MASSACHUSETTS INSTITUTE OF TECHNOLOGY
LINCOLN LABORATORY

DOPPLER MEAN VELOCITY ESTIMATION: SMALL SAMPLE ANALYSIS AND A NEW ESTIMATOR

E. S. CHORNOBOY

TECHNICAL REPORT 942

25 MARCH 1992

LEXINGTON MASSACHUSETTS

TECHNICAL REPORT STANDARD TITLE PAGE

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
TR-942	DOT/FAA/NR-92/5	

4. Title and Subtitle	5. Report Date
Doppler Mean Velocity Estimation: Small Sample Analysis and a New Estimator	25 March 1992
	6. Performing Organization Code

7. Author(s)	8. Performing Organization Report No.
Edward S. Chornoboy	TR-942

9. Performing Organization Name and Address	10. Work Unit No. (TRAIS)
Lincoln Laboratory, MIT P.O. Box 73 Lexington, MA 02173-9108	11. Contract or Grant No. DTFA 01-84-Z-02030

12. Sponsoring Agency Name and Address	13. Type of Report and Period Covered
Department of Transportation Federal Aviation Administration Systems Research and Development Service Washington, DC 20591	Technical Report
	14. Sponsoring Agency Code

15. Supplementary Notes

This report is based on studies performed at Lincoln Laboratory, a center for research operated by Massachusetts Institute of Technology under Air Force Contract F19628-90-C-0002.

16. Abstract

Optimal Doppler velocity estimation, under the constraint of small sample size, is explored for a standard Gaussian signal measurement model and thematic maximum likelihood (ML) and Bayes estimation. Because the model considered depends on a vector parameter [velocity, spectrum width, and signal-to-noise ratio (SNR)], the exact formulation of an ML or Bayes solution involves a system of equations that is neither uncoupled nor explicit in form. Historically, iterative methods have been the most suggested approach to solving the required equations. In addition to being computationally intensive, it is unclear whether iterative methods can be constructed to perform well given a small-sample size and low signal strength.

This report takes a different approach and seeks to construct approximate (ML and Bayes) estimators based on the notion of using constrained adaptive models to deal with nuisance parameter removal. A Monte Carlo simulation is used to determine small-sample estimator statistics and to demonstrate true performance bounds in the case of known nuisance values. Performance comparisons between these optimal forms and other standard estimators [pulse pair (PP) and a frequency domain wind profiler (WP) method] are presented. Performance sensitivity of the optimal algorithms, with respect to uncertainty in the values of model nuisance parameters, is explored and provides the foundation for the recommendation to seek an adaptive method.

An adaptive estimation method, closely allied with the derived ML and Bayes formulas, is developed using information theoretic methods to constrain the adaptation process. The improved, near optimal performance of the new Doppler velocity estimator is documented by comparison to derived optimal bounds and to the performance of the PP method.

17. Key Words		18. Distribution Statement
adaptive	velocity	Document is available to the public through the National Technical Information Service, Springfield, VA 22161.
Doppler	maximum likelihood	
estimation	Bayes	

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	78	

FORM DOT F 1700.7 (8-72) Reproduction of completed page authorized

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF ILLUSTRATIONS
(Continued)

LIST OF ILLUSTRATIONS
(Continued)

LIST OF TABLES

1. INTRODUCTION

Mean velocity (i.e., first spectral moment) estimation is central to pulsed-Doppler weather-radar processing. However, some radars, such as the Terminal Doppler Weather Radar (TDWR) and Airport Surveillance Radar (ASR), operate under conditions that test the limits of current capabilities. For example, high scan rate and fine (azimuth) sampling requirements yield a small number of pulse samples per gate[1] (i.e., small-sample size) while clear-air conditions result in low signal-to-noise levels. Hence, coupled with general interest for optimal velocity estimation, specific needs motivate an investigation of improved performance under the constraints of small-sample size and low signal strength.

As is often the case, the issue of practicability limits the degree to which optimal performance can be attained. Further complicating matters in this report is an assumed small-sample constraint that makes the definition of optimality itself problematic. Indeed, the natural appeal to the (asymptotic) optimality of maximum likelihood (ML) could fail; verification is clearly required. The Cramèr-Rao (CR) performance bound could be too optimistic, likely unattainable by any estimator and therefore useless in gauging what room there is for improvement. Because few theoretical results exist to serve as guides, numerical methods typically must be relied upon for the relevant analysis.

This report presents a study of the velocity estimation problem and provides arguments for frequency-domain smoothing (autocorrelation-lag weighting) as a means of achieving improved velocity estimation performance. Although the emphasis is on small-sample analysis, the results are not restricted to the small-sample case. Note that the Bayes (conditional mean) estimator (defined by the usual formulation) minimizes the mean-squared estimation error over all estimators. This property holds regardless of sample size and hence rationalizes the use of Bayes performance as a benchmark for optimality in this small-sample case. Unfortunately, even assuming the conventional Gaussian signal model, the conditional mean is computationally impractical. It nevertheless provides useful insight to the development of improved (and practicable) frequency-domain velocity estimators.

At the heart of this study is a Monte Carlo analysis examining small-sample performance for selected mean-frequency estimators. In addition to the pulse pair (PP) estimator and a typical Fast Fourier Transform (FFT) estimator [i.e., a periodogram based estimator derived from wind-profiler (WP) literature], maximum likelihood (ML) and Bayes estimators (these two based on a Gaussian signal model) are examined and compared. Sample size is fixed to a small ($M = 20$) value, and performance at low signal-to-noise ratios (SNR) is also a point of focus. The results of

[1]For TDWR, on the order of 40 pulse samples are coherently processed per output velocity estimate; for ASR-9, on the order of 18 (although data sharing has been used to extend this value to 27).

these analyses are used to recommend an optimal smoothing (weighting) strategy for periodogram (autocorrelation-lag) based estimation.

As stated, the motivating interest is improved velocity estimation for Doppler weather radars, and the content of this report is clearly oriented toward weather-radar processing. However, the results of Section 6 should be of a more general interest, presenting a new and novel method for dealing with the fundamental problem of nuisance parameter removal under the constraint of a limited sample size.

2. BACKGROUND

2.1 Statistical Model

For this report, it is assumed that a frequently adopted Gaussian model (for single-source weather returns in additive noise) is appropriate (see, for example, Doviak and Zrnić [1]). This model is characterized by the sample covariance relation

$$r_m = \mathcal{S} e^{-8(\pi\sigma_v m\tau/\lambda)^2} e^{-j4\pi\bar{v}m\tau/\lambda} + \mathcal{N}\delta_m, \tag{1}$$

where the model parameters $\bar{v}$ and σ_v represent mean Doppler velocity and spectrum width, λ is the radar wavelength, and $\mathcal{S}$ and $\mathcal{N}$ respectively represent signal and noise power magnitudes. Complex-valued radar returns, $\mathbf{Z}^T = [z_0\ z_1 \ldots z_{M-1}]$,[2] corresponding to a single range cell are assumed to represent M equally-spaced samples (separated in time by pulse separation τ) of a complex Gaussian process with covariance matrix $\mathbf{R} = E[ZZ^\dagger]$. In view of Equation (1), $\mathbf{R}$ can be given the parametric representation

$$\mathbf{R} = \mathbf{D}\,[\mathcal{S}\mathbf{G} + \mathcal{N}\mathbf{I}]\mathbf{D}^* = \mathcal{S}\mathbf{D}\mathbf{G}\mathbf{D}^* + \mathcal{N}\mathbf{I}, \tag{2}$$

where

$$\mathbf{D} = \mathbf{D}(\omega) = \mathrm{diag}[1\ e^{-j\omega}\ \ldots\ e^{-j(M-1)\omega}],$$

$$\mathbf{G} = \mathbf{G}(\sigma) = \begin{bmatrix} \rho(0) & \rho(1) & \cdots & \rho(M-1) \\ \rho(1) & \rho(0) & \cdots & \rho(M-2) \\ \vdots & \vdots & \ddots & \vdots \\ \rho(M-1) & \rho(M-2) & \cdots & \rho(0) \end{bmatrix},$$

$$\mathbf{I} = \mathrm{diag}[1\ 1 \ldots 1],$$

$$\rho(m) = e^{-\frac{1}{2}(\pi m\sigma)^2},\quad \omega = -\frac{\bar{v}}{v_{Nyq}}\pi,\quad \sigma = \frac{\sigma_v}{v_{Nyq}},\quad \text{and}\quad v_{Nyq} = \frac{\lambda}{4\tau}.$$

---

[2]Notationally, a superscript "T" will be used to indicate a matrix transpose, a superscript "*" will be used to indicate complex conjugate, and a superscript "†" will be used to indicate conjugate transpose.

It is assumed that $\mathcal{N}$ can be reliably estimated and treated as known, and for convenience $\mathcal{S}$ and $\mathcal{N}$ are combined into an unknown SNR $\eta = \mathcal{S}/\mathcal{N}$. The complex data vector $\mathbf{Z}$ is therefore governed by the probability density

$$p(\mathbf{Z} \mid \Theta) = \pi^{-M} |\mathbf{R}|^{-1} e^{-\mathbf{Z}^\dagger \mathbf{R}^{-1} \mathbf{Z}}, \tag{3}$$

and estimation of the Doppler velocity ω must be done in the context of the unknown parameter vector $\Theta^T = [\omega \; \sigma \; \eta]$.

2.2 Experimental Design

2.2.1 Weather-Radar Simulation

Weatherlike Doppler signals were simulated as described by Zrnić [2]. For all simulations, digital sampling of Gaussian shaped power spectra was emulated, including Nyquist folding, until all tail values with power greater than 0.01 were accounted for. Given an M point input power spectrum, Zrnić's method generates a random vector of M consecutive process samples. A brief study of sample estimator performance was made to examine (and thereby ensure limits for) uncertainties due to power spectral sampling and Monte Carlo sample size. As a result of this investigation, a 1024 point power spectrum representation was used and, from the corresponding output process vector, a block of 20 contiguous samples was extracted for use as a data sample (the remaining 1004 points were discarded). Also, a total of 10,000 (independent) realizations were used for each assessment of sample mean and standard error.

Error results are presented in both normalized and unnormalized form. In view of the definitions in Equation (2), spectrum width and standard error values are normalized to the magnitude of the Nyquist velocity v_{Nyq}. Bias errors are normalized to the magnitude of the true underlying signal velocity. Unnormalized values are presented in the context of weather-radar processing using a 10.4 cm radar operating with a pulse repetition frequency of 1000 s^{-1}. (The unambiguous Nyquist velocity therefore equals 26 m/s.)

Provided also, for reference, are curves corresponding to the CR bounds that were computed using the well-known result[3]

$$f_{i,j} = \mathrm{tr}\left\{ \mathbf{R}^{-1} \frac{\partial \mathbf{R}}{\partial \theta_i} \mathbf{R}^{-1} \frac{\partial \mathbf{R}}{\partial \theta_j} \right\}, \tag{4}$$

[3]A derivation is provided as an appendix.

where $\mathbf{F} = [f_{i,j}]$ represents the Fisher information matrix

$$\mathbf{F} = E\left[\left(\frac{\partial \ln p(\mathrm{Z}|\Theta)}{\partial \Theta}\right)^T \left(\frac{\partial \ln p(\mathrm{Z}|\Theta)}{\partial \Theta}\right)\right].$$

The diagonal entries of $\mathbf{F}$ provide the desired bounds:

$$E[(\hat{\theta}_i - \theta_i)^2] \geq \overline{f}_{i,i}, \tag{5}$$

where $\mathbf{F}^{-1} = [\overline{f}_{i,j}]$ and $\hat{\theta}_i$ is any unbiased estimator of component θ_i of Θ.

2.2.2 Test-Signal Parameters

Performance evaluations/comparisons are based on estimation error that clearly can vary as a function of the true parameter value. Error quantification for all possible values of the (true) parameter is not feasible. Evaluations are based, therefore, on comparisons for selected parameter values.

Target velocities of 0, 5 (0.192 v_{Nyq}), and 13 m/s (0.500 v_{Nyq}) are used in the present evaluation. Preliminary studies obtained similar error profiles for velocities in the range 0 to 0.5 v_{Nyq}. Velocities greater than 0.5 v_{Nyq} are not examined to avoid the added complexity of spectral folding.

Values for weather spectrum width can range from 0 to v_{Nyq}. For coherent processing, only a small range is of interest. A further simplification is made by considering only extreme points of this range providing two classes of input signal: narrow and wide spectrum width. Table 1 contains

TABLE 1

Model Sample Correlation vs Spectral Width

Sample Correlation: $e^{-\frac{1}{2}(\pi m \sigma)^2}$			
	σ		
m	0.040	0.100	0.200
1	0.992	0.952	0.821
2	0.969	0.821	0.454
3	0.931	0.641	0.169

the correlation values for the first three lag indices in the model [Equation (2)] corresponding

to three values of normalized spectrum width. For a normalized width of 0.2, there is significant decorrelation between samples separated by two or more places. (Improved estimation using higher-order lag-product statistics would seem unlikely beyond this point.) The normalized value 0.2 is used to represent wide spectrum width signals. Given the setting of the simulation, this signal corresponds to a weather signal with a spectrum width of 5 m/s. At the other end, the value 0.038 (corresponding to a 1 m/s spectrum width) is representative of narrow spectrum width signals.

Initially, performance comparisons are made for discrete SNR values in the range -10 to 10 dB. These (initial) results are the basis for a later shortening of the examined range to -4 to 10 dB.

3. ESTIMATOR EQUATIONS

3.1 ML and BAYES Formulations

The ML solution corresponding to Equation (3) is obtained, in principle, from joint maximization of the "likelihood" function

$$L(\Theta) = -\ln|\mathbf{R}| - \mathbf{Z}^{\dagger}\,\mathbf{R}^{-1}\,\mathbf{Z}, \tag{6}$$

but this has limited practical application because the resulting system of equations is neither explicit nor separable with respect to the components of Θ. The Bayes estimator that minimizes the mean-squared error $E[(\hat{\omega}-\omega)^2]$ is the mean of the conditional (posterior) density $p(\Theta|\mathbf{Z})$ (see, for example, Van Trees [3]) where the posterior density is derived from application of the Bayes formula:

$$p(\Theta|\mathbf{Z}) = \frac{p(\mathbf{Z}|\Theta)\,p(\Theta)}{\int p(\mathbf{Z}|\Theta)\,p(\Theta)\,d\Theta}. \tag{7}$$

This too has limited practical use as, notationally, "$d\Theta$" = "$d\omega\,d\sigma\,d\eta$"; the resulting estimator requires substantial numerical integration to cover the parameter space volume. Approximations of some form are clearly required in order to proceed further. The approach considered here first steps back and examines a simpler model for which practical (ML and Bayes) solution is possible and, later, examines adaptation of the resulting solutions to the model of Equation (3).

Examining the (somewhat) degenerate situation wherein one assumes the spectrum width (σ) and signal-to-noise (η) parameters known is, of itself, very useful. Under these conditions the aforementioned computational blocks are largely avoided, and one achieves optimal estimators that form a foundation for further analysis.

In the case of ML estimation, knowing σ and η reduces likelihood function, Equation (6), to the term $\mathbf{Z}^{\dagger}\mathbf{R}^{-1}\mathbf{Z}$ defining the ML solution

$$\hat{\omega}_{ML} = \underset{\omega\in[-\pi,\pi)}{\arg\min} \sum_{i=0}^{M-1}\sum_{k=0}^{M-1} z_i^{*}\,\gamma_{i,k}\,z_k\,e^{j\omega(i-k)}. \tag{8}$$

In the above, the coefficients $\gamma_{i,k}$ are the elements of the matrix Γ resulting from the convenient redefinition

$$\Gamma = \Gamma(\sigma,\eta) = [\mathcal{S}\mathbf{G} + \mathcal{N}\mathbf{I}]^{-1}. \tag{9}$$

Equation (8) is not explicit for ω; however, a change of variable and rearrangement results in an interesting frequency domain form:

$$\hat{\omega}_{ML} = \underset{\omega\in[-\pi,\pi)}{\arg\min} \sum_{m=-(M-1)}^{M-1} \hat{r}_m^{(\Gamma)}\, e^{-j\omega m}, \tag{10a}$$

more simply,

$$\hat{\omega}_{ML} = \underset{\omega\in[-\pi,\pi)}{\arg\min}\, \mathrm{Re}\left\{\sum_{m=0}^{M-1} \hat{r}_m^{(\Gamma)}\, e^{-j\omega m}\right\}, \tag{10b}$$

where $\hat{r}_m^{(\Gamma)}$ is the Γ-weighted autocorrelation estimate defined by

$$\hat{r}_m^{(\Gamma)} = \sum_{i=0}^{M-m-1} z_i^*\, \gamma_{i,i+m}\, z_{i+m} \tag{11}$$

From inspection of Equation (10) it is clear that a solution to this ML equation can be obtained by implementing a standard FFT transform of the weighted autocorrelation estimates $\hat{r}_m^{(\Gamma)}$. The transform samples Equation (10) with a resolution of $2\pi/N_{FFT}$ (N_{FFT} is the FFT size) and thus, by zero filling, computes $\hat{\omega}_{ML}$ to a desired resolution.

