrelated between disjoint bands. The spectrum over disjoint bands provides a nonredundant representation of a signal [1]. Thus we can use spectra to distinguish between different classes of signals. Spectral classification has been applied to classifying speech [2], [3], images [4]–[6], trees [7], and sea beds [8], [9].

The variance of a spectral estimate of a stochastic process is essential for a classifier's formulation and performance [10], [11]. The overall variance of a spectral estimate originates from two sources: the within-class spectral uncertainty and the variance introduced in the spectral estimation procedure. An analytical derivation of the total variance is not seen in the literature. An insight of the composition of the total variance is needed for formulating a spectral classifier, analyzing its performance, and devising approaches to reduce the total variance and thus improving the classifier's performance.

In this paper, we analyze the formation of the total variance of a periodogram-based spectral estimate of a stochastic process with spectral uncertainty. The total variance is derived under some assumptions. The result gives a clear and concise expression of the contributions of the two variance sources. Thus it provides a quantitative guide for devising approaches to reduce the total variance. Using the total variance expression, we formulate a linear spectral classifier based on Fisher's separability metric [10]. As an application example, the classifier is used to classify two oceanographic processes.

Manuscript received January 30, 2004. This work was supported in part by the Office of Naval Research under Grants N00014-95-1-1316 and N00014-97-1-0470, in part by the MIT Sea Grant College Program under Grant NA46RG0434, in part by the Ford Professorship of Ocean Engineering, and in part by the David and Lucile Packard Foundation. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Athina Petropulu.

Y. Zhang is with the Monterey Bay Aquarium Research Institute, Moss Landing, CA 95039 USA (e-mail: yzhang@mbari.org).

A. B. Baggeroer is with the Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA 02139 USA (e-mail: abb@boreas.mit.edu).

J. G. Bellingham is with the Monterey Bay Aquarium Research Institute, Moss Landing, CA 95039 USA (e-mail: jgb@mbari.org).

Digital Object Identifier 10.1109/TSP.2005.859346

II. Total Variance of a Periodogram-Based Spectral Estimate of a Stochastic Process With Spectral Uncertainty

Due to the computational efficiency of fast Fourier transform, the periodogram and its variants [12] are the most commonly used methods for estimating the power spectrum density (PSD) of a stochastic process. Suppose $y(t)$ is one realization of a stochastic process $Y(t)$. Its periodogram is defined as [13]

$$S_w(f|T) = \frac{1}{T}\left|\mathcal{F}[y(t)w(t)]\right|^2 \tag{1}$$

where $w(t)$ is a window function of duration T. $\mathcal{F}[\cdot]$ stands for the Fourier transform.

The periodogram $S_w(f|T)$ is an asymptotically unbiased estimate of the true PSD $S_Y(f)$, but its standard deviation is as large as its asymptotic mean [14]. To reduce the PSD estimate's variance, one can do frequency-domain smoothing (the Daniell method [12]) with an interval of $1/T$, and take the smoothed periodogram as the PSD estimate

$$\hat{S}_Y(f) = \sum_{n=-N_s}^{N_s} h_n S_w\left(f + \frac{n}{T}|T\right) \tag{2}$$

where $2N_s+1$ is the total number of frequency points included in the smoothing, and h_n is the weighting function.

The expectation of $\hat{S}_Y(f)$ formulated in (2) is [14]

$$E[\hat{S}_Y(f)|S_Y(f)] = \sum_{n=-N_s}^{N_s}\left\{h_n\int_{-\infty}^{\infty} S_Y(\xi)\left|W\left[\xi-\left(f+\frac{n}{T}\right)\right]\right|^2 d\xi\right\}. \tag{3}$$

Under the assumption that $Y(t)$ is a Gaussian process, the covariance of $\hat{S}_Y(f)$ is [14] as shown in (4) at the bottom of the next page, where $W(f)$ is the Fourier transform of the window function $w(t)$ and h_n is the weight for frequency-domain smoothing. Notation "$\ldots|S_Y(f)$" stresses that the expectation and the covariance are conditioned on the true PSD $S_Y(f)$.

The variance of $\hat{S}_Y(f)$ is obtained directly from (4) as shown in (5) at the bottom of the page.

When $S_Y(f)$ is smooth across the smoothing bandwidth $(2N_s+1)/T$, $S_Y(f)$ can be pulled out of the integration in (3)

$$E[\hat{S}_Y(f)|S_Y(f)] \approx S_Y(f) \sum_{n=-N_s}^{N_s} \left\{ h_n \int_{-\infty}^{\infty} \left| W\left[\xi - \left(f + \frac{n}{T}\right)\right] \right|^2 d\xi \right\} = S_Y(f) \left(\sum_{n=-N_s}^{N_s} h_n \right) \left[\int_{-\infty}^{\infty} |W(f)|^2 df \right]. \quad (6)$$

Under the constraint $\left(\sum_{n=-N_s}^{N_s} h_n\right)\left[\int_{-\infty}^{\infty} |W(f)|^2 df\right] = 1$, (6) becomes

$$E[\hat{S}_Y(f)|S_Y(f)] \approx S_Y(f). \quad (7)$$

Thus as long as $S_Y(f)$ is smooth over the smoothing bandwidth, and the window and the weighting functions satisfy the above constraint, $\hat{S}_Y(f)$ is approximately an unbiased estimate of $S_Y(f)$.

Similarly, under the smoothness assumption, (5) becomes[1] (8), shown at the bottom of the page, where $1/\nu_{\text{eff}}$ represents the quantity in the curly braces and ν_{eff} is called the "effective number of degrees of freedom." $\hat{S}_Y(f)$ is shown [14] to approximately obey a χ^2 distribution with $2\nu_{\text{eff}}$ degrees of freedom

$$2\nu_{\text{eff}} \frac{\hat{S}_Y(f)}{S_Y(f)} \sim \chi^2(2\nu_{\text{eff}}). \quad (9)$$

Equation (8) shows that the periodogram-based PSD estimate has an inherent variance that is proportional to the square of the true PSD. By (4), we see that the PSD estimates are approximately uncorrelated between frequencies farther apart than $B_w + B_h$, where B_w is the bandwidth of $W(f)$ and $B_h = 2N_s/T$

$$\text{Cov}[\hat{S}_Y(f_1), \hat{S}_Y(f_2)|S_Y(f_1), S_Y(f_2)] \approx 0 \quad \text{when } |f_2 - f_1| \geq B_w + B_h. \quad (10)$$

ν_{eff} is dependent on the window function $w(t)$ and the frequency-domain smoothing function h_n. When $w(t)$ is a boxcar window, and an L-point smoothing with uniform weights is done in the frequency domain, we have $\nu_{\text{eff}} = L$. The variance of $\hat{S}_Y$ is thus reduced by a factor of L. The cost paid is the smearing of the PSD estimate within the smoothing bandwidth, so the total number of frequency points that provide uncorrelated PSD estimates is effectively reduced. The duration of $w(t)$ determines the bandwidth of $W(f)$. With a longer duration, the window's bandwidth is smaller, so that the total number of frequency points that provide uncorrelated PSD estimates increases. The shape of $w(t)$ determines the width of $W(f)$'s mainlobe and the height of its sidelobes. Bartlett, Hanning, and Hamming windows have lower sidelobes than that of a boxcar window, but their mainlobes are wider. A lower sidelobe means less interference from nearby frequencies, but a wider mainlobe means a larger bandwidth of $W(f)$, which leads to fewer frequency points that provide uncorrelated PSD estimates.

For stochastic processes that belong to the same class, their underlying physical properties can vary in a range. For example, the PSD of ocean internal waves' vertical flow velocity is dependent on the buoyancy frequency [15], which is a function of the vertical profile of water density. The buoyancy frequency typically varies in a range of several cycles per hour. Corresponding to different buoyancy frequencies, the PSDs are different. As the buoyancy frequency is uncertain, the PSD of the internal waves' vertical flow velocity correspondingly bears an uncertainty. Thus the variability of the underlying physical properties causes spectral uncertainty of the stochastic processes within the same class. Corresponding to given values of the physical

[1]Zero-frequency is special: $\text{Var}[\hat{S}_Y(0)|S_Y(0)] \approx 2S_Y^2(0)/\nu_{\text{eff}}$. This specialty is properly treated in computations in this paper, but omitted in expressions herein for the sake of conciseness.

$$\text{Cov}[\hat{S}_Y(f_1), \hat{S}_Y(f_2)|S_Y(f_1), S_Y(f_2)] = \sum_{m=-N_s}^{N_s} \sum_{n=-N_s}^{N_s} \left\{ h_m h_n \left| \int_{-\infty}^{\infty} S_Y(\xi) W^*\left[\xi - \left(f_1 + \frac{m}{T}\right)\right] W\left[\xi - \left(f_2 + \frac{n}{T}\right)\right] d\xi \right|^2 \right\} \quad (4)$$

$$\text{Var}[\hat{S}_Y(f)|S_Y(f)] = \sum_{m=-N_s}^{N_s} \sum_{n=-N_s}^{N_s} \left\{ h_m h_n \left| \int_{-\infty}^{\infty} S_Y(\xi) W^*\left[\xi - \left(f + \frac{m}{T}\right)\right] W\left[\xi - \left(f + \frac{n}{T}\right)\right] d\xi \right|^2 \right\} \quad (5)$$

$$\text{Var}[\hat{S}_Y(f)|S_Y(f)] \approx [S_Y(f)]^2 \left\{ \sum_{m=-N_s}^{N_s} \sum_{n=-N_s}^{N_s} h_m h_n \left| \int_{-\infty}^{\infty} W^*\left[\xi - \left(f + \frac{m}{T}\right)\right] W\left[\xi - \left(f + \frac{n}{T}\right)\right] d\xi \right|^2 \right\} = \frac{[S_Y(f)]^2}{\nu_{\text{eff}}} \quad (8)$$

properties of a stochastic process, its PSD $S_Y(f)$ is a deterministic quantity at any frequency f. Over the variation range of the physical properties, however, $S_Y(f)$ is actually a random variable at any frequency. We denote the mean of $S_Y(f)$ by $E[S_Y(f)] = M_Y(f)$. Note that the expectation and the covariance of $\hat{S}_Y(f)$ derived so far are conditioned on $S_Y(f)$. The successive decomposition of the variance of $\hat{S}_Y(f)$ is illustrated in Fig. 1.