A similar computational reduction occurs for the Bayes formulation. If one assumes a uniform prior[4] for ω, the (one-dimensional) Bayes estimator can be written as

$$\hat{\omega}_{Bayes} = \frac{\int \omega\, e^{-\mathbf{Z}^\dagger \mathbf{D}\Gamma\mathbf{D}^*\mathbf{Z}}\, d\omega}{\int e^{-\mathbf{Z}^\dagger \mathbf{D}\Gamma\mathbf{D}^*\mathbf{Z}}\, d\omega}. \tag{12}$$

The above FFT computation, which yields the ML solution, also provides the exponent in Equation (12) and therefore a starting point for computing the posterior mean.

Comment 1. If, in Equation (11), $\gamma_{i,i+m} \equiv 1$, then $\hat{r}_m^{(\Gamma)}$ is equivalent to the (biased) sample autocorrelation estimate, and an FFT implementation of Equation (10) is equivalent to the (windowed) power spectral density estimate obtained using the Bartlett window. Hence, there is

[4]For this report only noninformative priors are considered, which, for ω, can be stated as $p(\omega) = 1/2\pi$, $-\pi \le \omega < \pi$.

also an equivalence between the FFT computation for Equation (10) and the periodogram spectral estimate

$$\mathcal{I}_M(\omega) \stackrel{\text{def}}{=} \frac{1}{M}\left|\sum_{m=0}^{M-1} z_m\, e^{-j\omega m}\right|^2. \tag{13}$$

The resulting ML estimator would therefore be equivalent to the Periodogram Maximization estimator considered by Mahapatra and Zrnić [4]. (The apparent contradiction in which a minimization of the power spectrum in Equation (10) defines the ML estimate is accounted for by the sign of the weights $\gamma_{i,i+m}$.)

Comment 2. If $\gamma_{i,i+m} \equiv \gamma_m$, then Equation (10) implements a classic smoothed periodogram estimate. However, Γ (the inverse of a Toeplitz matrix) is only asymptotically Toeplitz (though Γ is persymmetric: $\gamma_{i,j} = \gamma_{M-j,M-i}$). This nevertheless does suggest an alternative frequency domain solution and approximation based on implementation of a smoothed periodogram estimate.

3.1.1 Smoothed Periodogram Approximation

For this report, the weighted autocorrelation, Equation (11), is viewed as posing no computational difficulty (the results presented here assume such an implementation). However, a Toeplitz approximation to Γ provides a useful heuristic within which to view the relevance of the parameters σ and η. This, in turn, may suggest approaches to the more general case where σ and η are unknown.

In treating Γ as Toeplitz, one can compute smoothing weights by taking averages along the diagonals of Γ:

$$\overline{\gamma}_m = \frac{-1}{M-m-1} \sum_{i=0}^{M-m-1} \gamma_{i,i+m}$$

(The minus sign is introduced here so that the ML solution of Equation (10) can be associated with finding the maximum of the power spectral density estimate.) This maximum likelihood window function, indexed by η and σ, can be used to define an approximate ML solution

$$\hat{\omega}_{AML} = \arg\max_{\omega\in[-\pi,\pi)} \sum_{m=-(M-1)}^{M-1} \overline{\gamma}_m\, \hat{r}_m\, e^{-j\omega m}, \tag{14}$$

where $\hat{r}_m$ represents the normal (unweighted) autocorrelation estimate.

The relationship between $\overline{\gamma}_m$ and the parameters σ and η is documented by Figure 1 where frequency-domain smoothing windows (Fourier transforms of the coefficients $\overline{\gamma}_m$; interpolated for

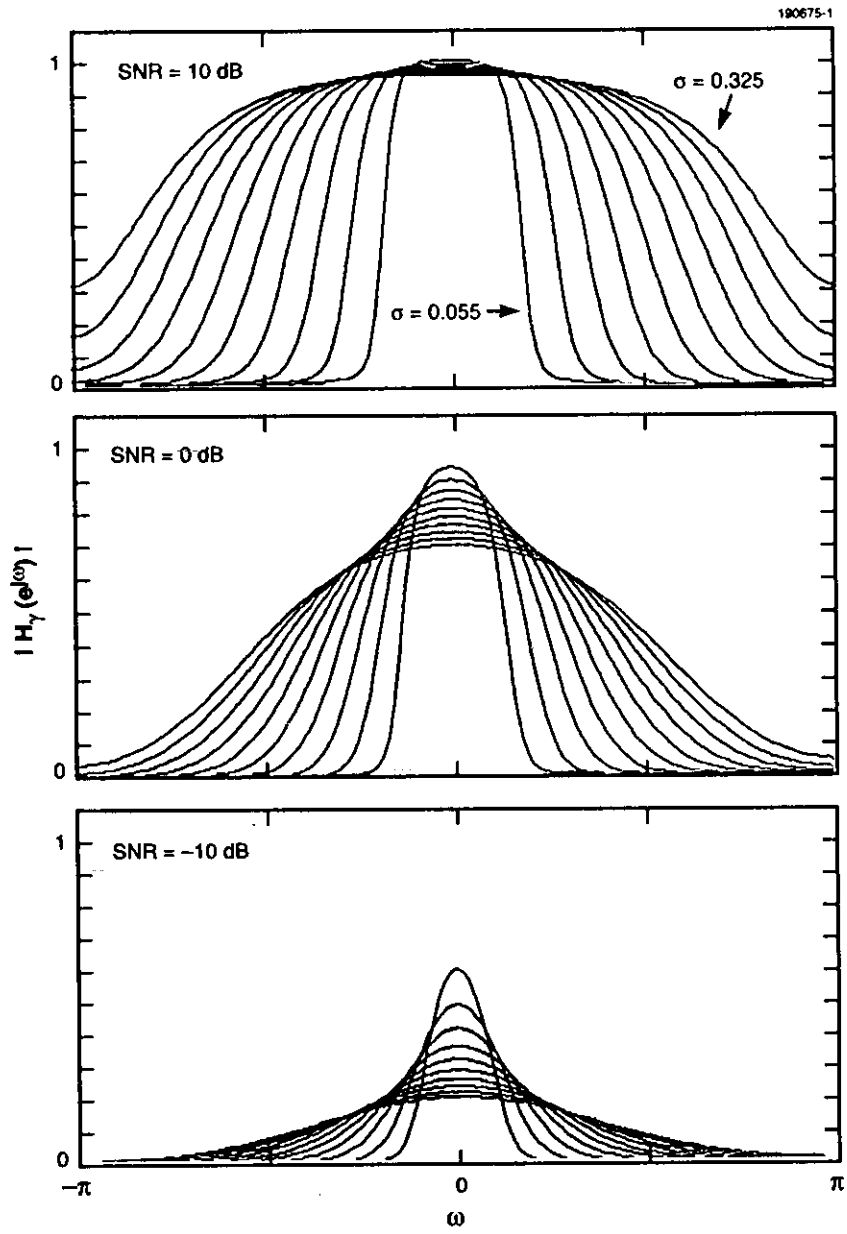

Figure 1. ML matched filter windows: σ and η dependence. Frequency domain smoothing windows obtained by Fourier transforming the weighting coefficients $\overline{\gamma}_m$ are plotted for three values of weather SNR and a range of weather spectrum width values.

clarity) are presented for three SNR values: -10, 0, and 10 dB, and a range of normalized spectrum widths. At 10 dB SNR and for low to moderate spectrum widths the smoothing windows appear to be, practically speaking, square-window averagers. This suggests the generalized view that the spectrum width parameter σ primarily affects window width. As SNR decreases, the dominant feature change is a scaling of the windows to smaller magnitudes (although shape rounding and bandwidth broadening, as measured by 3 dB width, are clearly evident). It is helpful to add to the generalization by stating that the windows are scaled according to the SNR parameter η. Predictions that result from these simplifications are as follows. Clearly [from Equation (10)], scaling the window does not affect ML estimation, and hence the value of the SNR parameter η would not be expected to have a strong effect on the performance of an ML implementation. However, in the case of the Bayes implementation, scaling coupled with exponentiation implicitly introduces a form of signal isolation. That is, by its nonlinear nature, the exponentiation in Equation (12) enhances (isolates) large spectral lines in the smoothed periodogram, and scaling of the window (by η) modulates this isolation by adjusting the power of exponentiation. Hence, one would anticipate that, at least, the bias of the Bayes algorithm would be affected by the η value used.[5]

3.2 Time-Domain Processing: Pulse Pair (PP)

The PP estimator is standard in weather radar applications, and its formulation can be obtained from many sources. Originally proposed in the context of the independent sampling of PP measurements (for example, see Miller et al. [5], time-domain formulas for PP estimators have been routinely applied to vector measurements such as those represented by the sample **Z**. For the purpose of this report, the PP estimator is defined by the equation

$$\hat{\omega}_{PP} = \arg \sum_{i=0}^{M-2} z_i^* \, z_{i+1}. \tag{15}$$

When $M = 2$, this also defines the ML estimate, as can be seen by comparison with Equation (8). Clearly, as σ increases and the correlation between samples decreases one would also predict $\hat{\omega}_{PP}$ to approach the performance of $\hat{\omega}_{ML}$.

3.3 Frequency-Domain Processing: Wind Profiling (WP)

Among the various published frequency-domain methods is one that has been developed for use in WP networks—a task with inherently low SNR values (see May and Strauch [6]). The WP

[5]The presence of a uniform noise floor in the spectral density estimate induces a bias toward the value zero when the mean of the density is computed.

estimator of May and Strauch [6] is periodogram derived and has led to general claims of improved Doppler mean estimation in the case of low SNR; hence, it is of interest for the present report.

The estimator, WP, was derived with one major departure from the description given in May and Strauch [6] (their First Moment or FM algorithm). Computational aspects of this algorithm are presented below by way of a comparison with the proposed (one-dimensional) ML and Bayes algorithms. Note, however, that there is one important difference between the WP algorithm defined here and the algorithm described in May and Strauch [6]. The WP network makes generous use of *periodogram averaging* as a means of stabilizing the periodogram estimates (Bartlett's procedure; see the next section). Because the emphasis of this report is on estimation using a fixed small-sample size, no such averaging is possible. (This report does not consider the possibility of data averaging across range gates or neighboring radials.)

Estimate Stabilization. At the heart of the WP algorithm is an implementation of Bartlett's procedure for estimating the power spectral density. That is, the data segment for analysis (length M data segment corresponding to a single range cell observation) is divided into q equal length subsegments, each of which is used to obtain an M/q-length periodogram estimate. The q power spectral estimates are averaged to improve the stability of the power spectral estimate. The Bayes implementation [Equation (12)] also can be viewed as having a power spectral estimator—a smoothed periodogram—at the heart of its procedure. Because Bartlett's method per se is not appropriate for the small single-sample case (the primary interest here) it may be argued that smoothing windows therefore are necessary to stabilize the estimates. (It should be remarked that, generally speaking, the goals of mean Doppler velocity estimation and power spectral density estimation are not one and the same. Hence, general arguments for improving power spectral estimation do not necessarily carry over to improved velocity estimation.)

Signal Isolation. After obtaining an estimate of the power spectral density, the WP method proceeds with an ad hoc attempt to isolate signal from noise. A noise floor for the spectral estimate is determined, and censoring (zeroing) of all spectral coefficients beyond the first crossing of the noise floor, when proceeding from that frequency index with maximum power, is performed. The noise power level is also subtracted from the remaining interval of nonzero values, and a mean frequency value is computed by constructing a density from the remaining spectral coefficients and computing its mean. As previously mentioned for the Bayes implementation, the combination of window scaling and exponentiation can be given the heuristic interpretation of signal-from-noise isolation.

4. PERFORMANCE WITH KNOWN σ AND η

This section presents a baseline analysis corresponding to a one-dimensional parameter space (i.e., σ and η assumed known). These Monte Carlo results provide optimal performance measures for each method. In the case of the Bayes implementation, results for known σ and η also provide a greatest lower bound for the standard error of estimating mean Doppler velocity.[6] In later sections, the Bayes curves determined here are employed with the label "Bayes Bound" for velocity estimation.

4.1 Zero Mean Velocity

Figure 2 presents the simplest comparison: estimation of a zero-mean Doppler weather target (eliminating, for the moment, the contribution of bias[7] in standard error comparisons). The figure plots standard error vs input SNR for two cases: a narrow input spectrum width ($\sigma = 0.038\, v_{Nyq}$) and a wide input spectrum width ($\sigma = 0.192\, v_{Nyq}$). The corresponding CR lower bounds are included in each panel.

4.1.1 Narrow Spectrum Widths

ML, PP, and WP. For narrow spectrum widths ML, PP, and WP estimates all exhibit similar functional relationships with respect to SNR. A uniform ranking of these estimators (across SNR) cannot be deduced from the data: ML is clearly best at high SNR (7 to 10 dB) values but the WP method appears relatively better at low SNR values (less than 0 dB).

CR Bound. For narrow spectrum widths, the CR bound is well below the error curve of any of the above three estimators—lower by nearly a factor of two at the high SNR end. This discrepancy between CR bound and observed performance, which is even more substantial for low SNR values, could erroneously lead to speculation that much improved performance is possible.

Bayes Bound. The curve for the Bayes standard error shows the CR bound to be overly optimistic. However, the Bayes estimator clearly exhibits a performance gain for SNR values in the range 0 to 10 dB (at 10 dB, the Bayes standard error is respectively 0.72, 0.66, and 0.52%

---

[6]Comparisons involving the ML, Bayes, and WP performance statistics must concede an error (bias) that results from the finite length FFT implementation. For these frequency domain estimators, the sampling resolution of $2/N_{FFT}\ v_{Nyq}$ is exhibited as a bias in the range $\pm 1/N_{FFT}\ v_{Nyq}$ (and therefore maps an interval about the error values reported here). All frequency domain computations in this report were computed using an FFT length of 64.

[7]For this zero mean Doppler signal, each of the three frequency domain methods was found to exhibit an estimated bias below the resolution defined by the 64 point FFT implementation (i.e., $< 1/32$), regardless of the η value.

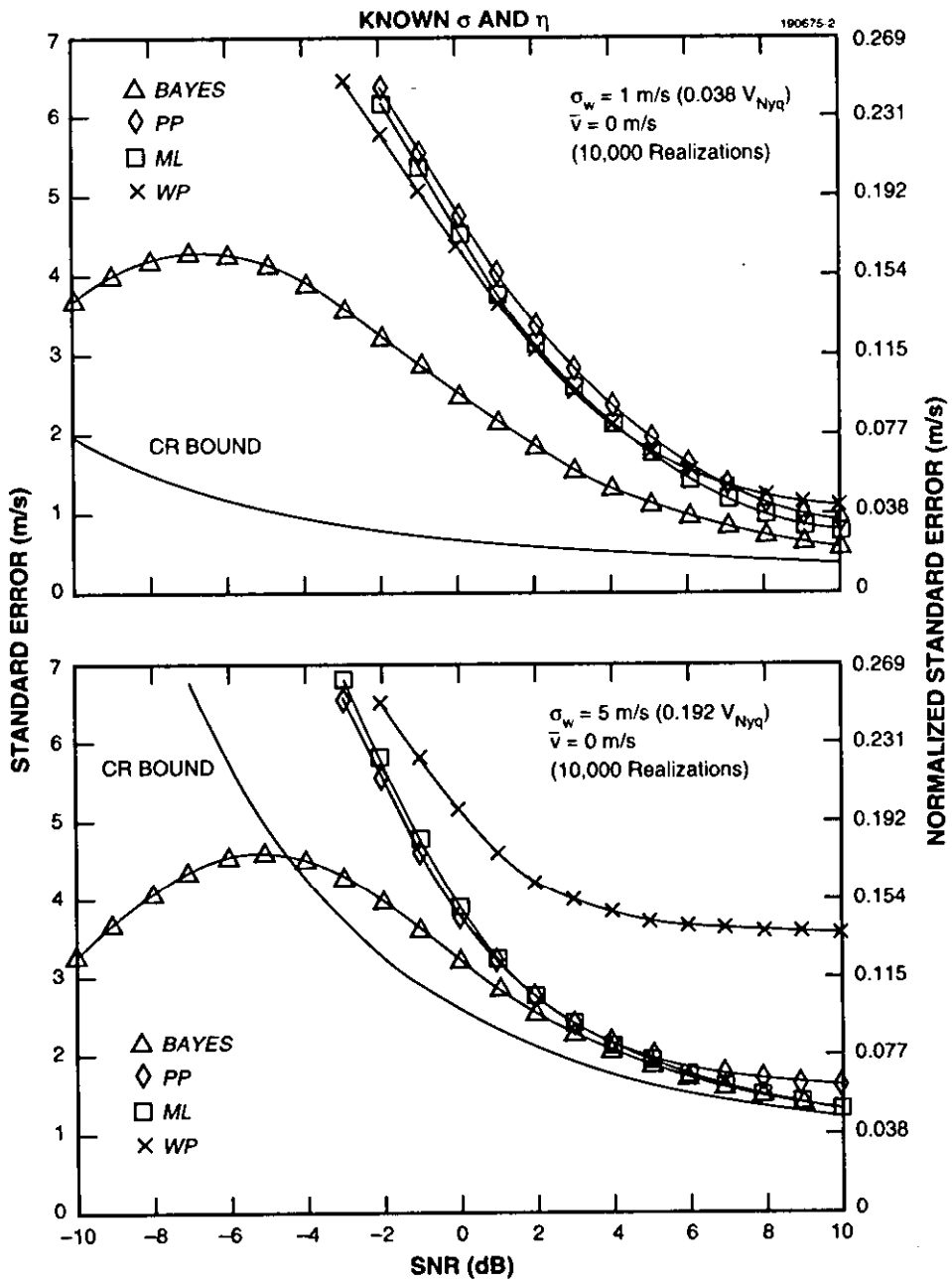

Figure 2. Optimal velocity estimation standard error for zero-mean Doppler weather. Both panels plot estimated standard error for four methods of Doppler velocity estimation: ML, Bayes, WP derived, and PP. The upper panel corresponds to weather having a narrow spectrum width and the lower corresponds to weather having a wide spectrum width.

that of the ML, PP, and WP estimators; at 0 dB, the Bayes standard error is respectively 0.58, 0.60, and 0.54% that of the others). For SNR values less than 0 dB, adequate separation of signal from noise becomes more difficult and the Bayes standard error *drops* as its estimate becomes increasingly biased toward zero. Comparisons among the estimators for SNR values below 0 dB should therefore include bias error, and this is included in the sections to follow.