Using (7), we have

$$\begin{aligned} E[\hat{S}_Y(f)] &= E\big[E[\hat{S}_Y(f)|S_Y(f)]\big] \\ &\approx E[S_Y(f)] \\ &= M_Y(f). \end{aligned} \tag{11}$$

Thus, as long as $S_Y(f)$ is smooth over the smoothing bandwidth, and the window and the weighting functions satisfy the above-mentioned constraint, $\hat{S}_Y(f)$ is approximately an unbiased estimate of $M_Y(f)$. The variance of $\hat{S}_Y(f)$ at any frequency f is derived as follows, where argument "(f)" is omitted for conciseness, as shown in (12) at the bottom of the page.

We integrate the three terms individually and then sum them.

1) Integration of the first term in (12)

$$\begin{aligned} &\int_{-\infty}^{\infty} (\hat{S}_Y - S_Y)^2\, p_{\hat{S}_Y}(\hat{S}_Y)\, d\hat{S}_Y \\ &= \int_{-\infty}^{\infty} (\hat{S}_Y - S_Y)^2 \left[\int_{-\infty}^{\infty} p_{\hat{S}_Y|S_Y}(\hat{S}_Y|S_Y)\, p_{S_Y}(S_Y) dS_Y\right] d\hat{S}_Y \\ &= \int_{-\infty}^{\infty} p_{S_Y}(S_Y) \left[\int_{-\infty}^{\infty} (\hat{S}_Y - S_Y)^2 p_{\hat{S}_Y|S_Y}(\hat{S}_Y|S_Y) d\hat{S}_Y\right] dS_Y \end{aligned} \tag{13}$$

where by (7) and (8), and the definition of conditional variance, we get

$$\int_{-\infty}^{\infty} (\hat{S}_Y - S_Y)^2 p_{\hat{S}_Y|S_Y}(\hat{S}_Y|S_Y)\, d\hat{S}_Y \approx \mathrm{Var}[\hat{S}_Y|S_Y] \approx \frac{S_Y^2}{\nu_{\mathrm{eff}}}. \tag{14}$$

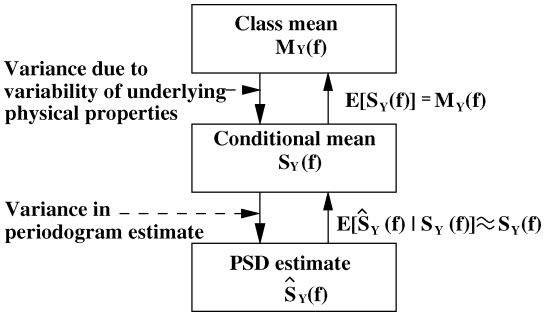

Fig. 1. Relationship between a periodogram-based PSD estimate and the class mean of PSD.

Incorporating (14) into (13), we have (15) shown at the bottom of the page.

By $E[S_Y] = M_Y$, integration of the first term equals M_Y^2; that of the second term equals $\mathrm{Var}[S_Y]$; that of the third term vanishes (note that $\int_{-\infty}^{\infty} S_Y\, p_{S_Y}(S_Y)\, dS_Y = E[S_Y]$). Equation (15) then becomes

$$\int_{-\infty}^{\infty} (\hat{S}_Y - S_Y)^2 p_{\hat{S}_Y}(\hat{S}_Y) d\hat{S}_Y \approx \frac{1}{\nu_{\mathrm{eff}}} \big\{M_Y^2 + \mathrm{Var}[S_Y]\big\}. \tag{16}$$

2) Integration of the second term in (12)

$$\begin{aligned} &\int_{-\infty}^{\infty} (S_Y - M_Y)^2\, p_{\hat{S}_Y}(\hat{S}_Y)\, d\hat{S}_Y \\ &= \int_{-\infty}^{\infty} (S_Y - M_Y)^2 \left[\int_{-\infty}^{\infty} p_{\hat{S}_Y|S_Y}(\hat{S}_Y|S_Y)\, p_{S_Y}(S_Y) dS_Y\right] d\hat{S}_Y \\ &= \int_{-\infty}^{\infty} (S_Y - M_Y)^2\, p_{S_Y}(S_Y) \left[\int_{-\infty}^{\infty} p_{\hat{S}_Y|S_Y}(\hat{S}_Y|S_Y) d\hat{S}_Y\right] dS_Y \\ &= \int_{-\infty}^{\infty} (S_Y - M_Y)^2\, p_{S_Y}(S_Y)\, dS_Y \\ &= \mathrm{Var}[S_Y] \end{aligned} \tag{17}$$

where we note that $\int_{-\infty}^{\infty} p_{\hat{S}_Y|S_Y}(\hat{S}_Y|S_Y) d\hat{S}_Y = 1$ and $E[S_Y] = M_Y$.

$$\begin{aligned} \mathrm{Var}[\hat{S}_Y] &= E\left[\big\{\hat{S}_Y - E[\hat{S}_Y]\big\}^2\right] \approx E[(\hat{S}_Y - M_Y)^2] \\ &= \int_{-\infty}^{\infty} (\hat{S}_Y - M_Y)^2\, p_{\hat{S}_Y}(\hat{S}_Y)\, d\hat{S}_Y = \int_{-\infty}^{\infty} \big[(\hat{S}_Y - S_Y) + (S_Y - M_Y)\big]^2\, p_{\hat{S}_Y}(\hat{S}_Y)\, d\hat{S}_Y \\ &= \int_{-\infty}^{\infty} \big[(\hat{S}_Y - S_Y)^2 + (S_Y - M_Y)^2 + 2(\hat{S}_Y - S_Y)(S_Y - M_Y)\big]\, p_{\hat{S}_Y}(\hat{S}_Y)\, d\hat{S}_Y \end{aligned} \tag{12}$$

$$\begin{aligned} &\int_{-\infty}^{\infty} (\hat{S}_Y - S_Y)^2\, p_{\hat{S}_Y}(\hat{S}_Y)\, d\hat{S}_Y \approx \frac{1}{\nu_{\mathrm{eff}}} \int_{-\infty}^{\infty} S_Y^2\, p_{S_Y}(S_Y)\, dS_Y \\ &= \frac{1}{\nu_{\mathrm{eff}}} \int_{-\infty}^{\infty} \big\{S_Y - E[S_Y] + E[S_Y]\big\}^2\, p_{S_Y}(S_Y)\, dS_Y \\ &= \frac{1}{\nu_{\mathrm{eff}}} \int_{-\infty}^{\infty} \Big\{\big\{E[S_Y]\big\}^2 + \big\{S_Y - E[S_Y]\big\}^2 + 2E[S_Y]\big\{S_Y - E[S_Y]\big\}\Big\}\, p_{S_Y}(S_Y)\, dS_Y \end{aligned} \tag{15}$$

3) Integration of the third term in (12) results in (18), shown at the bottom of the page, noting that

$$\int_{-\infty}^{\infty}(\hat{S}_Y - S_Y)\, p_{\hat{S}_Y|S_Y}(\hat{S}_Y|S_Y)\, d\hat{S}_Y = E[(\hat{S}_Y - S_Y)|S_Y] \approx S_Y - S_Y = 0.$$

Summing (16)–(18), (12) finally reduces to

$$\mathrm{Var}[\hat{S}_Y] \approx \frac{1}{\nu_{\mathrm{eff}}} M_Y^2 + \left(1 + \frac{1}{\nu_{\mathrm{eff}}}\right) \mathrm{Var}[S_Y]. \tag{19}$$

Equation (19) expresses the total variance of the PSD estimate $\hat{S}_Y(f)$. It is composed of the periodogram's inherent variance (the first term) and the variance due to the spectral uncertainty caused by the variability of the underlying physical properties of the process (the second term). One can take two approaches to reduce $\mathrm{Var}[\hat{S}_Y]$: 1) increasing ν_{eff} by time-domain segmentation and/or frequency-domain smoothing and 2) suppressing the spectral uncertainty in each class so as to lower $\mathrm{Var}[S_Y]$. Let us observe (19) in two extreme cases.

1) If $\nu_{\mathrm{eff}} \gg 1$, $\mathrm{Var}[\hat{S}_Y]$ approaches $\mathrm{Var}[S_Y]$. The PSD estimate's variance is dominated by the spectral uncertainty. $\mathrm{Var}[S_Y]$ is actually the lower bound of $\mathrm{Var}[\hat{S}_Y]$.
2) When the spectral uncertainty is negligible, i.e., $\mathrm{Var}[S_Y] \approx 0$, we have $\mathrm{Var}[\hat{S}_Y] \approx (1/\nu_{\mathrm{eff}}) M_Y^2$. The PSD estimate's variance is then dominated by the periodogram's inherent variance.