4.1.2 Wide Spectrum Widths

For wide spectrum widths also, the general observations of the previous section apply, but of particular note for these input signals is the following. ML and PP performances are predictably close, although ML is better at higher SNR values and appears to approach the CR bound (as SNR increases), whereas the PP curve appears to plateau at a level distinctly above the CR bound. From inspection of Table 1, one may surmise that at a normalized spectrum width of 0.2, the decorrelation between samples is such that very little improvement can be realized from using higher lag terms in the estimation process. In confirmation, there is less difference between performance of PP and ML and the CR and Bayes bounds; however, note that WP stands alone as a clearly suboptimal estimate. This exceptionally marked degradation in WP performance persists to high SNR values and confirms a wide-spectrum-width weakness identified by other investigators. The ML estimator appears to achieve the lower Bayes bound for SNR values in the 5 to 10 dB range, and both ML and Bayes improve upon PP performance over this range (at 10 dB, the Bayes standard error is respectively 0.98, 0.81, and 0.37% that of the ML, PP, and WP estimators). For SNR values 0 to 5 dB, ML and PP performances are essentially equivalent, and both depart appreciably from the Bayes bound as SNR approaches 0 dB (at 0 dB, the Bayes standard error is respectively 0.83, 0.83, and 0.64% that of ML, PP, and WP).

As a secondary note, one should observe that there is no conflict in the fact that the CR bound and Bayes error curves cross (there is an implied crossing of these two curves for the narrow spectrum width case as well). This only serves as a reminder that the CR bound applies to the performance of unbiased estimators, which the Bayes estimator, generally speaking, is not.

4.2 Nonzero Mean Velocity

Standard error and bias results for nonzero mean velocities ($\overline{v} = 5$ m/s and $\overline{v} = 13$ m/s) and narrow and wide spectrum width cases, as above, are presented in Figure 3. Standard error results are summarized in the upper half of each panel, and bias results are summarized in the lower.

4.2.1 Standard Error

For SNR values greater than 0 dB, there is general agreement (within the resolution of the Monte Carlo parameterization) with the results of Figure 2. Below 0 dB, there is a notable departure (most evident for the Bayes results) due to inclusion of bias error.

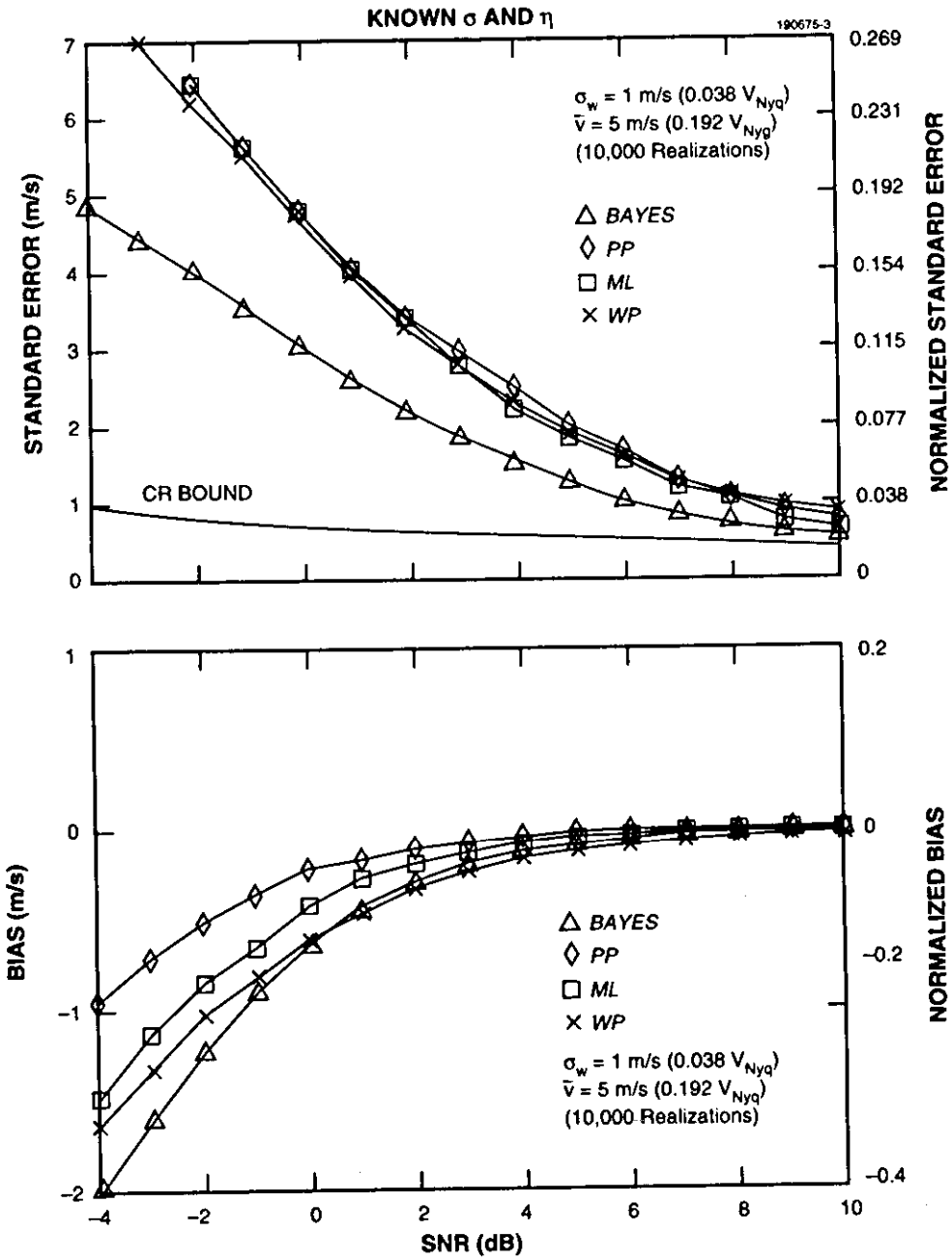

Figure 8. Optimal performance: standard error and bias results for nonzero velocities. (a) $\bar{v} = 5$ *m/s,* $\sigma_w = 0.038\, v_{Nyq}$.

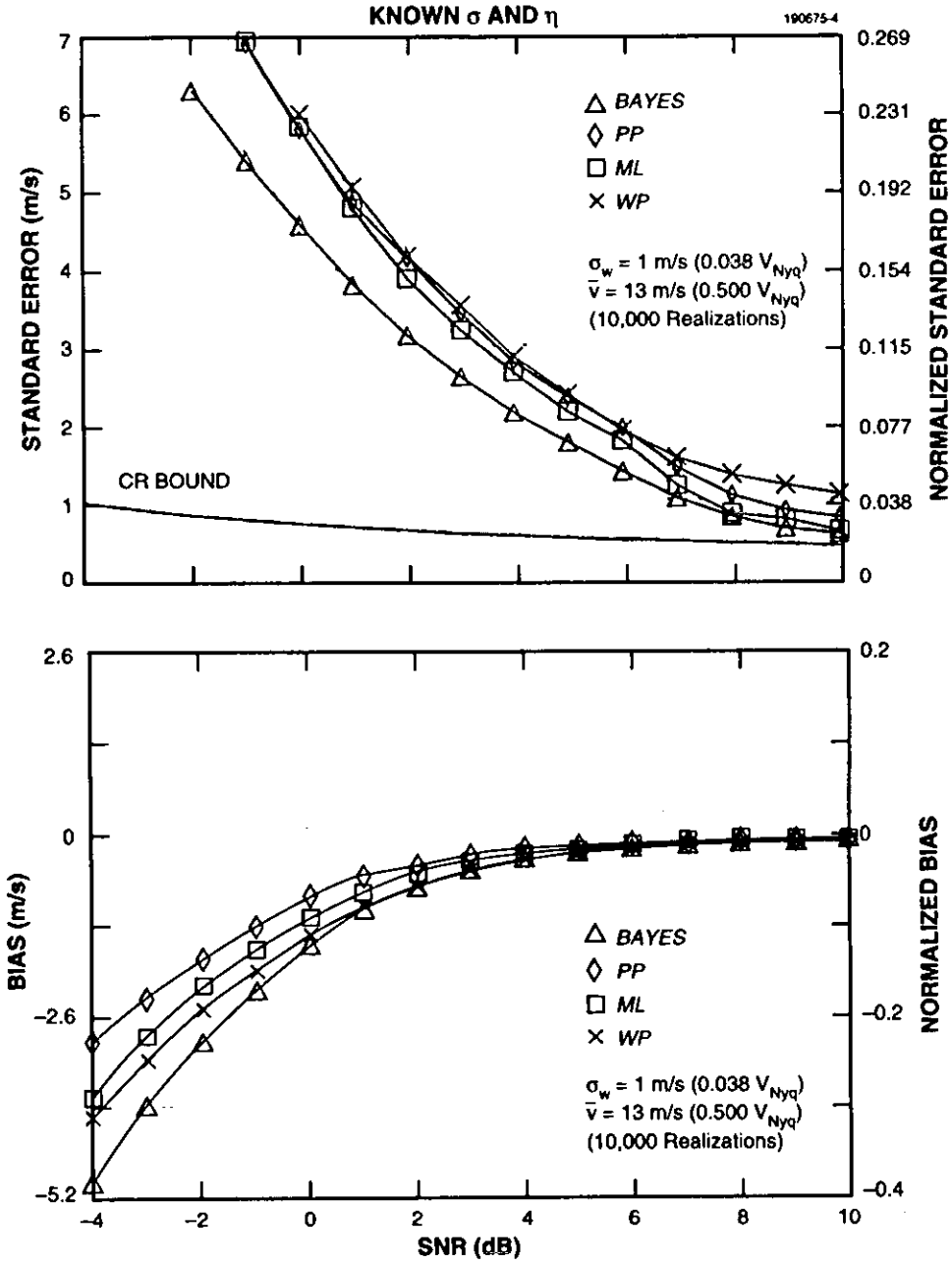

Figure 8. Optimal performance: standard error and bias results for nonzero velocities. (b) $\bar{v} = 13$ *m/s,* $\sigma_w = 0.038\, v_{Nyq}$.

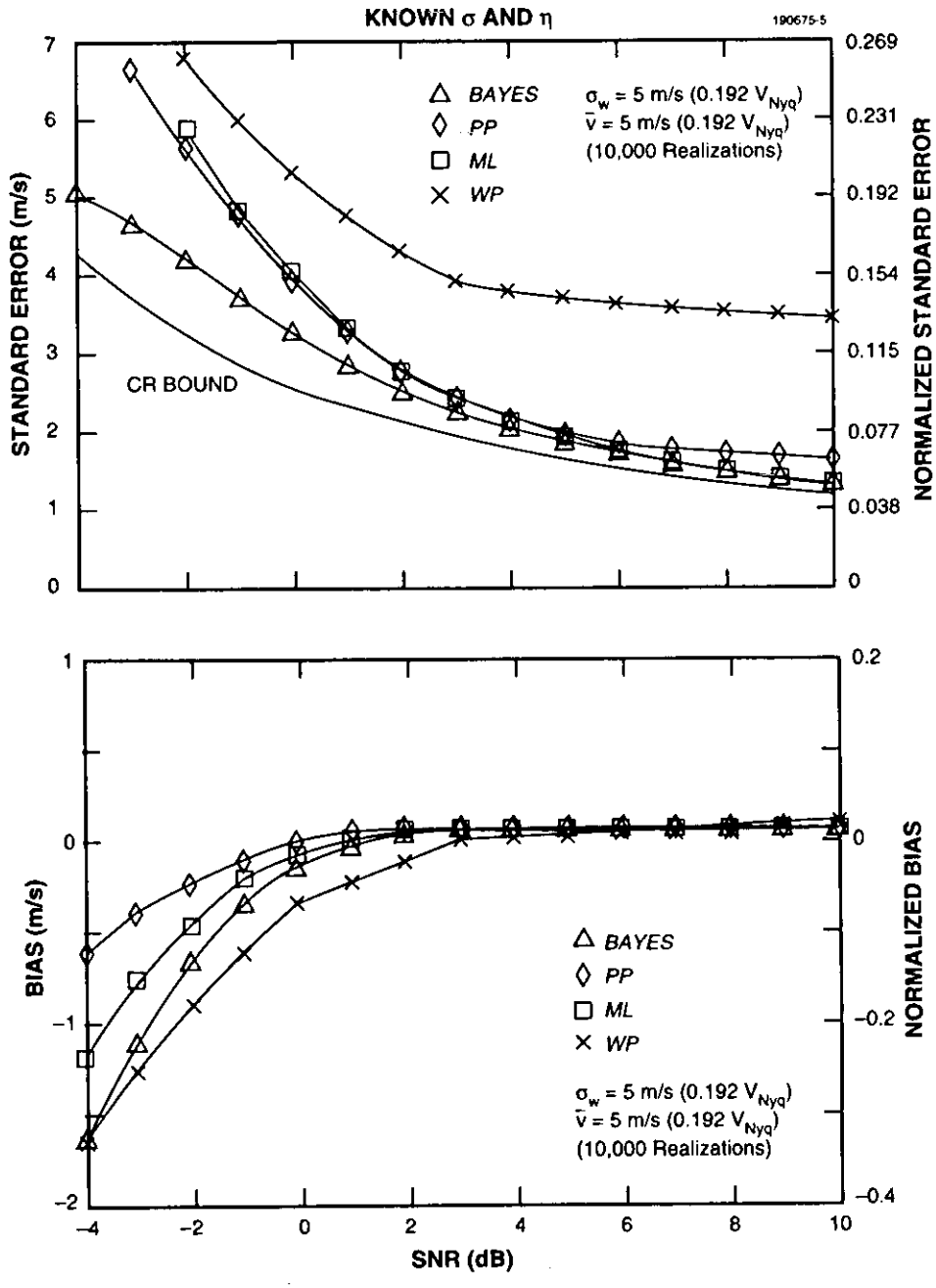

Figure 8. Optimal performance: standard error and bias results for nonzero velocities. (c) $\bar{v} = 5$ *m/s,* $\sigma_w = 0.192\ v_{Nyq}$.