III. Application to Linear Spectral Classifier Design

A. Feature Extraction Based on Fisher's Criterion

The objective of classification is to determine to which class a given observation belongs [1], [10]. One obtains an observation vector through a measurement process. By a decision rule, the observation vector is assigned to one of the postulated classes. In this paper, we consider two-class problems.

Feature extraction is to choose those components that are most effective for separating classes. It is a process of transforming the observation vector Y to a lower dimensional feature vector Z for classification. The feature extraction algorithm is crucial to the classifier design: if Z carries a significant distinction between classes, the classifier's performance will be good. Selection of the algorithm is based on a class separability criterion. The Bayes error [1] is theoretically the ideal criterion for class separability, but computational difficulties curb its applicability [10].

In this paper, we adopt Fisher's criterion [10]. It represents clear physical meanings and leads to a linear classifier. The linear classifier becomes the Bayes classifier under the condition of Gaussian distributions with equal covariance matrices [1]. When that condition is not met, the linear classifier's performance is inferior to that of the Bayes classifier. In practice, the linear classifier's simplicity and robustness often compensate for its loss in performance [10].

Fisher's criterion is expressed by a within-class scatter matrix and a between-class scatter matrix [10]. The within-class scatter matrix A_{w_Y} depicts the scatter of samples around their respective class means

$$A_{w\_Y} = \sum_{i=1}^{2} P_i E[(Y - M_i)(Y - M_i)^T | H_i] = \sum_{i=1}^{2} (P_i \Sigma_i) \tag{20}$$

where Y is the observation vector (a column vector); $P_i = Pr\{H_i\}$ is the prior probability of class i (H_1 and H_2 denote the two classes); $M_i = E[Y|H_i]$ is the mean vector of Y in class i; and Σ_i is the covariance matrix of Y in class i.

The between-class scatter matrix A_{b_Y} measures the "distance" between the two classes

$$A_{b\_Y} = P_1 P_2 (M_2 - M_1)(M_2 - M_1)^T. \tag{21}$$

Fisher's class separability metric is defined as

$$J_Y = tr(A_{w\_Y}^{-1} A_{b\_Y}) = P_1 P_2\, tr\left\{\left[\sum_{i=1}^{2}(P_i \Sigma_i)\right]^{-1} (M_2 - M_1)(M_2 - M_1)^T\right\}. \tag{22}$$

The trace of a matrix equals the sum of its eigenvalues. So (22) implies that a good separability requires a large between-class scatter and a small within-class scatter. This makes sense.

Consider a linear mapping that transforms the observation vector Y to a lower dimensional feature vector Z: $Z = C^T Y$, where Y is $n \times 1$; Z is $m \times 1$ ($m < n$ for dimension reduction); the transformation matrix C is $n \times m$. The class separability metric in the Z-space is $J_Z = tr(A_{w\_Z}^{-1} A_{b\_Z})$. The optimum transformation matrix C_{optimum} should maximize J_Z. Since the rank of A_{b_Y} equals one, C_{optimum} actually reduces to a vector V [10]

$$C_{\mathrm{optimum}} = V = \beta A_{w\_Y}^{-1} (M_2 - M_1) \tag{23}$$

where β is an arbitrary nonzero coefficient.

$$\begin{aligned}
&\int_{-\infty}^{\infty} 2(\hat{S}_Y - S_Y)(S_Y - M_Y)\, p_{\hat{S}_Y}(\hat{S}_Y)\, d\hat{S}_Y \\
&= 2\int_{-\infty}^{\infty} (S_Y - M_Y)(\hat{S}_Y - S_Y) \left[\int_{-\infty}^{\infty} p_{\hat{S}_Y|S_Y}(\hat{S}_Y|S_Y)\, p_{S_Y}(S_Y)\, dS_Y\right] d\hat{S}_Y \\
&= 2\int_{-\infty}^{\infty} (S_Y - M_Y)\, p_{S_Y}(S_Y) \left[\int_{-\infty}^{\infty} (\hat{S}_Y - S_Y)\, p_{\hat{S}_Y|S_Y}(\hat{S}_Y|S_Y)\, d\hat{S}_Y\right] dS_Y \approx 0
\end{aligned} \tag{18}$$

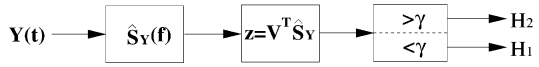

Fig. 2. Diagram of a linear spectral classifier.

Consequently, V transforms the original observation vector Y to a scalar feature z. It can be proved [10] that $J_z = J_Y$. Under Fisher's separability criterion in (22), no separability information is lost due to the dimension reduction. The scalar feature z is a linear function of Y

$$z = V^T Y = \beta(M_2 - M_1)^T A_{w\_Y}^{-1} Y. \tag{24}$$

B. Architecture of a Linear Spectral Classifier

Now we apply (24) to spectral classification. The classifier's architecture is illustrated in Fig. 2. From temporal measurement $Y(t)$, a PSD estimate $\hat{S}_Y(f)$ is obtained at discrete frequencies, expressed as a random column vector $\hat{S}_Y(k)$, $k = 0, 1, \ldots, N-1$, where N is the total number of frequency points. For classification, the vector $\hat{S}_Y$ is transformed to a scalar feature $z = V^T \hat{S}_Y$, where V is the optimum transformation vector that maximizes Fisher's separability, expressed as (the coefficient β is dropped since it does not affect the classifier's performance)

$$V = A_{w\_Y}^{-1}(M_{Y2} - M_{Y1}) \tag{25}$$

where

$$M_{Yi} = E[\hat{S}_Y | H_i] \qquad i = 1, 2 \tag{26}$$

is the class mean and

$$A_{w\_Y} = \sum_{i=1}^{2} (P_i \Sigma_{Yi}) \tag{27}$$

is the within-class scatter matrix, where Σ_{Yi} is the covariance matrix in class i

$$\Sigma_{Yi} = E[(\hat{S}_Y - M_{Yi})(\hat{S}_Y - M_{Yi})^T | H_i] \qquad i = 1, 2. \tag{28}$$

Finally, the scalar feature z is compared with a threshold to make the classification decision

$$z = V^T \hat{S}_Y \underset{\hat{H}_1}{\overset{\hat{H}_2}{\gtrless}} \gamma \tag{29}$$

where threshold γ is selected by minimizing the total cost (in the Bayesian sense) or by satisfying a prescribed false alarm probability (in the Neyman–Pearson sense) [1], [16], [17].

C. Expressing the Linear Spectral Classifier in Terms of the Mean and the Variance of the PSD Estimate

The transformation vector V is the key to the classifier. To compute V as in (25), we need class mean vectors M_{Y1} and M_{Y2}, as well as the within-class scatter matrix A_{w_Y}. Class mean vectors are

$$M_{Y1}(f) = E[\hat{S}_Y(f) | H_1] \tag{30}$$

$$M_{Y2}(f) = E[\hat{S}_Y(f) | H_2]. \tag{31}$$

Matrix A_{w_Y} is composed of covariance matrices Σ_{Y1} and Σ_{Y2}, as expressed in (27). We pick N frequency points with an interval of $B_w + B_h$ so that the PSD estimates are uncorrelated according to (10). Covariance matrix Σ_{Yi} is consequently diagonal

$$\Sigma_{Yi} = diag\{\mathrm{Var}[\hat{S}_{Yi}(0)], \mathrm{Var}[\hat{S}_{Yi}(1)], \ldots, \mathrm{Var}[\hat{S}_{Yi}(N-1)]\} \qquad i = 1, 2. \tag{32}$$

With Σ_{Y1} and Σ_{Y2}, the within-class scatter matrix A_{w_Y} can be constructed by (27) as shown in (33) at the bottom of the page.

Incorporating (33) into (25), we have

$$\begin{aligned} V &= A_{w\_Y}^{-1}(M_{Y2} - M_{Y1}) \\ &= \left[\frac{M_{Y2}(0) - M_{Y1}(0)}{P_1 \mathrm{Var}[\hat{S}_{Y1}(0)] + P_2 \mathrm{Var}[\hat{S}_{Y2}(0)]} \quad \cdots \right. \\ &\qquad \left. \frac{M_{Y2}(N-1) - M_{Y1}(N-1)}{P_1 \mathrm{Var}[\hat{S}_{Y1}(N-1)] + P_2 \mathrm{Var}[\hat{S}_{Y2}(N-1)]} \right]^T. \end{aligned} \tag{34}$$

The numerators in (34) show that the classifier exploits the difference between the means of the two classes of spectra. The bigger that difference, the easier the classification. The denominators, however, restrain the extent of exploiting that difference. Larger spectrum variances mean larger uncertainty of spectra, hence less confidence in the spectral difference. Fig. 3 illustrates the impact of larger spectrum variances on the transformation vector, where P_1 and P_2 are assumed equal, and σ_{Yi} represents $\sqrt{\hat{S}_{Yi}}$. Larger variances increase the denominators in (34), and hence reduce the magnitude of V. Consequently, the difference between the two mean spectra is discounted in classification, which in turn lowers the classifier's performance.

$$\begin{aligned} A_{w\_Y} &= \sum_{i=1}^{2} (P_i \Sigma_{Yi}) \\ &= diag\left\{ P_1 \mathrm{Var}[\hat{S}_{Y1}(0)] + P_2 \mathrm{Var}[\hat{S}_{Y2}(0)], \quad \ldots, \quad P_1 \mathrm{Var}[\hat{S}_{Y1}(N-1)] + P_2 \mathrm{Var}[\hat{S}_{Y2}(N-1)] \right\} \end{aligned} \tag{33}$$

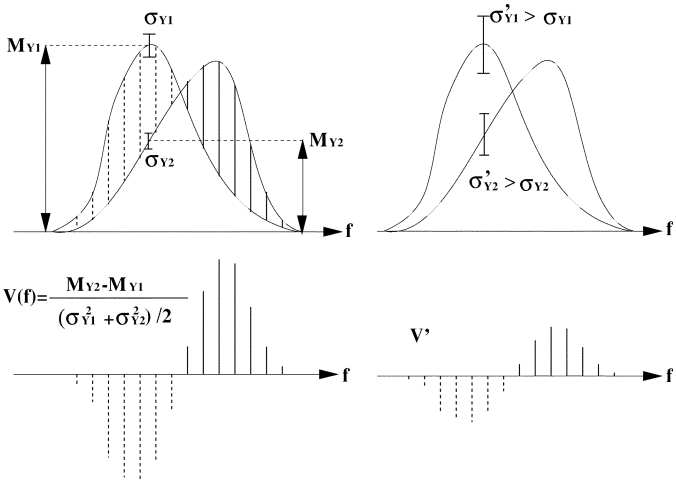

Fig. 3. Larger spectrum variances reduce the magnitude of the transformation vector V.

Incorporating (19) into (34), we get (35) shown at the bottom of the page, and accordingly by (29), the statistic of the linear spectral classifier is expressed as shown in (36) at the bottom of the page.

D. Performance of the Linear Spectral Classifier

For performance comparison purposes, we do not introduce spectral uncertainties into S_{Y1} and S_{Y2}. Hence the variance of $\hat{S}_Y$ is just the periodogram's inherent variance that is proportional to S_{Y1} or S_{Y2}.

Equation (9) shows that under some assumptions, $\hat{S}_Y(f)$ approximately obeys a χ^2 distribution with $2\nu_{\text{eff}}$ degrees of freedom. When ν_{eff} is large, $\hat{S}_Y(f)$ approaches a Gaussian distribution

$$\hat{S}_Y(f) \sim \mathcal{N}\left(S_Y(f), \frac{[S_Y(f)]^2}{\nu_{\text{eff}}}\right). \tag{37}$$

To do classification, we pick N frequency points with an interval of $B_w + B_h$ so that $\hat{S}_Y(k)$ $(k = 1, 2, \ldots, N-1)$ are uncorrelated. Then as $\hat{S}_Y(k)$ approximately obeys a Gaussian distribution, $\hat{S}_Y(k)$ can be further considered to be approximately independent. Thus at a large ν_{eff}, the probability density functions of vector $\hat{S}_Y$ in class 1 and class 2 can be approximated by

$$p_{\hat{S}_Y|H_1}(\hat{S}_Y|H_1) \approx \frac{\exp\left\{-\sum_{k=0}^{N-1} \frac{[\hat{S}_Y(k)-S_{Y1}(k)]^2}{\frac{2S_{Y1}^2(k)}{\nu_{\text{eff}}}}\right\}}{(2\pi)^{N/2}\left[\frac{\prod_{k=0}^{N-1} S_{Y1}(k)}{(\nu_{\text{eff}})^{N/2}}\right]} \tag{38}$$

$$p_{\hat{S}_Y|H_2}(\hat{S}_Y|H_2) \approx \frac{\exp\left\{-\sum_{k=0}^{N-1} \frac{[\hat{S}_Y(k)-S_{Y2}(k)]^2}{\frac{2S_{Y2}^2(k)}{\nu_{\text{eff}}}}\right\}}{(2\pi)^{N/2}\left[\frac{\prod_{k=0}^{N-1} S_{Y2}(k)}{(\nu_{\text{eff}})^{N/2}}\right]}. \tag{39}$$

The likelihood ratio test (LRT) [1] is

$$\frac{p_{\hat{S}_Y|H_2}(\hat{S}_Y|H_2)}{p_{\hat{S}_Y|H_1}(\hat{S}_Y|H_1)} \underset{\hat{H}_1}{\overset{\hat{H}_2}{\gtrless}} \eta. \tag{40}$$

Incorporating (38) and (39) into (40), and taking the natural logarithm of both sides, we derive the LRT, as shown in (41) at the bottom of the page, where $\mu = (2/\nu_{\text{eff}}) \ln\left[\eta \prod_{k=0}^{N-1} S_{Y2}(k) / \prod_{k=0}^{N-1} S_{Y1}(k)\right]$.

$$V \approx \left[\frac{M_{Y2}(0) - M_{Y1}(0)}{\frac{1}{\nu_{\text{eff}}}\left\{P_1[M_{Y1}(0)]^2 + P_2[M_{Y2}(0)]^2\right\} + \left(1 + \frac{1}{\nu_{\text{eff}}}\right)\left\{P_1 \text{Var}[S_{Y1}(0)] + P_2 \text{Var}[S_{Y2}(0)]\right\}} \quad \cdots \right.$$
$$\left. \frac{M_{Y2}(N-1) - M_{Y1}(N-1)}{\frac{1}{\nu_{\text{eff}}}\left\{P_1[M_{Y1}(N-1)]^2 + P_2[M_{Y2}(N-1)]^2\right\} + \left(1 + \frac{1}{\nu_{\text{eff}}}\right)\left\{P_1 \text{Var}[S_{Y1}(N-1)] + P_2 \text{Var}[S_{Y2}(N-1)]\right\}}\right]^T \tag{35}$$

$$z = V^T \hat{S}_Y \approx \sum_{k=0}^{N-1} \left\{ \frac{M_{Y2}(k) - M_{Y1}(k)}{\frac{1}{\nu_{\text{eff}}}\left\{P_1[M_{Y1}(k)]^2 + P_2[M_{Y2}(k)]^2\right\} + \left(1 + \frac{1}{\nu_{\text{eff}}}\right)\left\{P_1 \text{Var}[S_{Y1}(k)] + P_2 \text{Var}[S_{Y2}(k)]\right\}} \hat{S}_Y(k) \right\} \underset{\hat{H}_1}{\overset{\hat{H}_2}{\gtrless}} \gamma \tag{36}$$

$$r = \sum_{k=0}^{N-1} \frac{\hat{S}_Y^2(k)[S_{Y2}^2(k) - S_{Y1}^2(k)] - 2\hat{S}_Y(k) S_{Y1}(k) S_{Y2}(k)[S_{Y2}(k) - S_{Y1}(k)]}{S_{Y1}^2(k) S_{Y2}^2(k)} \underset{\hat{H}_1}{\overset{\hat{H}_2}{\gtrless}} \mu \tag{41}$$

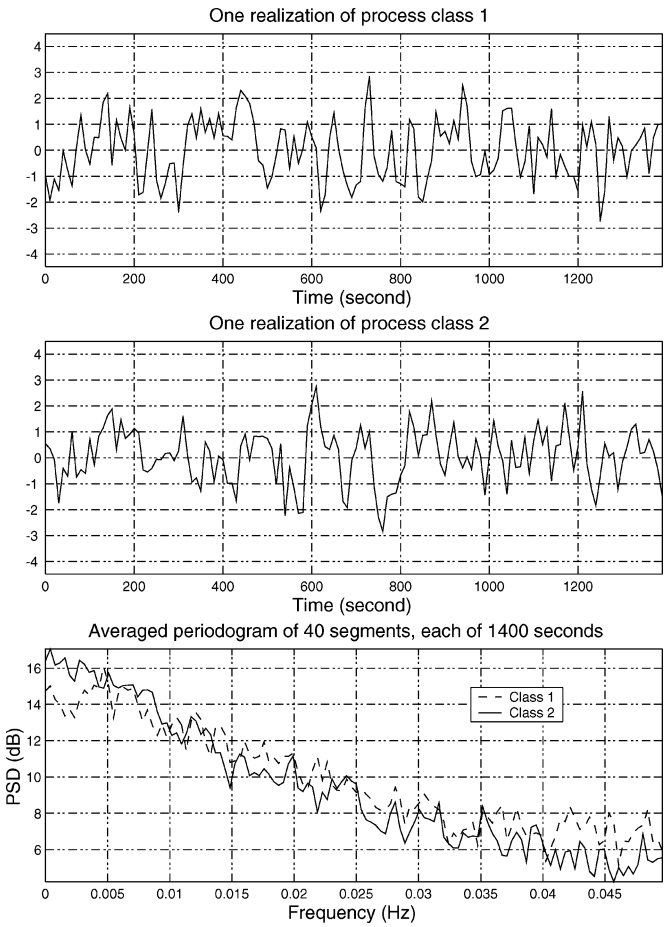

Fig. 4. A 1400-s segment of a realization of process class 1 and class 2 and the corresponding averaged periodograms.

Thus at a large ν_{eff}, the sufficient statistic is a quadratic function of $\hat{S}_Y(k)$. Let us compare the performance of the linear spectral classifier presented in Section III-C and that of the quadratic spectral classifier expressed by (41).

We employ first-order autoregressive (AR) models [18] to generate two classes of stochastic processes

$$y_1(n) = \alpha_1 y_1(n-1) + u_1(n) \tag{42}$$

$$y_2(n) = \alpha_2 y_2(n-1) + u_2(n) \tag{43}$$

where $\alpha_1 = 0.45$, $\alpha_2 = 0.55$. $u_1(n)$ and $u_2(n)$ are two independent white noise sequences.