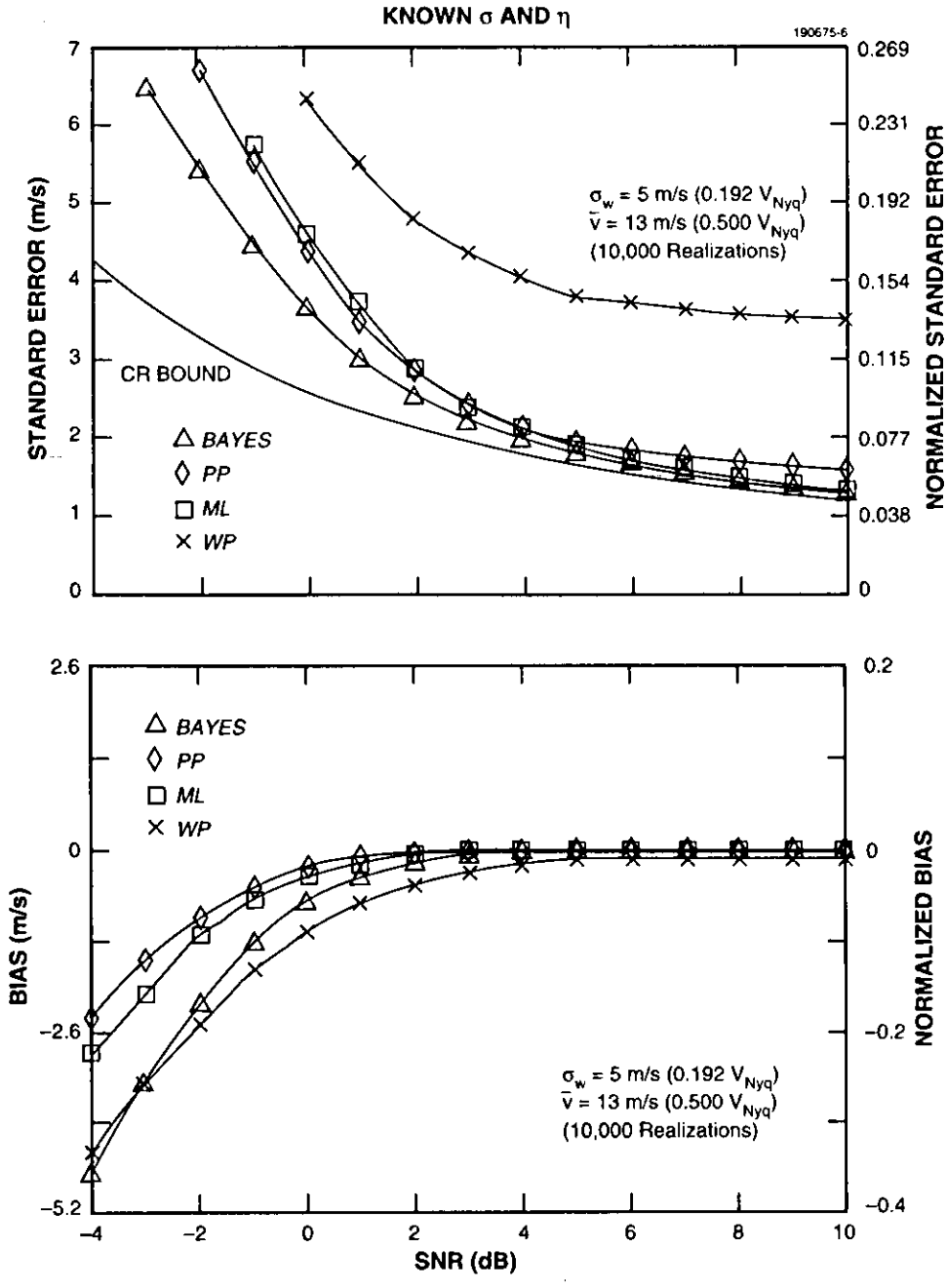

Figure 9. Optimal performance: standard error and bias results for nonzero velocities. (d) $\bar{v} = 13$ m/s, $\sigma_w = 0.192\ v_{Nyq}$.

4.2.2 Bias

For both narrow and wide spectrum widths there is a breakdown of all four methods as SNR decreases below 0 dB. For SNR values greater than 0 dB, the bias of each estimator appears to be within acceptable limits and generally near zero. A large proportion of the improved (standard error) performance of the Bayes method, for SNR < 0, is at the cost of increased bias (toward zero). Generally, the PP estimator appears to have the best (i.e., smallest) bias performance. Departure from zero-bias performance in the narrow spectrum width case occurs earlier (i.e., at higher SNR values) in comparison to the wide spectrum width case.

PP and ML results do not appear to change appreciably when the weather velocity is changed from 5 to 13 m/s. Although the theoretical CR bound does not depend on weather velocity, this is not true for the bound provided by the Bayes estimate. Performance for the Bayes and WP estimators, which both compute a spectral mean, deteriorate when weather velocity is increased to 13 m/s—a performance loss due to spectral folding. (Interestingly, moving weather velocity from 5 to 13 m/s has its most significant *bias* effect in increasing that of the PP estimator.)

4.3 Summary

4.3.1 Narrow Spectrum Widths

For narrow spectrum widths ML, PP, and WP methods all have similar performance characteristics (over the range of SNR values examined). Nevertheless, it can be argued that ML provides better performance at higher SNR values. Clearly, the indication is that higher SNR values are required to bring out a decisive advantage here from the ML implementation; a more extensive Monte Carlo analysis (including higher SNR values) would be needed to measure the extent of these improvements. As SNR decreases away from 0 dB, all methods begin to fail although the bias performance of PP is uniformly best. The CR bound is much lower than the performance of all, but the Bayes results show the information bound to be overly optimistic for this small-sample ($M = 20$) case. The asymptotic optimality of ML estimation was also demonstrated to be of no consequence for this small-sample case. The Bayes estimator demonstrates a markedly improved performance, but for SNR values below zero, this is at least in part, at the expense of increased bias. Bias does not appear to contribute appreciably to estimation error for SNR values above 0 dB. All four methods, however, do exhibit notable bias for SNR values in the range -5 to 0 dB. Bias comparisons appear to always favor PP estimation.

4.3.2 Wide Spectrum Widths

At wider spectrum widths ML and PP are for the most part similar, but the ML results show a slight improvement at higher SNR values (greater than 5 dB). Although Bayes performance represents the optimum, ML and PP are close to its bound (compare to the narrow spectrum case); all three estimators perform near the CR bound (again, in comparison to the narrow spectrum

case). However, the WP method clearly has an undesirable performance at wide input spectrum widths. An explanation for this is offered in Section 5.2.

5. PERFORMANCE WITH ARBITRARY σ AND η

The previous section indicated that a small performance improvement (as measured by standard error) might be obtained using an ML formulation (relative to PP and at high SNR values only), and that a much improved (and optimal) performance might result from a Bayes implementation. The task at hand, now, is to preserve these performance gains while addressing the issue that σ and η are never known exactly.

In the remainder, the Bayes and ML algorithms will process data using approximate values for the parameters σ and η and, in that sense, represent suboptimal algorithms. (However, for convenience, the labels Bayes and ML will still be used.) Before evaluating practical approaches to estimating σ and η it is useful to examine the effect of arbitrary σ and η values on Bayes and ML performance. In other words, for velocity estimation, first consider whether it is important that either σ or η be known at all. This examination is made by keeping one parameter fixed at its known value while varying the other among appropriate candidate values.

5.1 Sensitivity to Incorrect η

Figures 4 and 5 continue the narrow/wide spectrum width analysis of before by examining performance when data are processed assuming either one of two fixed η values: 0 or 10 dB. These choices represent logical test cases in the sense of asking whether reasonable performance can be obtained by categorically treating the data as either low or high SNR data. Figure 4 considers ML estimation for the narrow and wide spectrum width case; Figure 5 repeats the analysis for the Bayes estimator (comparisons for Doppler weather targets of 5 m/s (0.192 v_{Nyq}) and 13 m/s (0.500 v_{Nyq}) are presented). In this, and all following figures, the Bayes performance curve for the case of known σ and η (Section 4) is repeated as the "Bayes Bound." The results for PP estimation are also reproduced for continued comparison.

5.1.1 ML Algorithm

In the case of narrow spectrum width weather [Figures 4(a) and 4(b)], the parameter η as predicted (Section 3.1.1) has no apparent effect on ML performance. This is not quite the case, however, for wide spectrum width weather [Figures 4(c) and 4(d)] where mismatch between assumed and actual SNR results in increased estimation error. As will be seen also in the case of the Bayes algorithm, using a 10 dB value for the SNR parameter results in a performance loss (relative to PP) for input SNR values below 6 dB. Hence, the view of a smoothing window shape independent of η is, in places, an oversimplification.

5.1.2 Bayes Algorithm

From Figures 5(a-d), it is clear that the Bayes implementation, using an arbitrary fixed η, has a substantially altered performance. In neither case (0 or 10 dB window) did performance match

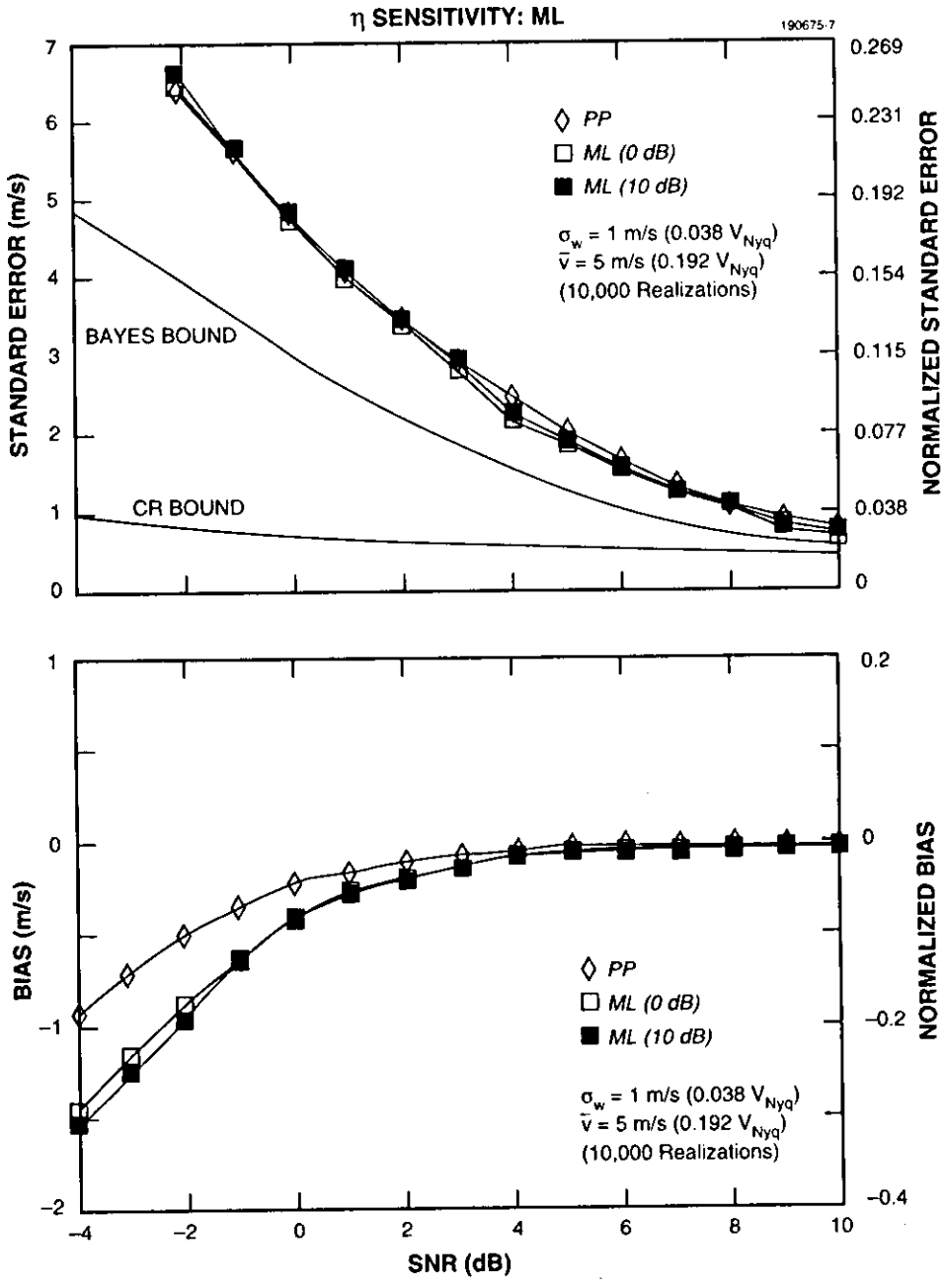

Figure 4. ML algorithm performance sensitivity: effect of signal-to-noise parameter η. (a) $\bar{v} = 5$ m/s, $\sigma_w = 0.038\, v_{Nyq}$.

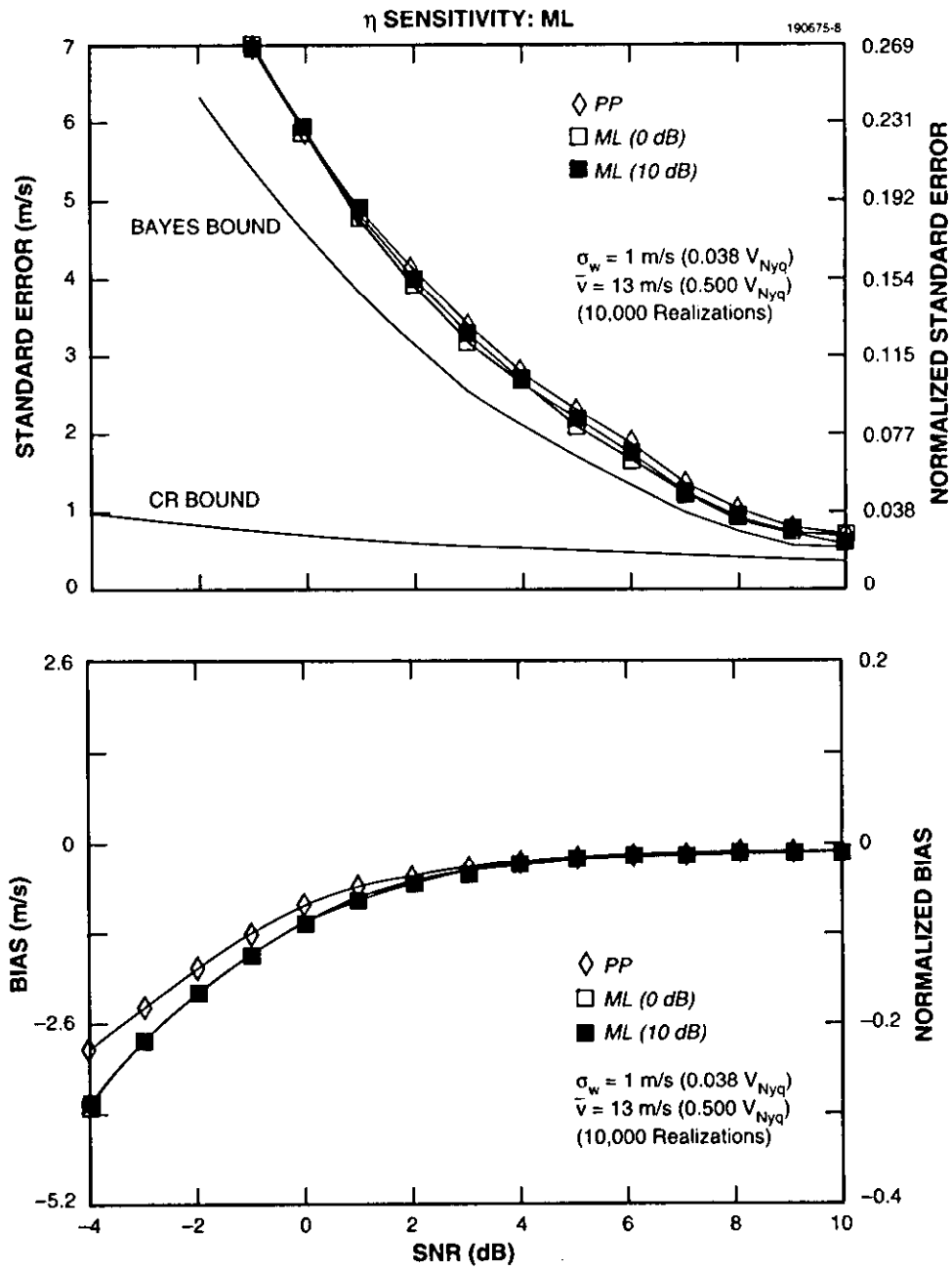

Figure 4. ML algorithm performance sensitivity: effect of signal-to-noise parameter η. (b) $\bar{v} = 13$ m/s, $\sigma_w = 0.038\ v_{Nyq}$.