Each realization of process class 1 or class 2 has a duration of 56 000 s with a sampling rate of 0.1 Hz. It is partitioned into 40 nonoverlapping segments, so each segment's duration is 1400 s. A periodogram is calculated for each segment. The average of the periodograms of all 40 segments is taken as the PSD estimate $\hat{S}_Y(f)$ for classification. Hence $\nu_{\text{eff}} = 40$. A 1400-s segment of a realization of process class 1 and class 2 and the corresponding averaged periodograms are displayed in Fig. 4.

Now we test the linear spectral classifier as shown in Fig. 2. To form the transformation vector V as expressed by (34), we pick frequency points with an interval of $B_w = 1/(1400\text{ s}) = 7.14 \times 10^{-4}$Hz. $P_1 = P_2 = 0.5$ is assumed. We feed 200 56 000-s time series of class 1 and another 200 of class 2 to the linear classifier. The histograms of z are displayed in Fig. 5. The corresponding receiver operating characteristic (ROC) curve [1] is shown by the "o" curve in Fig. 7, where P_F is the probability of declaring class 2 when class 1 is true and P_D is the probability of declaring class 2 when class 2 is indeed true.

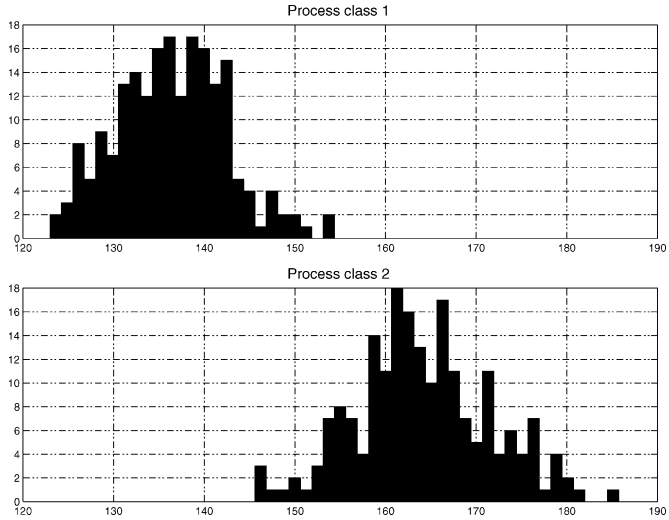

Fig. 5. Histograms of z of the linear spectral classifier.

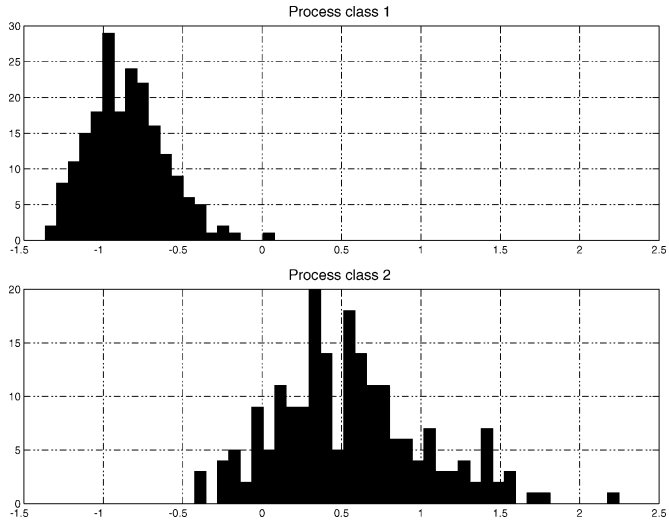

Fig. 6. Histograms of r of the quadratic spectral classifier.

For comparison, we now test the quadratic spectral classifier as expressed by (41), using the same data sets as for testing the linear spectral classifier. Frequency points picked are also the same. The histograms of the sufficient statistic r are displayed in Fig. 6. The corresponding ROC curve is shown by the "+" curve in Fig. 7.

We know that when both classes obey Gaussian distributions, the Bayes classifier is a quadratic classifier [10]. If furthermore, the two classes have equal covariance matrices, the Bayes quadratic classifier reduces to a linear classifier [10]. For the two classes of test processes, α_1 is close to α_2, so $S_{Y1}(f)$ is also close to $S_{Y2}(f)$, as seen in the third panel of Fig. 4. Consequently, the variances of $\hat{S}_Y(f)|H_1 = [S_{Y1}(f)]^2/\nu_{\text{eff}}$ and $\hat{S}_Y(f)|H_2 = [S_{Y2}(f)]^2/\nu_{\text{eff}}$ are close too. Hence the covariance matrices of class 1 and class 2 do not bear much

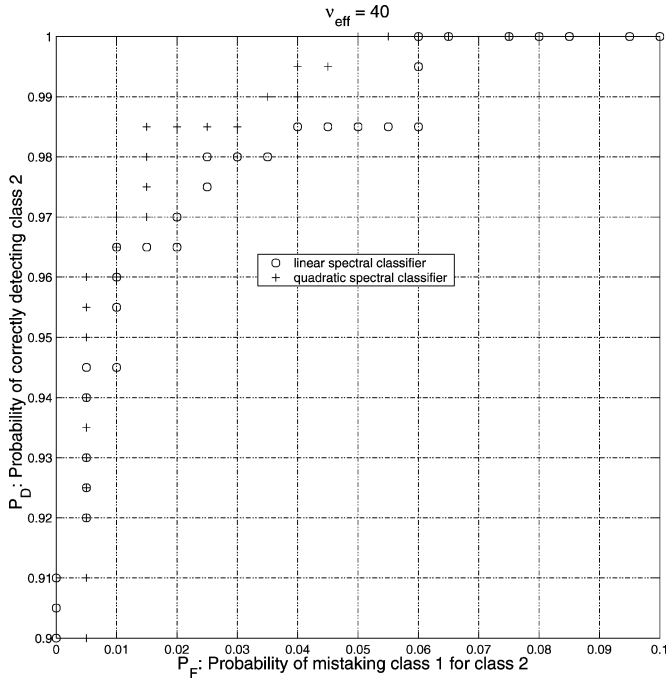

Fig. 7. Comparison of classification performance of the linear spectral classifier and the quadratic spectral classifier.

difference. Therefore, it is not surprising that the linear spectral classifier's performance is close to that of the quadratic spectral classifier. It is noted that if we used more distinct α_1 and α_2, both classifiers would produce nonoverlapping histograms for class 1 and class 2, so that both ROC curves would appear ideal, prohibiting performance comparison.

At a small ν_{eff}, it is improper to approximate a χ^2 distribution to a Gaussian distribution. Then the Bayes classifier is no longer quadratic, and its closed-form expression is difficult to formulate. Consequently, the above performance comparison of the linear spectral classifier and the quadratic spectral classifier does not apply well to small-ν_{eff} cases. Nonetheless, the above performance comparison indicates that the presented linear spectral classifier can be useful.

Note that the linear spectral classifier by (36) requires about N multiplications and N additions, while the quadratic spectral classifier by (41) requires about $2N$ multiplications and $2N$ additions. Hence the computational load of the linear classifier is only half that of the quadratic classifier.

IV. Using the Linear Spectral Classifier to Classify Two Oceanographic Processes

We use the linear spectral classifier to classify two oceanographic processes: ocean convection and internal waves. Convection is the transfer of heat by mass motion of fluid. It happens when the density distribution becomes unstable [19]. Strong winter cooling of ocean surface water causes it to become denser than the water beneath. The cooled surface water sinks and mixes with deeper water which enters the global ocean circulation [20]. This process releases heat from the overturned water to the atmosphere and thus maintains a moderate winter climate on the land. Hence ocean convection is important for global heat transfer.

Internal waves are the water's response to a disturbance to its equilibrium of hydrostatically stable density stratification, via the gravitational restoring force [21]. As opposed to convection, which occurs in a vertically mixed water column, internal waves are found in stably stratified water. Stable stratification is depicted by the buoyancy frequency (also called the Brunt–Väisälä frequency) [21] that is determined by the vertical profile of water density. Internal waves play an important role in mass and momentum transfer in the ocean. Their dynamics is essential for understanding ocean circulation, and temperature and salinity structures [15].

An oceanographic process $X(t,r)$ varies in both time and space. We assume that the studied process is temporally stationary within the survey duration and spatially homogeneous within the survey area. Then the oceanographic process can be described by its temporal-spatial PSD $S_X(\eta,\nu)$, where η is the frequency and ν is the wavenumber. Vertical flow velocities of convection and internal waves show distinct spectral features [22]. Thus we use the measurement of vertical flow velocity for classification. Based on the Massachusetts Institute of Technology (MIT) ocean convection model [20] and the Garrett–Munk internal waves model [15], $S_X(\eta,\nu)$s of convective vertical velocity and internal waves vertical velocity are shown in Fig. 8.

When an autonomous underwater vehicle (AUV) [23] conducts a line survey in an ocean field, as illustrated in Fig. 9, the vehicle makes a time-series measurement $Y(t)$. $Y(t)$ mingles temporal and spatial variations of the process, so we can call it a mingled measurement. Under the assumptions that $X(t,r)$ is temporally stationary and spatially homogeneous, $Y(t)$ is a stationary process. So $Y(t)$ can be depicted by its PSD $S_Y(f)$, which we call the mingled spectrum since it mingles spectral information of time and space. The mingled spectrum principle [22] reveals that $S_Y(f)$ is related to the frequency-wavenumber spectrum $S_X(\eta,\nu)$ by

$$S_Y(f)=\int_{-\infty}^{\infty} S_X\big((f+\nu u),\nu\big)d\nu \tag{44}$$

where u is the AUV's speed.

We apply (44) at $u=1$ m/s. The resultant mingled spectra of convective and internal waves vertical velocities are displayed in the upper panel of Fig. 10. The effect of a 1400-s boxcar window has been included in the computations.

From the viewpoint of spectral classification, $S_Y(f)$ is simply the PSD of the AUV's time series measurement $Y(t)$. To decide which class of oceanographic processes $Y(t)$ belongs to, we feed its periodogram-based PSD estimate $\hat{S}_Y(f)$ to the linear spectral classifier presented in Section III.

In the Labrador Sea experiment in 1998, we installed an acoustic Doppler velocimeter (ADV) in an MIT Odyssey IIB AUV to measure vertical flow velocity [24]. The AUV-borne ADV made a 1400-s measurement of vertical flow velocity at the 250-m depth in Mission B9804107, at a vehicle speed of about 1 m/s. In consideration of the instrument noise, the upper bound of the usable frequency range is about 0.009 Hz [24].

Note that the mean PSDs M_{Y1} and M_{Y2} in the upper panel of Fig. 10 are derived using model parameters listed in the "mean"

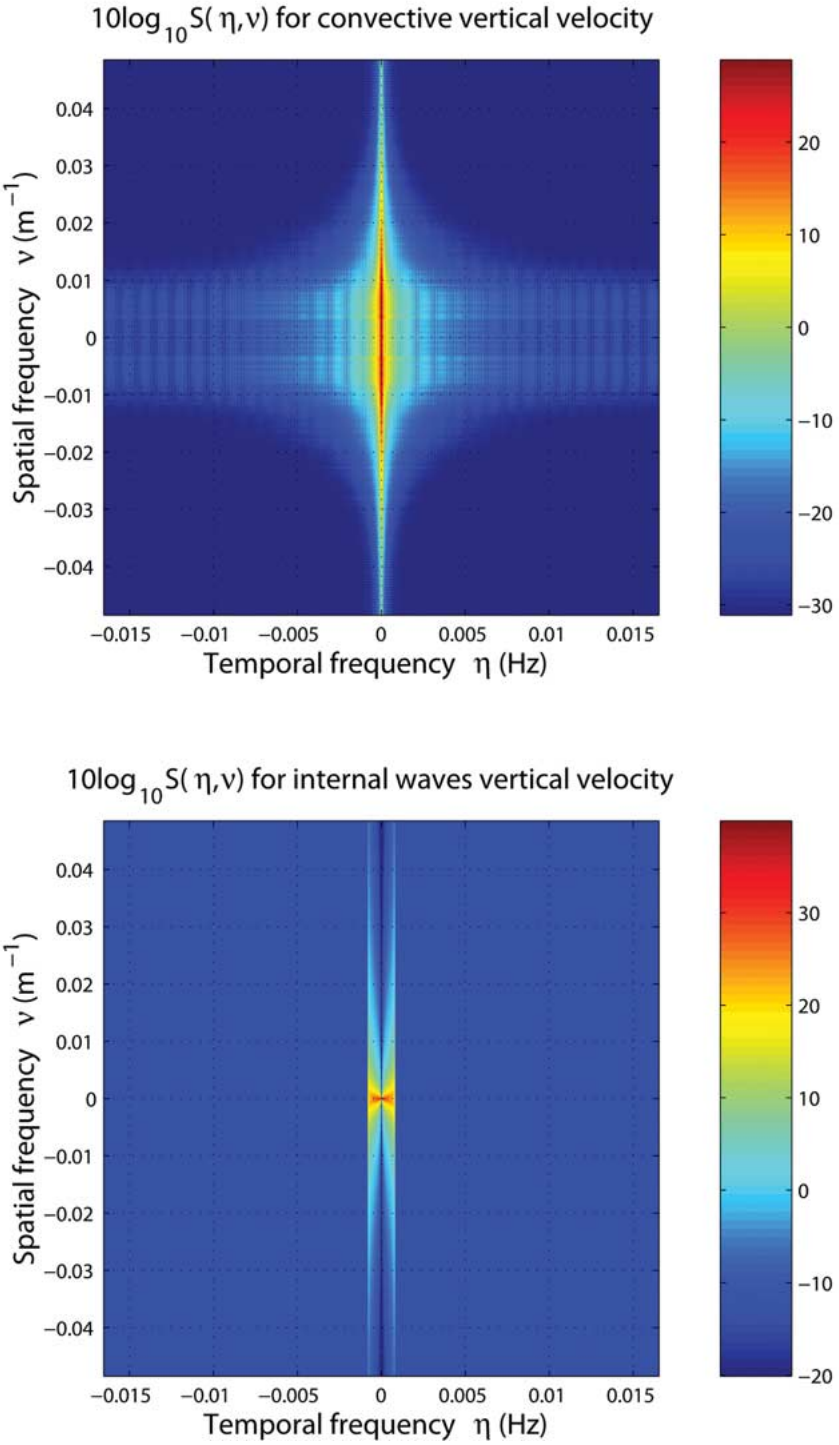

Fig. 8. Temporal-spatial PSD of vertical velocity of (upper) convection and (lower) internal waves, with parameters in the "mean" column of Table I. The unit of vertical bars = $10\log_{10}((\mathrm{m/s})^2/(\mathrm{Hz}\cdot\mathrm{m}^{-1}))$.

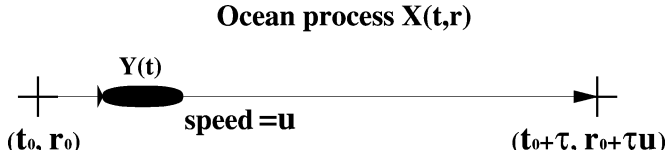

Fig. 9. Line survey of an autonomous underwater vehicle.

column of Table I: for internal waves, the buoyancy frequency $N(z) = 5.0\times10^{-3}$ rad/s; for convection, the surface heat flux $= 300\,\mathrm{W/m^2}$ and the mixed layer depth$= 350$ m. The values of the surface heat flux and the mixed-layer depth for convection are set using the meteorological and hydrographic measurements acquired during the AUV mission.

The values of the physical properties of internal waves and convection can vary due to environmental uncertainties. Internal waves' buoyancy frequency can be as high as 1.0×10^{-2} rad/s. We call internal waves under this setting an "extreme" internal wave case, and denote the resultant mingled spectrum

TABLE I
PARAMETERS FOR M_{Yi} AND $S_{Yi\_extreme}$

	Mean: M_{Yi}	Extreme: $S_{Yi\_extreme}$
Internal waves vertical velocity (i=1)	$N(z) = 5.0 \times 10^{-3}$ rad/s	$N(z) = 1.0 \times 10^{-2}$ rad/s
Convective vertical velocity (i=2)	surface heat flux=300 W/m^2 mixed layer depth=350 m	surface heat flux=900 W/m^2 mixed layer depth=1050 m

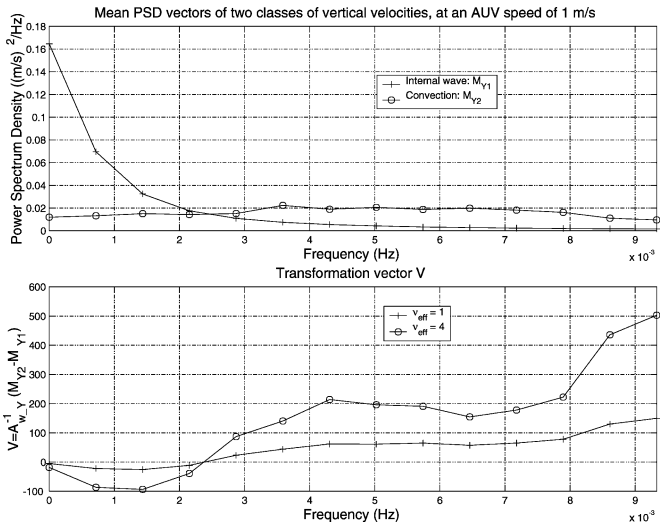

Fig. 10. Class mean vectors M_{Y1} and M_{Y2}, and the feature transformation vector V for AUV speed 1 m/s.

$S_{Y1\_extreme}$. Convection's surface heat flux and mixed layer depth can be as large as 900 W/m^2 and 1050 m, respectively, as listed in the "extreme" column of Table I. We denote the resultant mingled spectrum as $S_{Y2\_extreme}$.

We take $(M_{Y1} - S_{Y1\_extreme})^2$ and $(M_{Y2} - S_{Y2\_extreme})^2$ as the estimates of Var$[S_{Y1}]$ and Var$[S_{Y2}]$, i.e., at frequency k

$$\text{Var}[S_{Y1}(k)] \approx [M_{Y1}(k) - S_{Y1\_extreme}(k)]^2 \quad (45)$$

$$\text{Var}[S_{Y2}(k)] \approx [M_{Y2}(k) - S_{Y2\_extreme}(k)]^2. \quad (46)$$

The 1400-s duration of each time series is short. To retain enough frequency points for uncorrelated PSD estimates, we do no time-domain segmentation or frequency-domain smoothing, so $\nu_{\text{eff}} = 1$.

Ocean internal waves occur in a vertically stratified water column, while ocean convection occurs in a vertically mixed (i.e., unstable) water column. The probability of a water column being stable versus unstable is about equal. Therefore, we set $P_1 = P_2 = 0.5$. Plugging M_{Y1}, M_{Y2}, Var$[S_{Y1}]$, Var$[S_{Y2}]$, ν_{eff}, P_1, and P_2 into (35), we obtain the transformation vector V as shown by the "+" curve in the lower panel of Fig. 10.