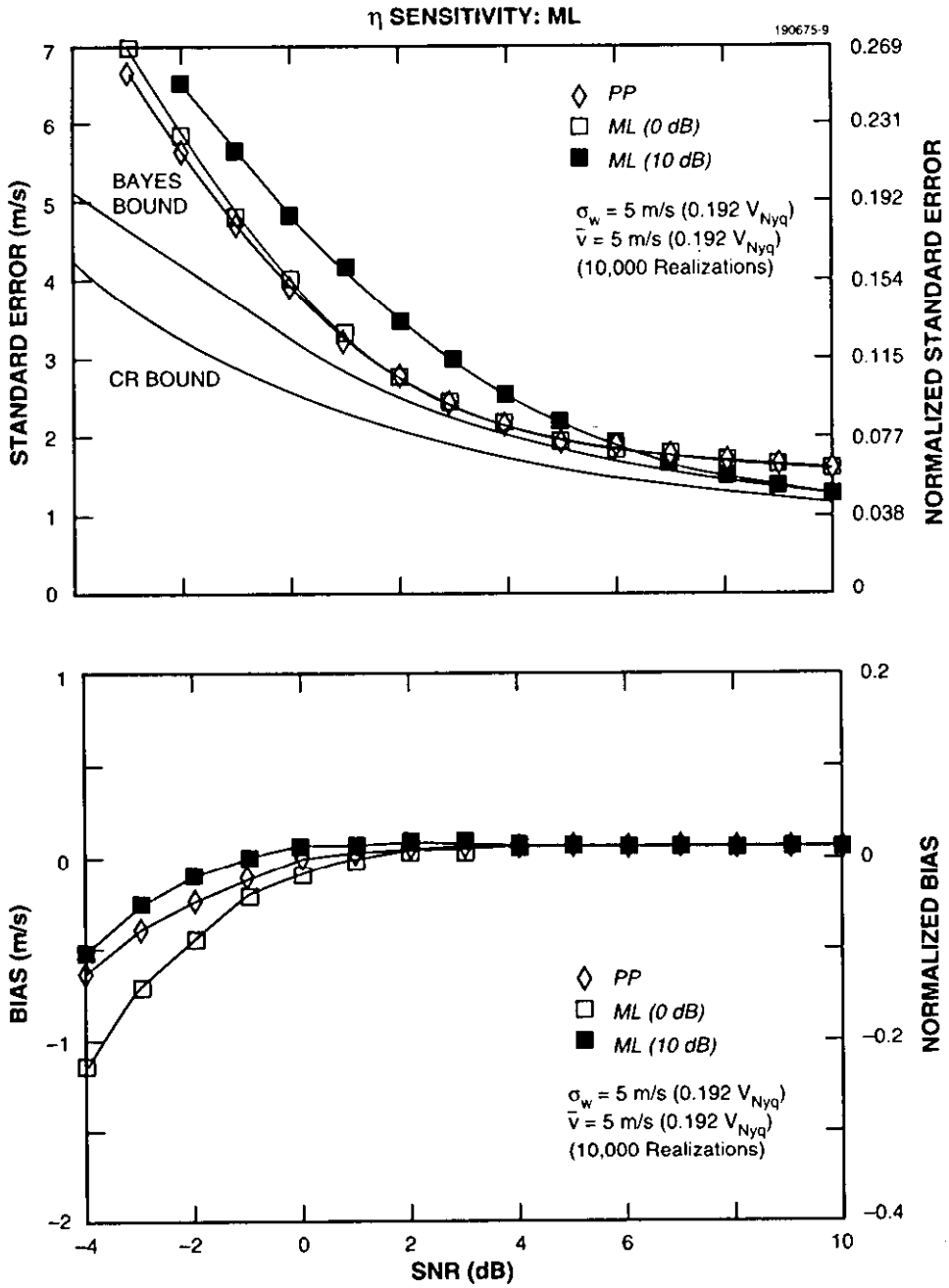

Figure 4. ML algorithm performance sensitivity: effect of signal-to-noise parameter η. (c) $\bar{v} = 5$ m/s, $\sigma_w = 0.192\, v_{Nyq}$.

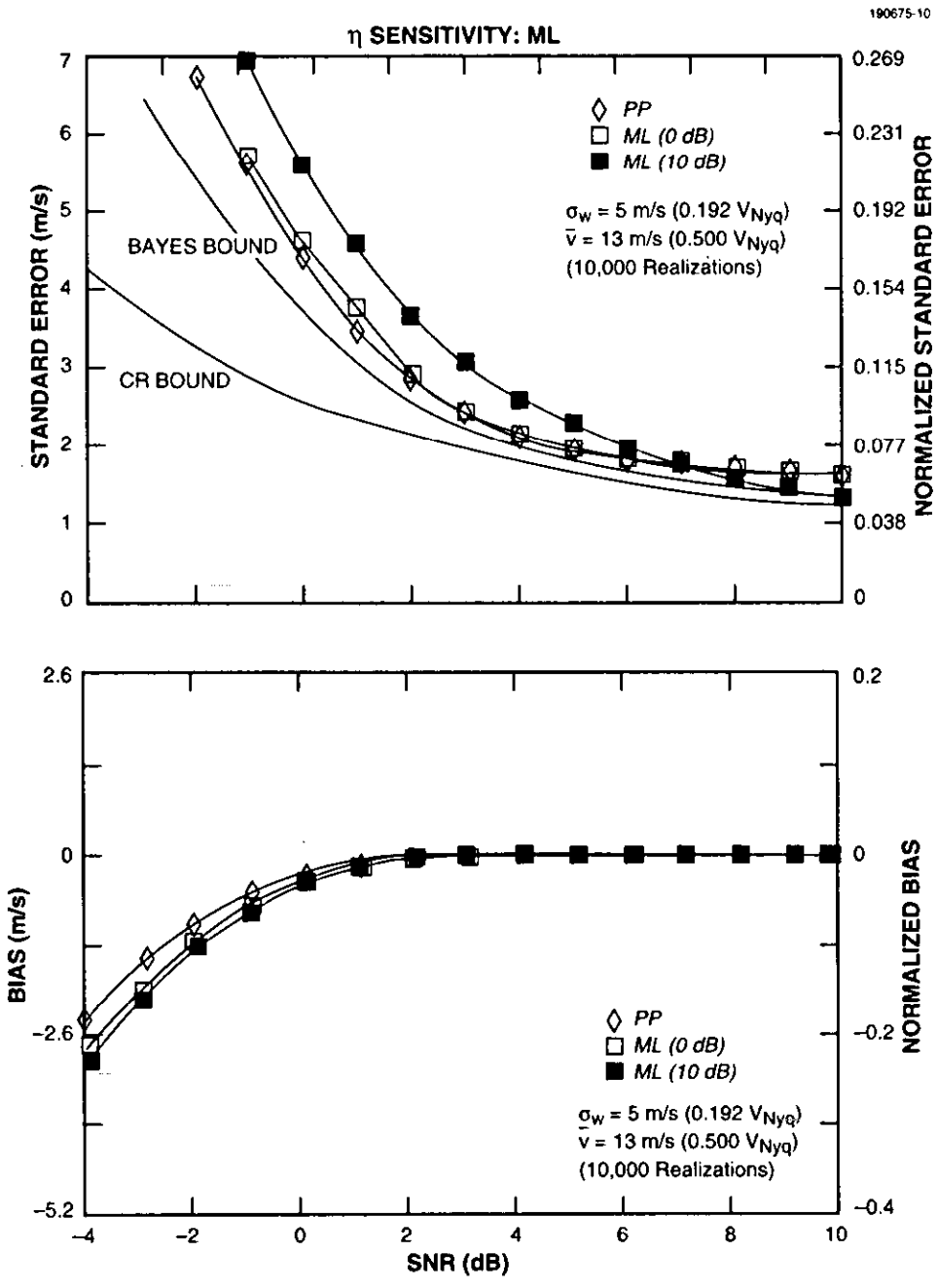

Figure 4. ML algorithm performance sensitivity: effect of signal-to-noise parameter η. (d) $\bar{v} = 13$ m/s, $\sigma_w = 0.192\ v_{Nyq}$.

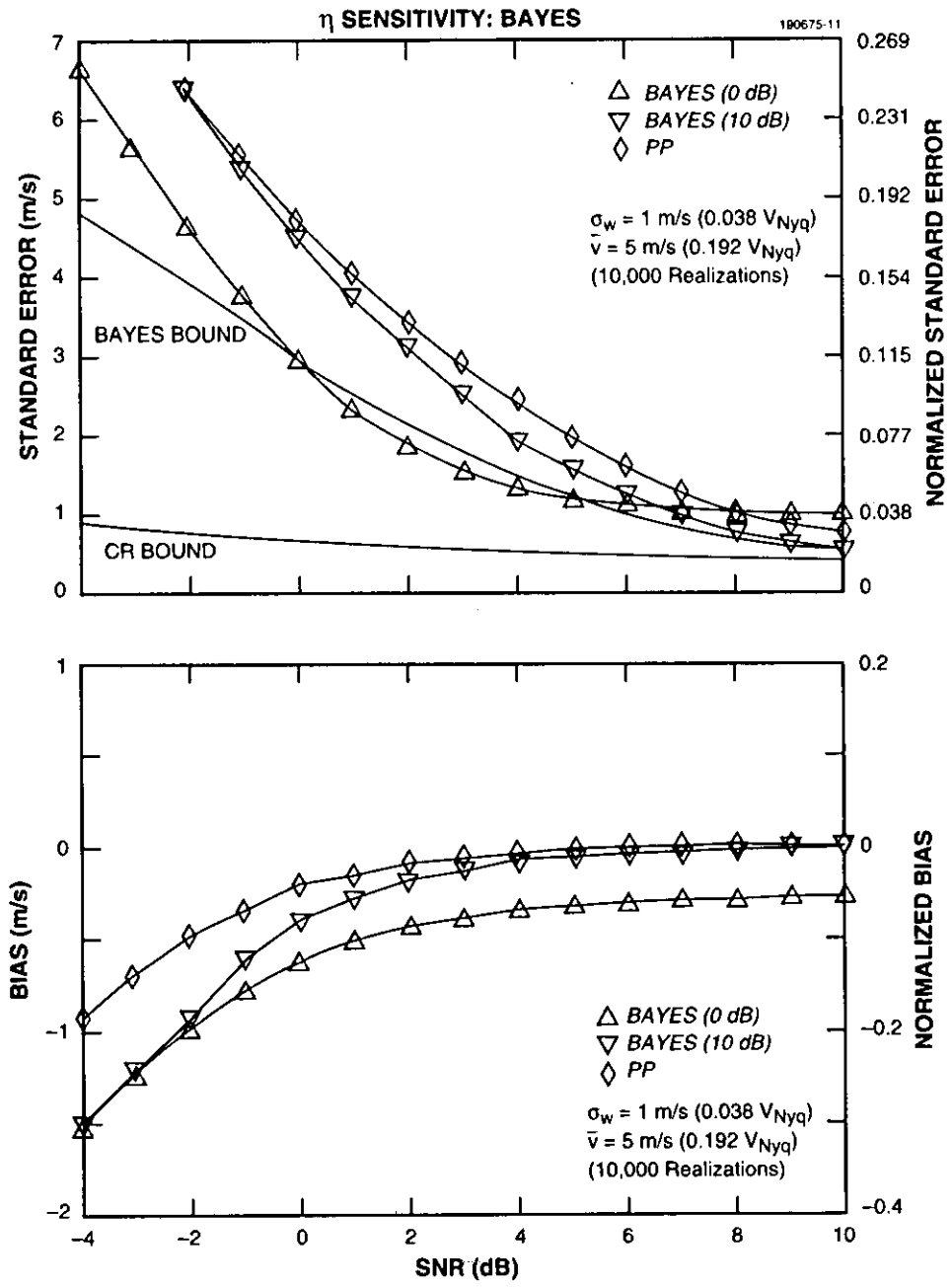

Figure 5. Bayes algorithm performance sensitivity: effect of signal-to-noise parameter η. *(a)* $\bar{v} = 5$ *m/s,* $\sigma_w = 0.038\ v_{Nyq}$.

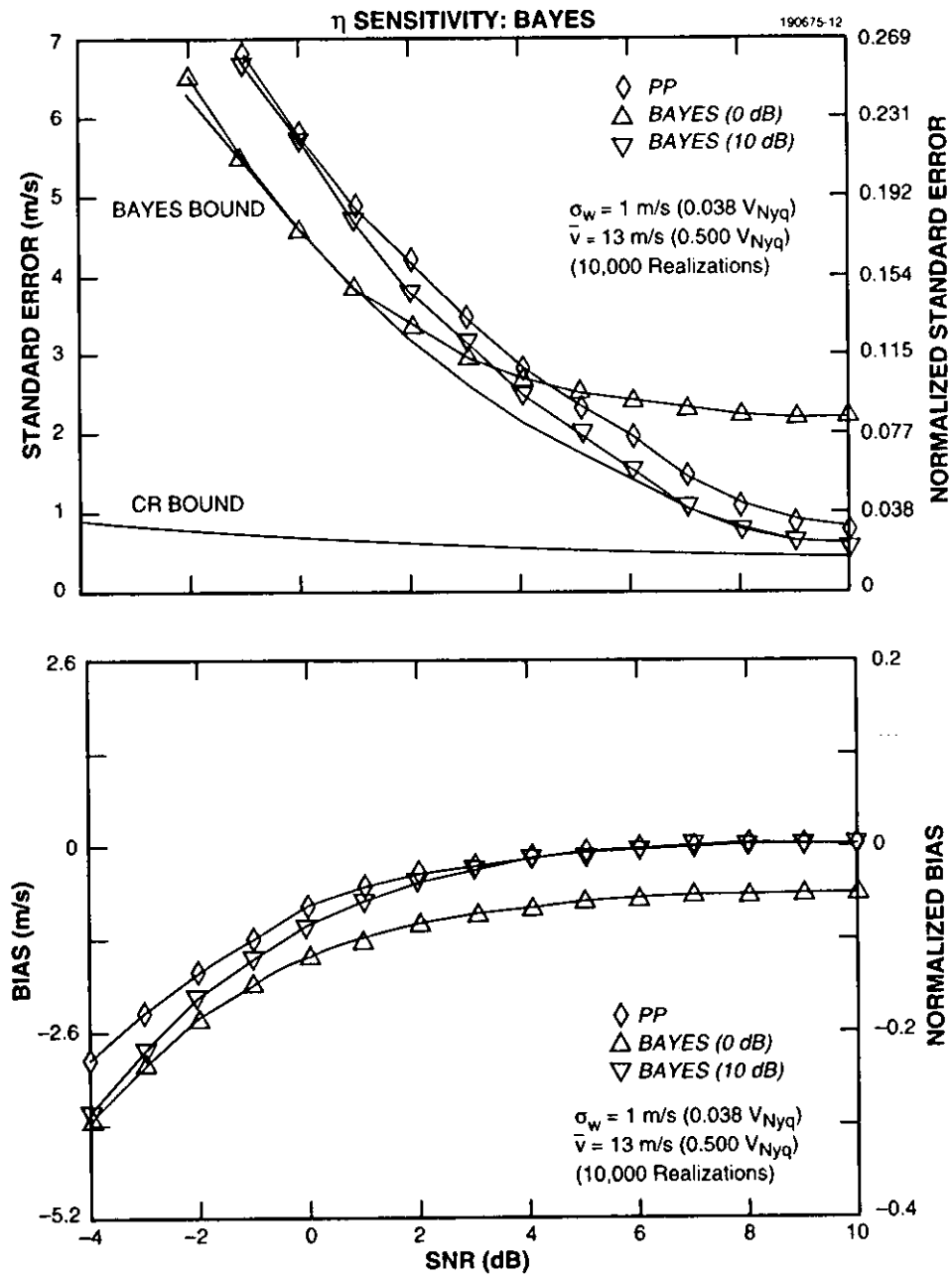

Figure 5. Bayes algorithm performance sensitivity: effect of signal-to-noise parameter η. *(b)* $\bar{v} = 13$ *m/s,* $\sigma_w = 0.038\ v_{Nyq}$.