Now we test the classifier by convection and internal waves vertical velocity data in simulated AUV line-surveys at a speed of 1 m/s, using the settings of physical property values listed in the "mean" column of Table I. The horizontal size of the convection model is 2000 by 2000 m with a grid size of 10 m, evolving over time. This provides 200 lines, each of 2000 m, in the x or y direction. An AUV traveling at 1 m/s covers 1400 m in 1400 s. We take one 1400-s data series from each of the 200 lines in the x direction alone, thus forming 200 1400-s time series. Two hundred time series of AUV data in the convection field and another 200 in the internal wave field are used in this test. For each time series, its PSD estimate $\hat{S}_Y$ is transformed to a scalar z by (36). The 1400-s measurement of the vertical flow velocity made by the AUV at the 250-m depth in the Labrador Sea [22] is also tested.

The histograms of z are shown in Fig. 11. In this test, the distributions of z in the two classes do not overlap. The stem in Fig. 11 marks the z value corresponding to the AUV's measurement. This experimental z value falls in the cluster of the model-based convection. The classifier thus declares that the AUV-measured vertical flow velocity was convective. This finding is supported by other independent oceanographic observations in the same experiment [22].

Next we test the classifier using the settings of physical property values listed in the "extreme" column of Table I. Compared with the "mean" versus "mean" case, the PSDs of the "extreme" convection's vertical velocity comes closer to that of the "extreme" internal waves [25]. We select this more difficult case to test the classifier's robustness. Two hundred lines of AUV data in the convection field and another 200 lines in the internal waves field are used in this test. Fig. 12 shows some overlap of z clusters between the two classes. This degradation in classification performance is expected. The ROC curves of the "mean" versus "mean" case and the "extreme" versus "extreme" case are compared in Fig. 13.

When longer time series are available, we can apply time-domain segmentation to increase ν_{eff} so as to reduce the periodogram's inherent variance. Now we lengthen each simulation time series to 5600 s $=$ 1400 s $\times$ 4. Each time series contains samples from three adjacent 2000-m lines in the convection model, one series followed by another, but with no overlap. This way, the x direction can provide $(2000 \text{ m} \times 200)/(5600 \text{ m}) = 71$ time series. We take another 29 5600-s time series from the y direction. We partition each 5600-s time series into four nonoverlapping segments. The average of the periodograms of the four segments is fed to the spectral classifier. Thus ν_{eff} equals four, which reduces the periodogram's inherent variance by a factor of four.

A larger ν_{eff} lowers Var$[\hat{S}_{Y1}]$ and Var$[\hat{S}_{Y2}]$ by (19). The magnitude of the transformation vector V is consequently larger (illustrated in Fig. 3). When $\nu_{\text{eff}} = 4$ at AUV speed 1 m/s, V is shown by the "o" curve in the lower panel of Fig. 10. Its magnitude is larger than when $\nu_{\text{eff}} = 1$.

When $\nu_{\text{eff}} = 4$, we test the classifier again by the "extreme" convection versus the "extreme" internal waves case. The classifier's performance is shown in Fig. 14. Compared with $\nu_{\text{eff}} = 1$

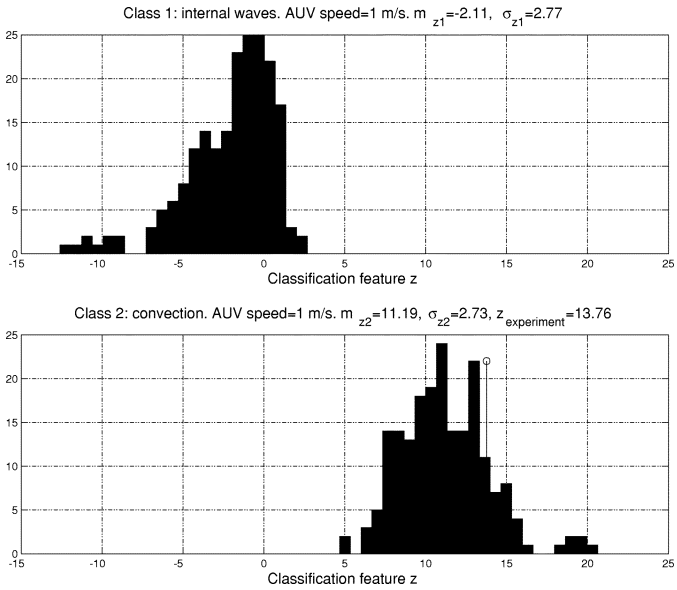

Fig. 11. At an AUV speed of 1 m/s and $\nu_{\text{eff}} = 1$, histograms of feature z for internal waves and convection. The value of $z_{experiment}$ corresponding to the AUV's measurement is marked by the stem's horizontal location.

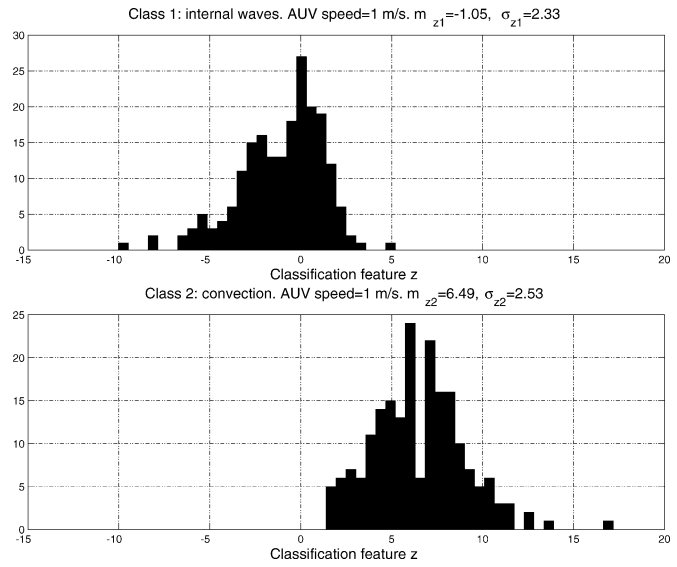

Fig. 12. At an AUV speed of 1 m/s and $\nu_{\text{eff}} = 1$, histograms of feature z for the "extreme" internal waves versus the "extreme" convection.

in Fig. 11, z clusters of the two classes are pulled apart. Hence the classifier's performance improves.

V. Conclusion

The overall variance of a spectral estimate originates from two sources: the within-class spectral uncertainty due to uncertain values of the underlying physical properties and the variance introduced in the spectral estimation procedure. In this paper, we derived the total variance of a periodogram-based spectral estimate under some assumptions. Using the total variance expression, we formulated a linear spectral classifier based on Fisher's separability metric. The classifier is applied to classifying vertical flow velocities of two oceanographic

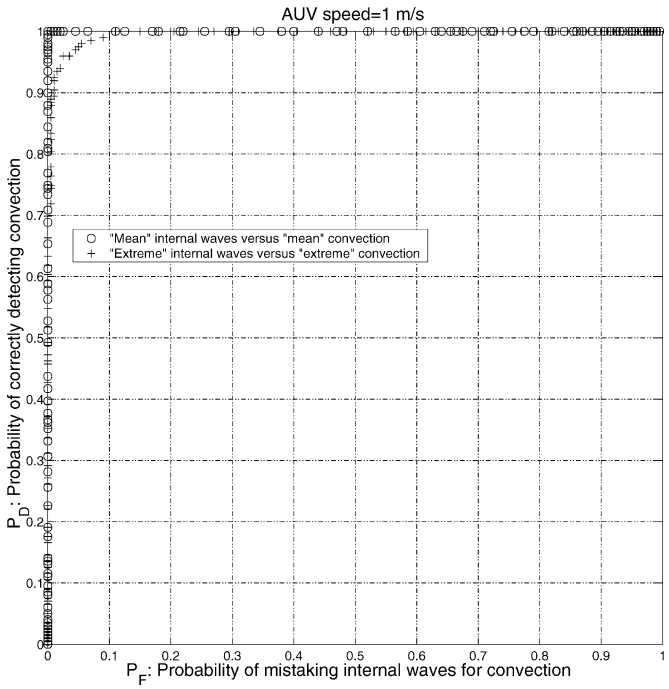

Fig. 13. Classification performance of the "mean" convection versus the "mean" internal waves, and the "extreme" convection versus the "extreme" internal waves.

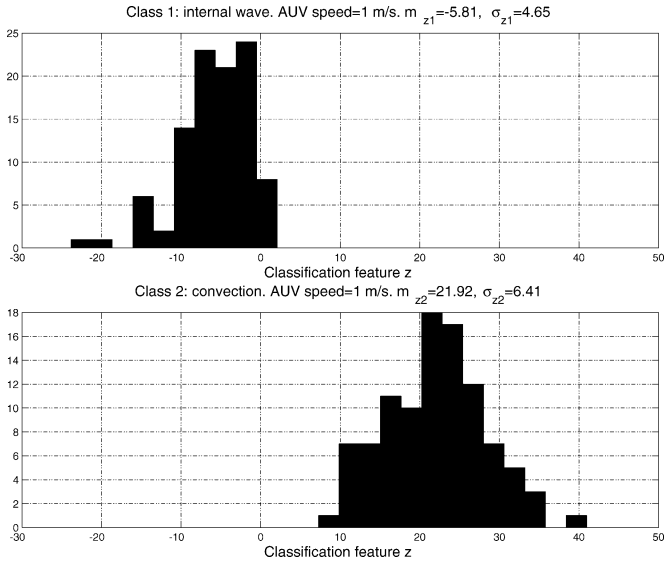

Fig. 14. At AUV speed 1 m/s and $\nu_{\text{eff}} = 4$, histograms of feature z for the "extreme" internal waves versus the "extreme" convection.

processes: ocean convection versus internal waves. The impact of the within-class spectral uncertainty on the classifier's performance is shown. It is also demonstrated that reducing the periodogram's inherent variance improves the classifier's performance.