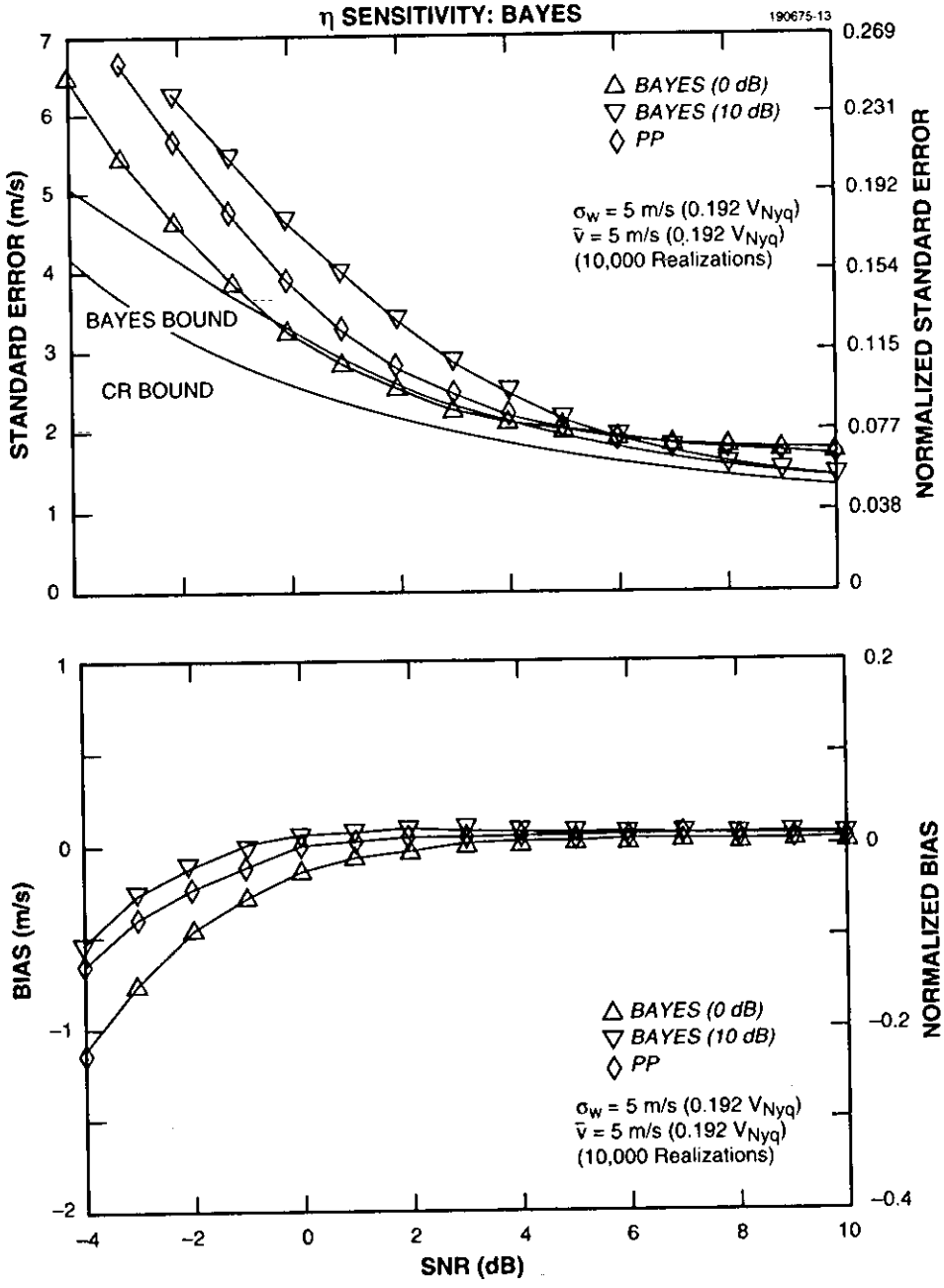

Figure 5. Bayes algorithm performance sensitivity: effect of signal-to-noise parameter η. *(c)* $\bar{v} = 5$ *m/s,* $\sigma_w = 0.192\ v_{Nyq}$.

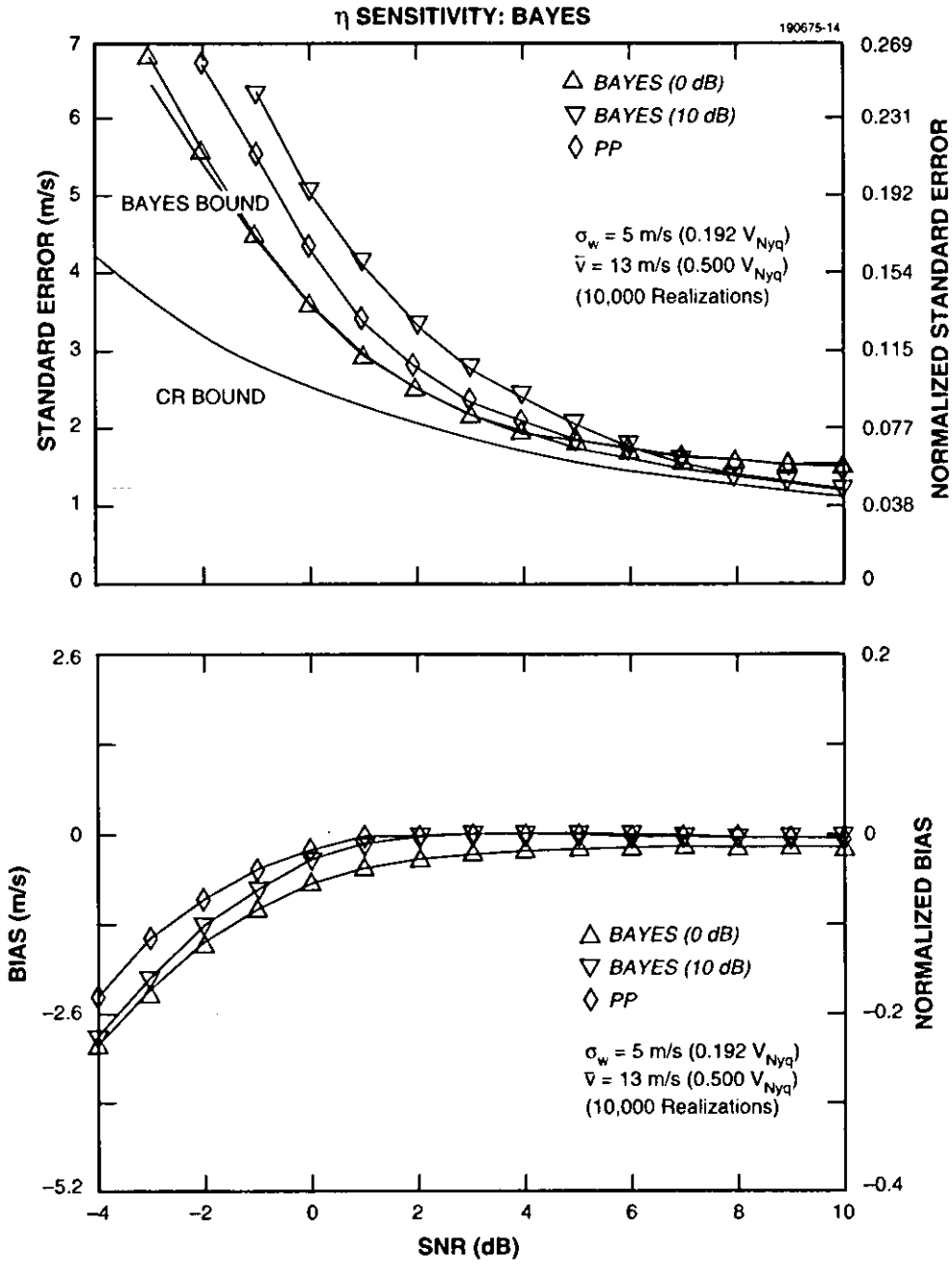

Figure 5. Bayes algorithm performance sensitivity: effect of signal-to-noise parameter η. *(d)* $\bar{v} = 13$ *m/s,* $\sigma_w = 0.192\ v_{Nyq}$.

the optimum Bayes Bound for all input SNR values; nor did it, at least, uniformly improve upon PP performance.

For the narrow spectrum width case of Figure 5(a), there is an (apparently) anomalous crossing of the Bayes Bound curve by the performance curve assuming $\eta = 0$ dB. This is not a contradiction because the Bayes Bound optimality applies only in the sense of average performance against the totality of all possible weather target velocities. Hence, it is possible to obtain lower standard errors for this *particular* weather velocity of 5 m/s. (Clearly, the estimator $\hat{v} \equiv 5$ m/s, an extremely degenerate case, has zero standard error and bias at this test point.) With the 0 dB window, note the severe compromise in estimator bias. This bias is reflected in the (standard error) performance loss, which is most notable at higher velocities [see Figure 5(b)]. Hence, for narrow spectrum width weather, processing the data with an assumed SNR of 0 dB incurs the penalty of increased bias resulting from inadequate signal isolation.

Assuming an η of 0 dB does make sense for wide spectrum width weather [Figures 5(c) and (d)]. For low input SNR values, the optimal Bayes Bound is matched, and at higher input SNR values, performance appears to be no worse than that of PP. The bias of this implementation is also very similar to that of PP.

At the other extreme, processing the data assuming $\eta = 10$ dB appears to better PP performance in the case of narrow spectrum width signals but at the cost of a performance loss (vs PP) at low input SNR values and wider spectrum widths [Figures 5(c) and 5(d)]. The bias performance with a 10 dB window, if anything, does improve upon that of the ideal (η known) case.

5.2 Sensitivity to Incorrect σ

For this series, the known values for η were used in the algorithm and a set of values for the parameter σ, ranging above and below the true value, were tested. In contrast to varying η, the range of σ values tested did not appreciably change the bias results. Therefore, bias curves for this set are not presented. Figures 6 and 7 summarize the standard error results for ML and Bayes algorithms respectively.

For the narrow input spectrum [Figures 6(a), 6(b), 7(a), and 7(b)], σ values ranging from $\sigma_{true}/4$ to $5\sigma_{true}$[8] were tested; for the wide input spectrum [Figures 6(c), 6(d), 7(c), and 7(d)], values ranging from $\sigma_{true}/5$ to $9/5\sigma_{true}$[9] were tested. In plotting the results, the performance region spanned by underestimating σ is marked in white; the performance region spanned by overestimating σ is indicated with dark shading. Performance curves for each of the tested values are included as thin lines; the lowest curve in each panel always represents the performance when

[8]Values 0.25, 0.5, 1.0, 2, 3, 4, and 5 m/s ($\sigma_{true} = 1$ m/s) were examined.

[9]Values 1, 2, 3, 4, 5, 6, 7, 8, and 9 m/s ($\sigma_{true} = 5$ m/s) were examined.

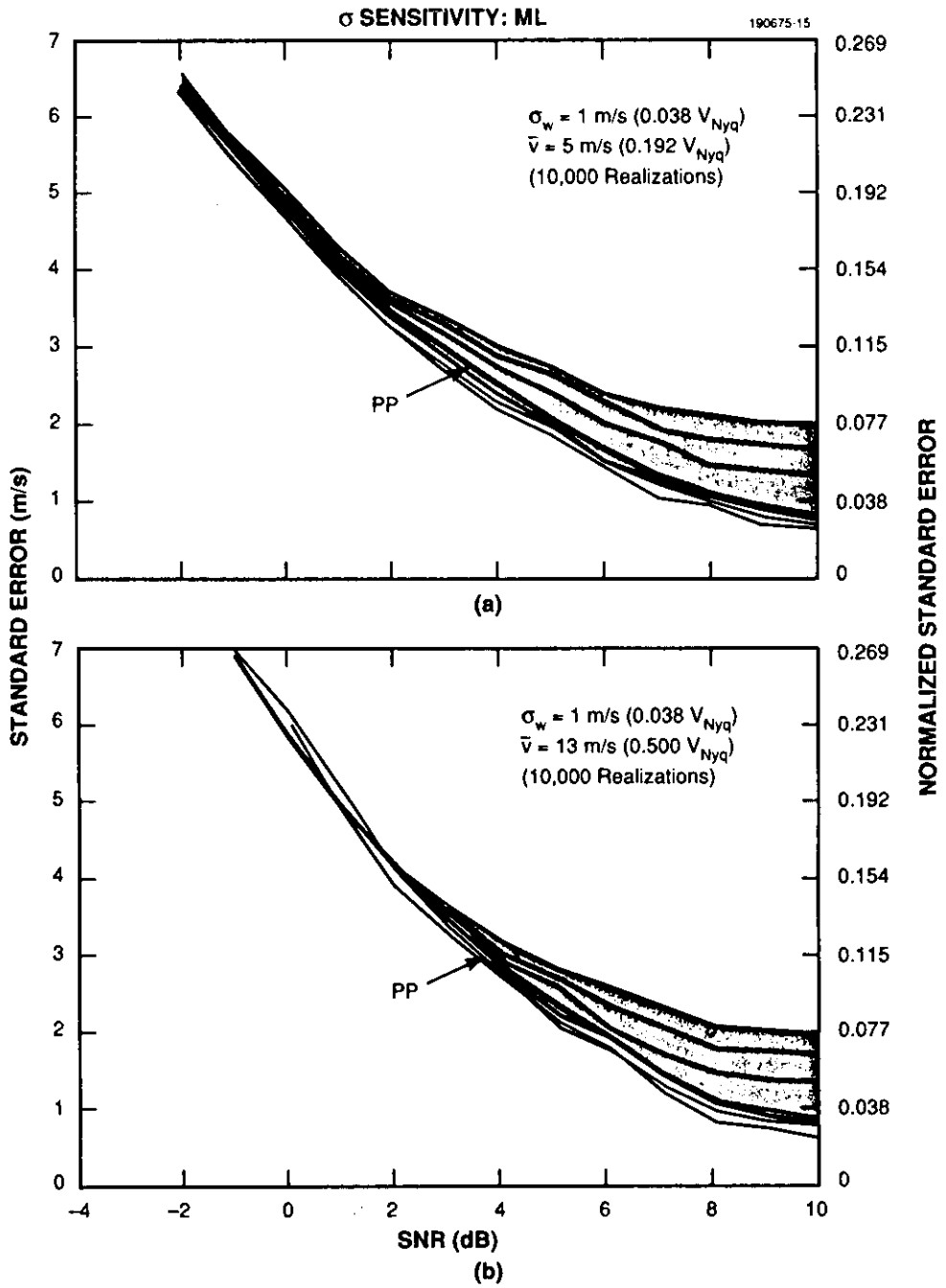

Figure 6.— ML algorithm performance sensitivity: effect of spectrum width parameter σ. (a) $\overline{v} = 5$ m/s, $\sigma_w = 0.038\, v_{Nyq}$. (b) $\overline{v} = 13$ m/s, $\sigma_w = 0.038\, v_{Nyq}$.

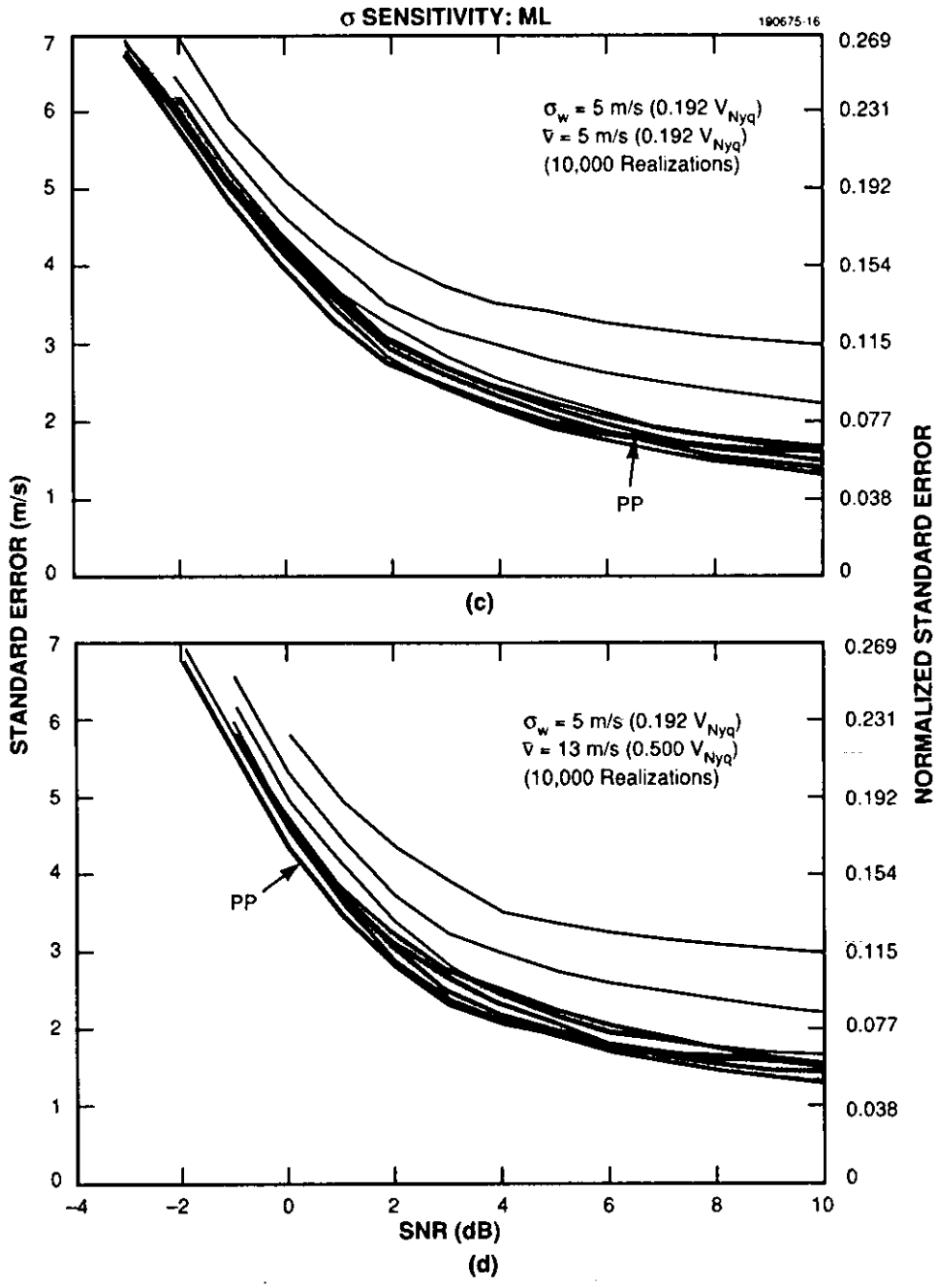

Figure 6. ML algorithm performance sensitivity: effect of spectrum width parameter σ. (c) $\bar{v} = 5$ m/s, $\sigma_w = 0.192\ v_{Nyq}$. (d) $\bar{v} = 13$ m/s, $\sigma_w = 0.192\ v_{Nyq}$.

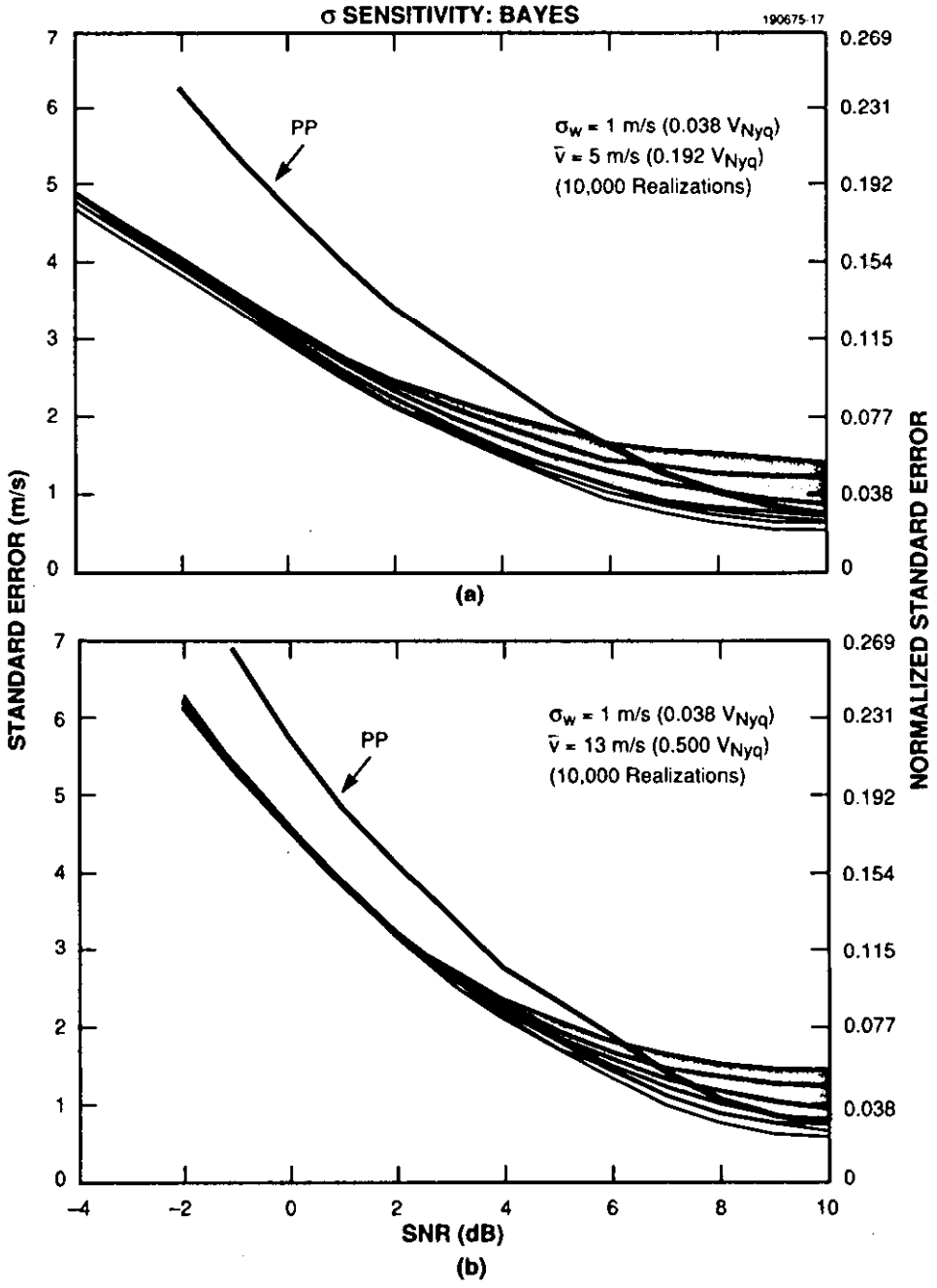

Figure 7. Bayes algorithm performance sensitivity: effect of spectrum width parameter σ. *(a)* $\bar{v} = 5$ *m/s,* $\sigma_w = 0.038\, v_{Nyq}$. *(b)* $\bar{v} = 13$ *m/s,* $\sigma_w = 0.038\, v_{Nyq}$.

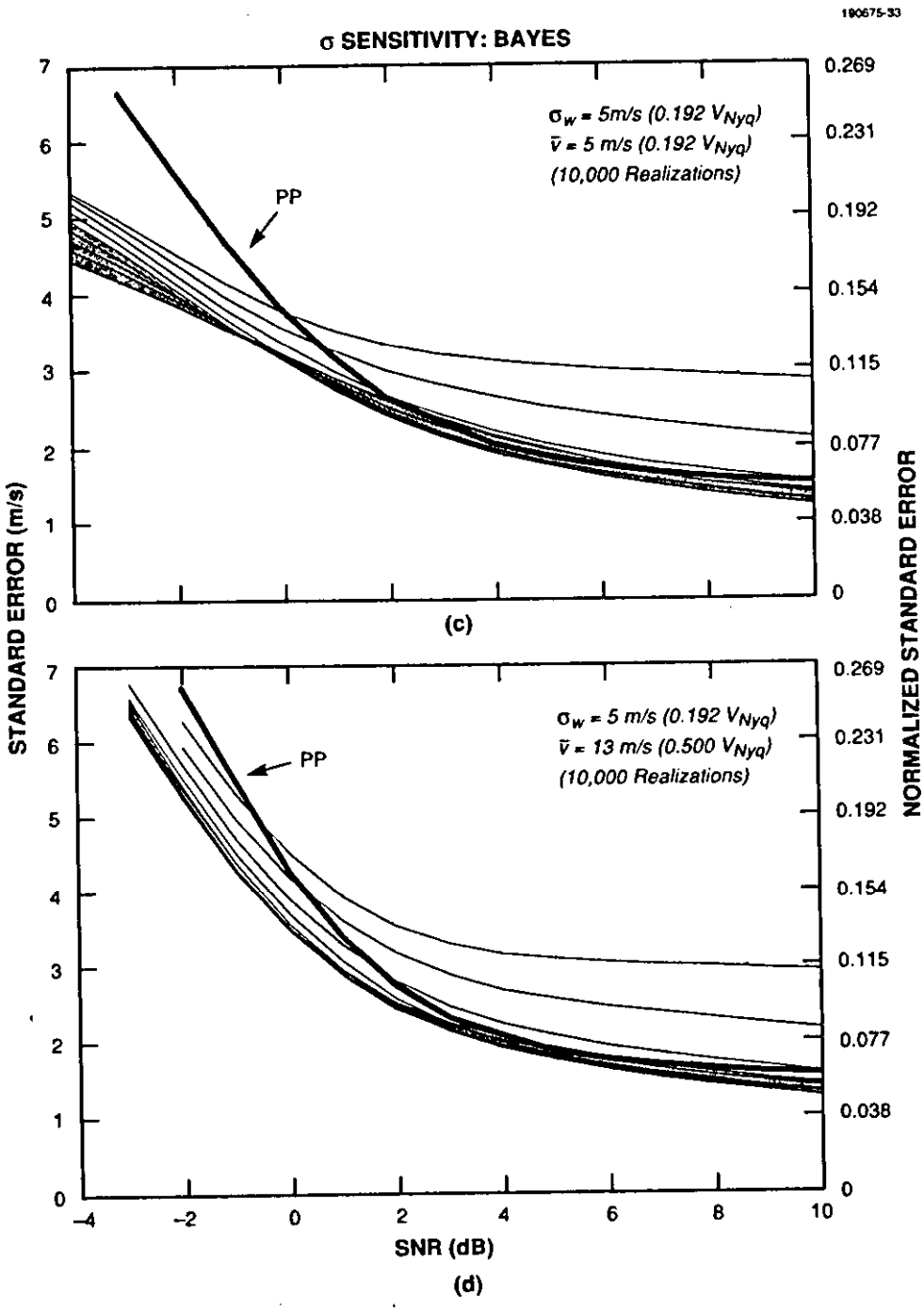

Figure 7. Bayes algorithm performance sensitivity: effect of spectrum width parameter σ. *(c)* $\bar{v} = 5$ *m/s,* $\sigma_w = 0.192\ v_{Nyq}$. *(d)* $\bar{v} = 13$ *m/s,* $\sigma_w = 0.192\ v_{Nyq}$.

there is a perfect match between the assumed σ value and that of the weather. The heavy curve in each panel is PP performance for reference.

In general, significant deviation from optimal performance occurs for σ parameter values grossly in error (using values 3, 4, and 5 m/s when σ_{true} = 1 m/s, and using values 1 and 2 m/s when σ_{true} = 5 m/s). Because ML performance, assuming σ and η known, is close to PP performance, not knowing the correct value for σ generally results in a significant performance loss [Figures 6(a-d)]. For narrow spectrum widths and SNR values in the range 0 to 5 dB, a marked performance gain still exists for the Bayes estimator, regardless of the value used for the parameter σ. The potential compromise at larger SNR values, however, indicates that the algorithm does not perform well with a σ value that is too distant from the underlying true value.

Comment. The WP method, prior to noise subtraction and censoring, can be viewed as a Bayes algorithm wherein the parameter σ is assumed to have an arbitrary narrow width (no frequency domain smoothing is done). The plots in Figure 7(c) and 7(d) confirm this notion as it can be seen that the Bayes performance curves approach those of the WP method (see Figure 3) when a narrow σ is assumed but the weather possesses a wide spectrum width.

5.3 Summary

It is unclear whether arbitrary fixed values for σ and η can guarantee an estimator performance that is consistently better than that of PP. Bayes performance relies heavily on knowledge of both σ and η, ML performance, more appreciably on σ.

In the smoothing window view of Section 3.1.1, it is not enough to smooth arbitrarily. It is important that the data be processed with an amount of smoothing that matches correlation strength between samples. Treating the data as being highly correlated (low σ) demonstrated a severe performance loss when weather signals in fact had wide spectrum widths. This explains why periodogram based algorithms, such as WP, do not perform well given wide spectrum width weather (too much weight is given to higher lag products). Unfortunately, treating the data as being largely uncorrelated (high σ) resulted in a corresponding performance loss with input signals having narrow spectrum widths and high SNR values. Nevertheless, the indications are good that improved performance (relative to PP) is possible with a smoothing methodology requiring less than perfect knowledge of σ and η.

The ML implementation at best only matches PP performance; any performance gain at higher SNR levels would clearly be compromised by inaccurate knowledge of σ. (Hence, the ML implementation will not be considered further in this report.) A successful Bayes implementation will require an adaptive selection of smoothing (weighting) coefficients, which is the focus of the next section.

6. PERFORMANCE WITH ADAPTIVE σ AND η

With an emphasis on the estimation of ω, one can adopt the view whereby σ and η are treated as nuisance parameters. Here, there exists a natural Bayesian method of treatment: removal through expectation. This logical recourse unfortunately does not lead to algorithm simplification (in the present case), requiring as much computation as that needed for solving the vector parameter problem. (It appears that the integrations required for nuisance parameter removal must be done numerically and also require the data vector to be in hand.) Coupled with this observation, the results of the previous sections motivate an approach that seeks to adapt the computational form of Equation (12) using (suboptimal) estimates for σ and η. Simple estimation of σ and η (using method of moments estimates described below) and direct substitution into Equation (9) and (12) (sample by sample) were found to yield a performance clearly worse than that of PP. This is not (entirely) unexpected because, like ω, σ and η are being estimated from a small sample and (ω-performance) sensitivity to large deviations from the true σ and η values can, on average, do more harm than good. Clearly, an approach is needed whereby the estimated values $\hat{\sigma}$ and $\hat{\eta}$, substituted into Equation (12) via Equation (9), are suitably constrained to minimize penalties that result from their inaccuracy. This section describes one such approach that was found to be successful.

The general processing strategy considered is as follows. Given a data sample $\mathbf{Z}$, suboptimal estimates of σ and η are first computed and used to select a weighting coefficient array (*matched filter*) Γ from a small, fixed, and predetermined family of matrices; the chosen matrix is used to process the data as per Equation (12) and provide a velocity estimate. The number of matrices required, their coefficient specification, and the criteria used to choose from among them is the subject, then, of this present section.

6.1 Constrained Inverse Filter (Γ) Selection

This section focuses on the nuisance pair (σ, η) as an element of a (parameter) set Ψ that is assumed to be partitioned into K disjoint pieces, i.e.,

$$\Psi = \bigcup_{k=0}^{K-1} \Psi_k, \text{ where } \Psi_i \cap \Psi_j = \emptyset \; (i \neq j).$$

Each region Ψ_k is assigned an optimal representor $\overline{(\sigma, \eta)}_k \in \Psi$, and a corresponding weighting matrix Γ_k is computed as per the definition [Equation (9)]. A representor for a given Ψ_k is determined by means of a minimization involving the directed divergence [7] (i.e., Kullback-Leibler information)

$$I(p:q) = E_p\left[\log \frac{p(\mathbf{Z})}{q(\mathbf{Z})}\right]. \tag{16}$$

The divergence $I(p : q)$ has a useful *interpretation* as a distance,[10] measuring an ability to discriminate probability density p (and, hence, its corresponding model) from alternative q (note: $I(p : q) \geq 0$ and $I(p : q) = 0 \Leftrightarrow p \equiv q$). Furthermore, the divergence [Equation (16)] is easily computed for the Gaussian case [Equation (3)]: if Γ_1 and Γ_2 correspond, respectively, to densities p_1 and p_2, then

$$I(p_1 : p_2) = \log|\Gamma_1| - \log|\Gamma_2| + \mathrm{tr}(\Gamma_2\Gamma_1^{-1} - I).$$

A somewhat natural approach, then, is to define the optimal representor for Ψ_k as that pair (σ,η) whose corresponding density [Equation (3)] minimizes the discrimination information (divergence) averaged over all densities q corresponding to parameter pairs in the set Ψ_k:

$$\overline{(\sigma,\eta)}_k \stackrel{\text{def}}{=} \arg\min_{(\sigma,\eta)\in\Psi} \int_{\Psi_k} I(q_\xi : \overline{p}_k)\, d\xi, \tag{17}$$

where $\overline{p}_k$ is the density corresponding to $\overline{(\sigma,\eta)}_k$ and ξ is an index to pairs (σ, η) in Ψ_k. In words, as representor for the set Ψ_k, select that density (model) which, in an average sense, is least distinguishable from the feasible densities (models) in Ψ_k. Note, no restriction is made that the point $\overline{(\sigma,\eta)}_k$ must be contained in Ψ_k.

Example. Consider the (trivial) case where it is assumed that $\Psi = (0, 0.25] \times (0, 20]$ and $K = 1$. That $K = 1$ is specified assumes all data can be processed with the weighting coefficients corresponding to one (arbitrary) set point $\overline{(\sigma,\eta)}_0$. For this case it is an easy matter to solve Equation (17),[11] and one obtains the solution $\overline{(\sigma,\eta)}_0 = (0.165, 8.4)$. Figure 8 shows the performance of the resulting estimation algorithm. Although performance is near optimal for wide spectrum widths and high SNR levels (locations near the set point), uniform improvement for the entire range of parameter values in Ψ is absent, and performance is severely compromised in places as well. One must conclude that this Ψ is too large to be represented by one set of weighting coefficients.