Acknowledgment

The authors would like to thank Dr. A. Williams III of the Woods Hole Oceanographic Institution for his important help in the Labrador Sea Experiment data analysis. The authors are thankful to Office of Naval Research (ONR) Program Manager

Dr. T. Curtin and MIT Sea Grant Director Prof. C. Chryssostomidis for their support. The MIT ocean convection model was kindly provided by MIT Prof. J. Marshall and Dr. H. Hill. The authors appreciate the valuable comments from the four reviewers.

REFERENCES

[1] H. L. Van Trees, *Detection, Estimation, and Modulation Theory, Part I*. New York: Wiley, 1968.
[2] H. K. Kim and H.-S. Lee, "Use of spectral autocorrelation in spectral envelope linear prediction for speech recognition," *IEEE Trans. Speech Audio Process.*, vol. 7, pp. 533–541, Sep. 1999.
[3] L. Bu and T.-D. Church, "Perceptual speech processing and phonetic feature mapping for robust vowel recognition," *IEEE Trans. Speech Audio Process.*, vol. 8, pp. 105–114, Mar. 2000.
[4] C.-I. Chang, "An information theoretic approach to spectral variability, similarity, and discrimination for hyperspectral image analysis," *IEEE Trans. Inform. Theory*, pp. 1927–1932, Aug. 2000.
[5] X. Yu, L. E. Hoff, I. S. Reed, A. M. Chen, and L. B. Stotts, "Automatic target detection and recognition in multiband imagery: A unified ML detection and estimation approach," *IEEE Trans. Image Process.*, vol. 6, pp. 143–156, Jan. 1997.
[6] G. Gelli and G. Poggi, "Compression of multispectral images by spectral classification and transform coding," *IEEE Trans. Image Process.*, vol. 8, pp. 476–489, Apr. 1999.
[7] B. Yu, I. M. Ostland, P. Gong, and R. Pu, "Penalized discriminant analysis of in situ hyperspectral data for conifer species recognition," *IEEE Trans. Geosci. Remote Sensing*, vol. 37, pp. 2569–2577, Sep. 1999.
[8] W. K. Stewart, D. Chu, S. Malik, S. Lerner, and H. Singh, "Quantitative seafloor characterization using a bathymetric sidescan sonar," *IEEE J. Ocean. Eng.*, vol. 19, pp. 599–610, Oct. 1994.
[9] Z. Reut, "Computer classification of sea beds by sonar," *Nature*, vol. 314, pp. 426–428, Apr. 1985.
[10] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, 2nd ed. New York: Academic, 1990.
[11] T. Hastie, A. Buja, and R. Tibshirani, "Penalized discriminant analysis," *Ann. Statist.*, vol. 23, no. 1, pp. 73–102, 1995.
[12] P. Stoica and R. Moses, *Introduction to Spectral Analysis*. Upper Saddle River, NJ: Prentice-Hall, 1997.
[13] A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1989.
[14] G. M. Jenkins and D. G. Watts, *Spectral Analysis and its Applications*. San Francisco, CA: Holden-Day, 1968.
[15] C. Garrett and W. Munk, "Internal waves in the ocean," *Annu. Rev. Fluid Mech.*, vol. 11, pp. 339–369, 1979.
[16] C. W. Therrien, *Decision, Estimation, and Classification: An Introduction to Pattern Recognition and Related Topics*. New York: Wiley, 1989.
[17] S. M. Kay, *Fundamentals of Statistical Signal Processing: Detection Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1998.
[18] ——, *Modern Spectral Estimation : Theory and Application*. Englewood Cliffs, NJ: Prentice-Hall, 1988.
[19] G. L. Pickard and W. J. Emery, *Descriptive Physical Oceanography: An Introduction*. New York: Pergamon, 1990.
[20] H. Jones and J. Marshall, "Convection with rotation in a neutral ocean: A study of open-ocean convection," *J. Phys. Ocean.*, vol. 23, no. 6, pp. 1009–1039, Jun. 1993.
[21] J. R. Apel, *Principles of Ocean Physics*. London, U.K.: Academic, 1987.
[22] Y. Zhang, A. B. Baggeroer, and J. G. Bellingham, "Spectral-feature classification of oceanographic processes using an autonomous underwater vehicle," *IEEE J. Ocean. Eng.*, vol. 26, pp. 726–741, Oct. 2001.
[23] J. G. Bellingham, "New oceanographic uses of autonomous underwater vehicles," *Marine Technol. Soc. J.*, vol. 31, no. 3, pp. 34–47, 1997.
[24] Y. Zhang, K. Streitlien, J. G. Bellingham, and A. B. Baggeroer, "Acoustic Doppler velocimeter flow measurement from an autonomous underwater vehicle with applications to deep ocean convection," *J. Atmospheric Ocean. Technol.*, vol. 18, no. 12, pp. 2038–2051, Dec. 2001.
[25] Y. Zhang, "Spectral feature classification of oceanographic processes using an autonomous underwater vehicle," Ph.D. dissertation, Massachusetts Institute of Technology and Woods Hole Oceanographic Institution, 2000.

Yanwu Zhang (S'95–M'00) was born in 1969 in Shaanxi Province, China. He received the B.S. degree in electrical engineering and the M.S. degree in underwater acoustics engineering from Northwestern Polytechnic University, Xi'an, China, in 1989 and 1991, respectively. He received the M.S. degree in electrical engineering and computer science from the Massachusetts Institute of Technology (MIT), Cambridge, in 1998 and the Ph.D. degree in oceanographic engineering from the MIT/Woods Hole Oceanographic Institution Joint Program in 2000.

From June to December 2000, he was a Systems Engineer working on medical image processing with the General Electric Company Research and Development Center, Niskayuna, NY. From 2001 to 2004, he was a Digital Signal Processing (DSP) Engineer and then a Senior DSP Engineer with Aware Inc., Bedford, MA, working on digital communications. Since December 2004, he has been with the Research and Development Division of the Monterey Bay Aquarium Research Institute, as an Engineering Research Specialist engaged in autonomous underwater vehicles (AUVs) and ocean observing systems. His current research interests are mainly in spatiotemporal signal processing and its applications to AUVs' sampling of oceanographic processes. He has published papers in IEEE JOURNAL OF OCEANIC ENGINEERING, IEEE TRANSACTIONS ON NEURAL NETWORKS, *Journal of Atmospheric and Oceanic Technology,* and conference proceedings. He is the author of a chapter on "Complex-Valued Generalized Hebbian Algorithm and Its Applications to Sensor Array Signal Processing" in *Complex-Valued Neural Networks: Theories and Applications* (Singapore: World Scientific, 2003).

Dr. Zhang is a member of Sigma Xi. He was a Finalist for MIT *Technology Review Magazine*'s 100 young innovators (TR100) in 1999.

Arthur B. Baggeroer (S'62–M'68–SM'87–F'89) received the B.S.E.E. degree from Purdue University, West Lafayette, IN, in 1963 and the Sc.D. degree from the Massachusetts Institute of Technology (MIT), Cambridge, in 1968.

He is a Ford Professor of Engineering in the Department of Electrical Engineering and Computer Science, MIT. He has been a Consultant to the Chief of Naval Research at the NATO SACLANT Center in 1977 and a Cecil and Ida Green Scholar at the Scripps Institution of Oceanography in 1990 while on sabbatical leaves. His research has been concerned with sonar array processing, acoustic telemetry, and, most recently, global acoustics and matched field array processing. He also has had a long affiliation with the Woods Hole Oceanographic Institution (WHOI) and was Director of the MIT/WHOI Joint Program from 1983 to 1988.

Dr. Baggeroer is a Fellow of the Acoustical Society of America (ASA). He received the IEEE Oceanic Engineering Society Distinguished Technical Achievement Award in 1991. He was an elected member of the Executive Council of the ASA from 1994 to 1997, and was awarded the Rayleigh-Helmholtz Medal from the ASA in 2003. He was elected to the National Academy of Engineering in 1995. He was awarded the Secretary of the Navy/Chief of Naval Operations Chair in Oceanographic Science in 1998.

James G. Bellingham received the B.S. and Ph.D. degrees in physics from the Massachusetts Institute of Technology (MIT), Cambridge, in 1984 and 1988, respectively.

He is Director of Engineering at the Monterey Bay Aquarium Research Institute (MBARI). Over the last 16 years he has led the design, high-level control development, and operation of autonomous underwater vehicles (AUVs). He is a Founder and Member of the Board of Directors of Bluefin Robotics Corporation, a manufacturer of AUVs for the military, commercial, and scientific markets.

Dr. Bellingham received the Lockheed Martin Award for Ocean Science and Engineering from the Marine Technology Society in 2004. He was a Steinbach Visiting Scholar at the Woods Hole Oceanographic Institution in 2004. He received the Antarctica Service Medal of the United States of America from the National Science Foundation and the Antarctica Service Medal from the Department of the Navy, both in 1993.