Figure 9(a) is a plot of the divergence for the above example. Divergence is near zero in the vicinity of the set point and curves upward (away from zero) as weather spectrum width and SNR deviate from the set point. Note, also, that the curvature is not uniform in direction—the most acute curvature occurs with respect to spectrum width as weather SNR becomes large. Clearly, the goal is to devise an adaptive method that has a composite divergence surface as flat and as near

---

[10]Technically, the divergence fails as a true distance because it does not satisfy the triangle inequality. This failing, however, does not prevent its use in the present application.

[11]Optimization was achieved by means of an implementation of the Nelder-Mead modified polytope (direct-search) algorithm (see, for example, Gill et al. [8]).

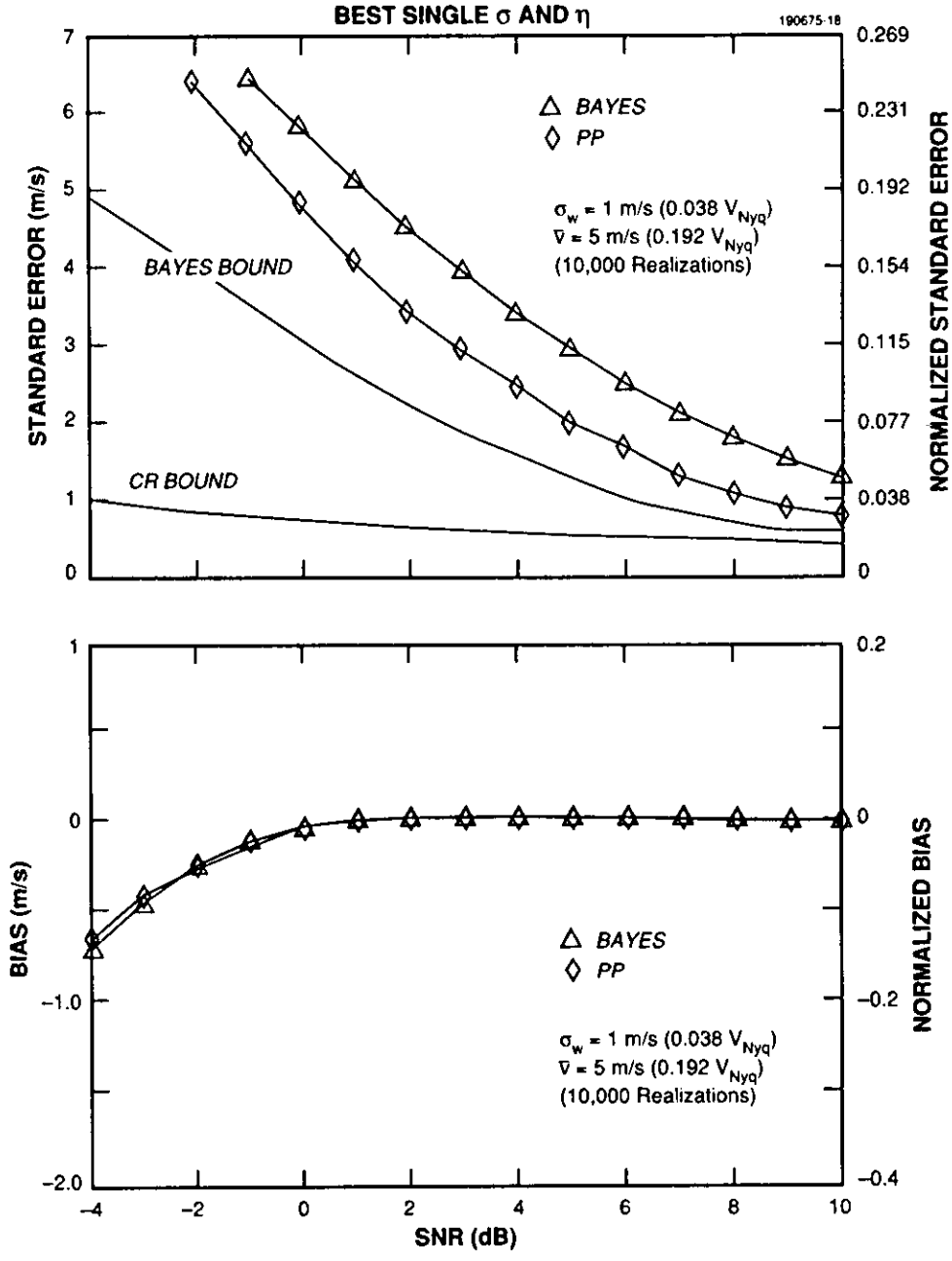

Figure 8. Bayes algorithm performance: suboptimal σ and η using best single-weighting coefficient set. (a) $\overline{v} = 5$ m/s, $\sigma_w = 0.038\, v_{Nyq}$.

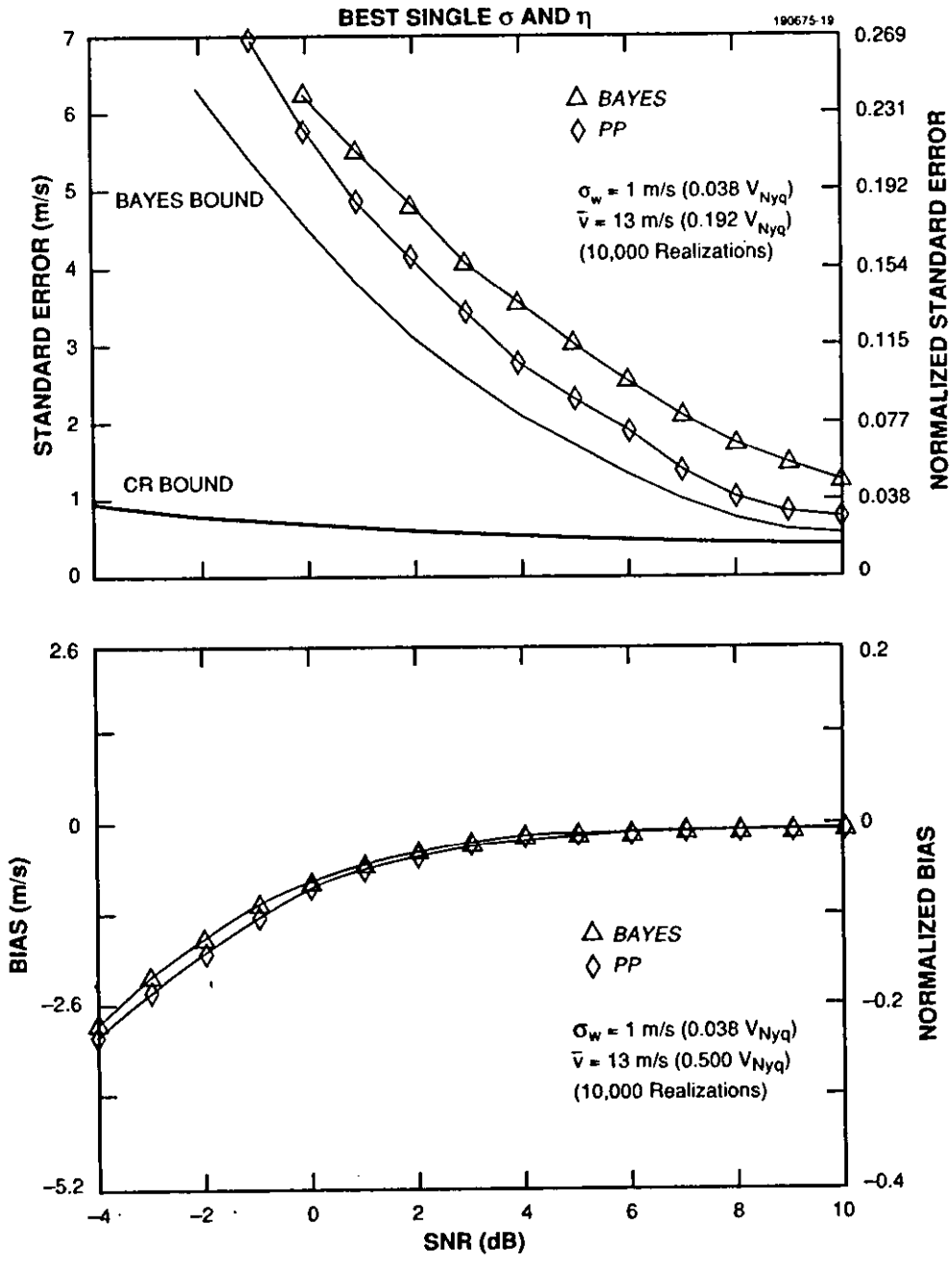

Figure 8. Bayes algorithm performance: suboptimal σ and η using best single-weighting coefficient set. (b) $\bar{v} = 13$ m/s, $\sigma_w = 0.038\ v_{Nyq}$.

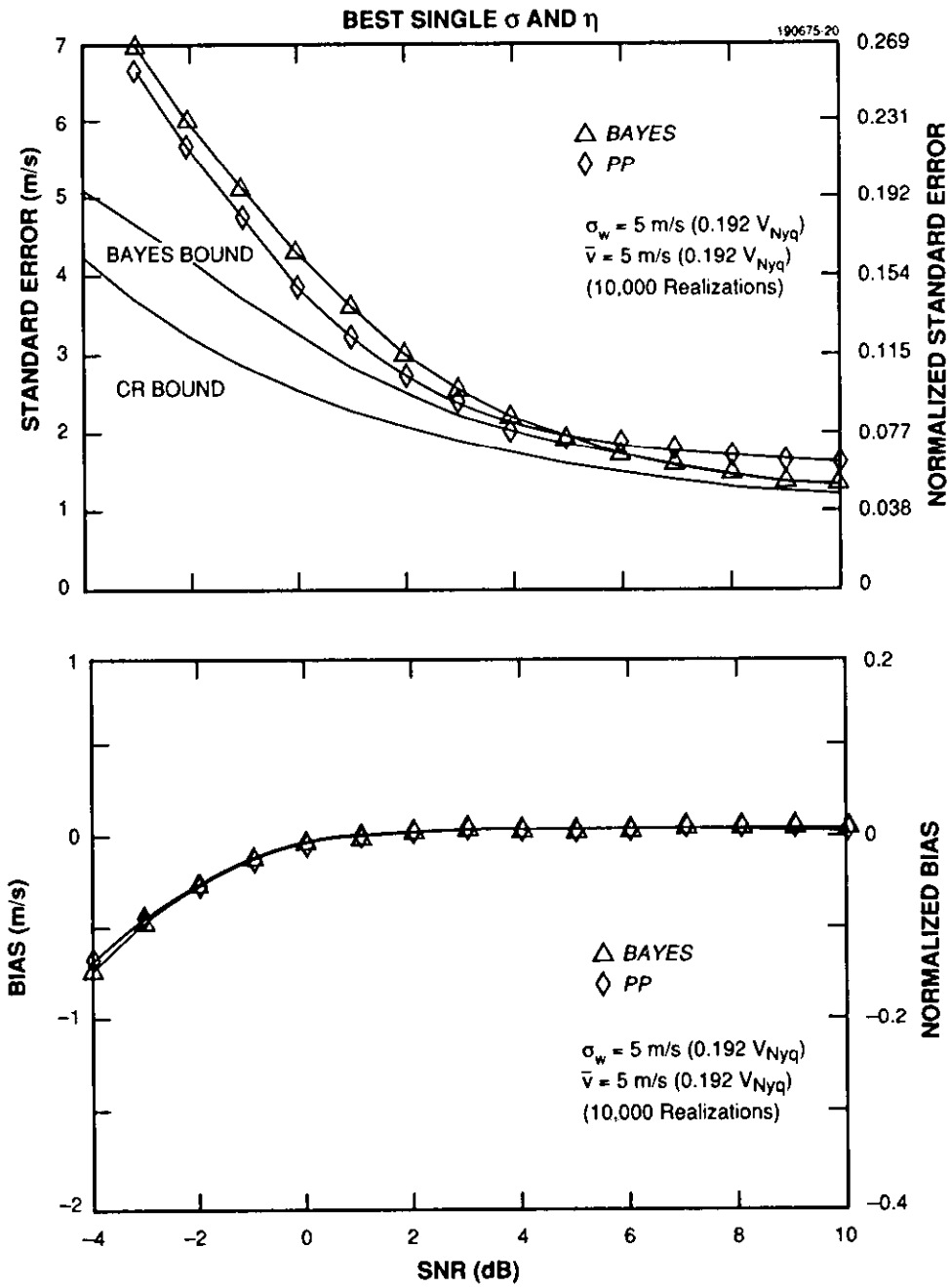

Figure 8. Bayes algorithm performance: suboptimal σ and η using best single-weighting coefficient set. (c) $\bar{v} = 5$ m/s, $\sigma_w = 0.192\, v_{Nyq}$.

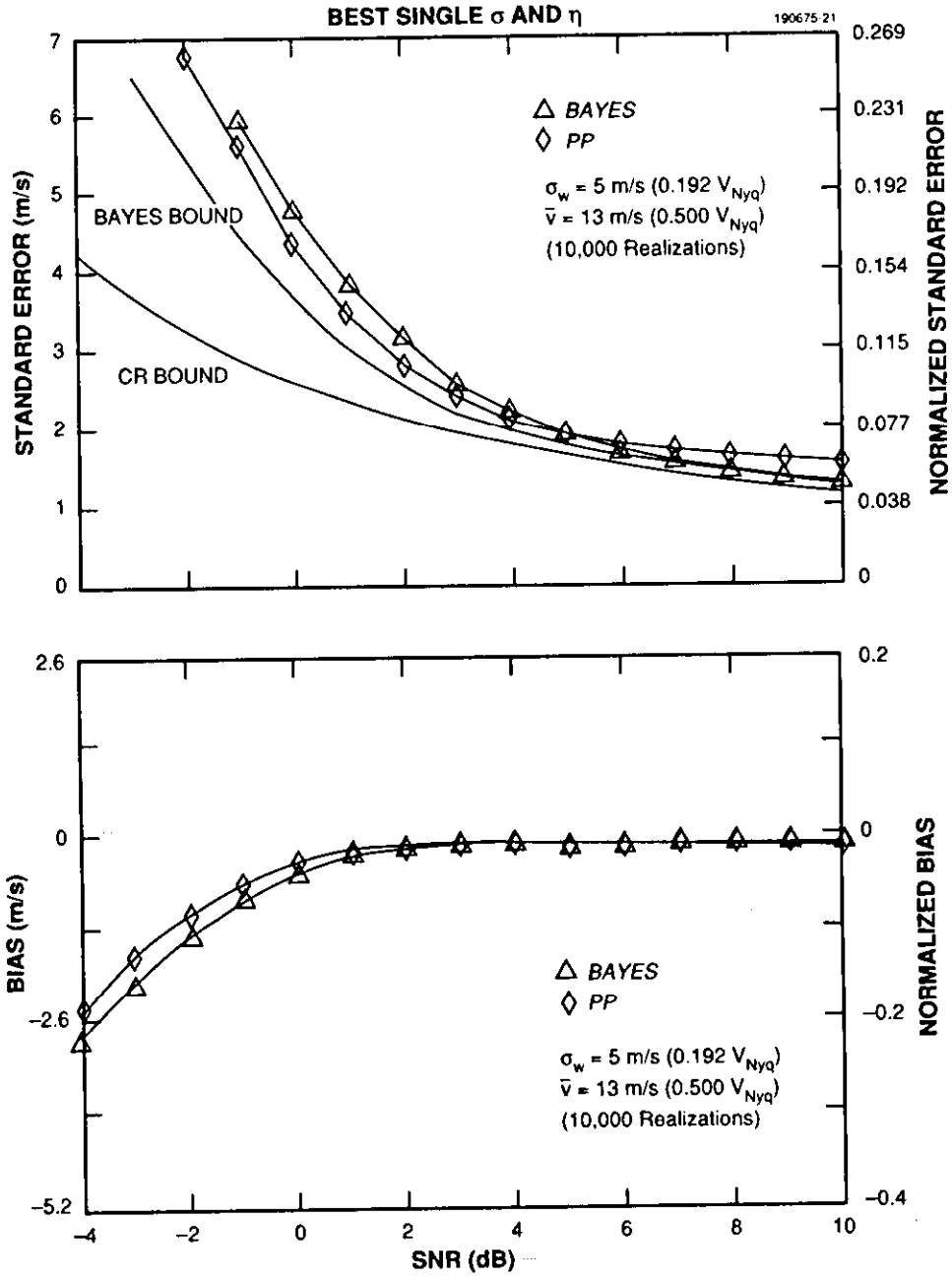

Figure 8. Bayes algorithm performance: suboptimal σ and η using best single-weighting coefficient set. (d) $\bar{v} = 13$ m/s, $\sigma_w = 0.192\, v_{Nyq}$.

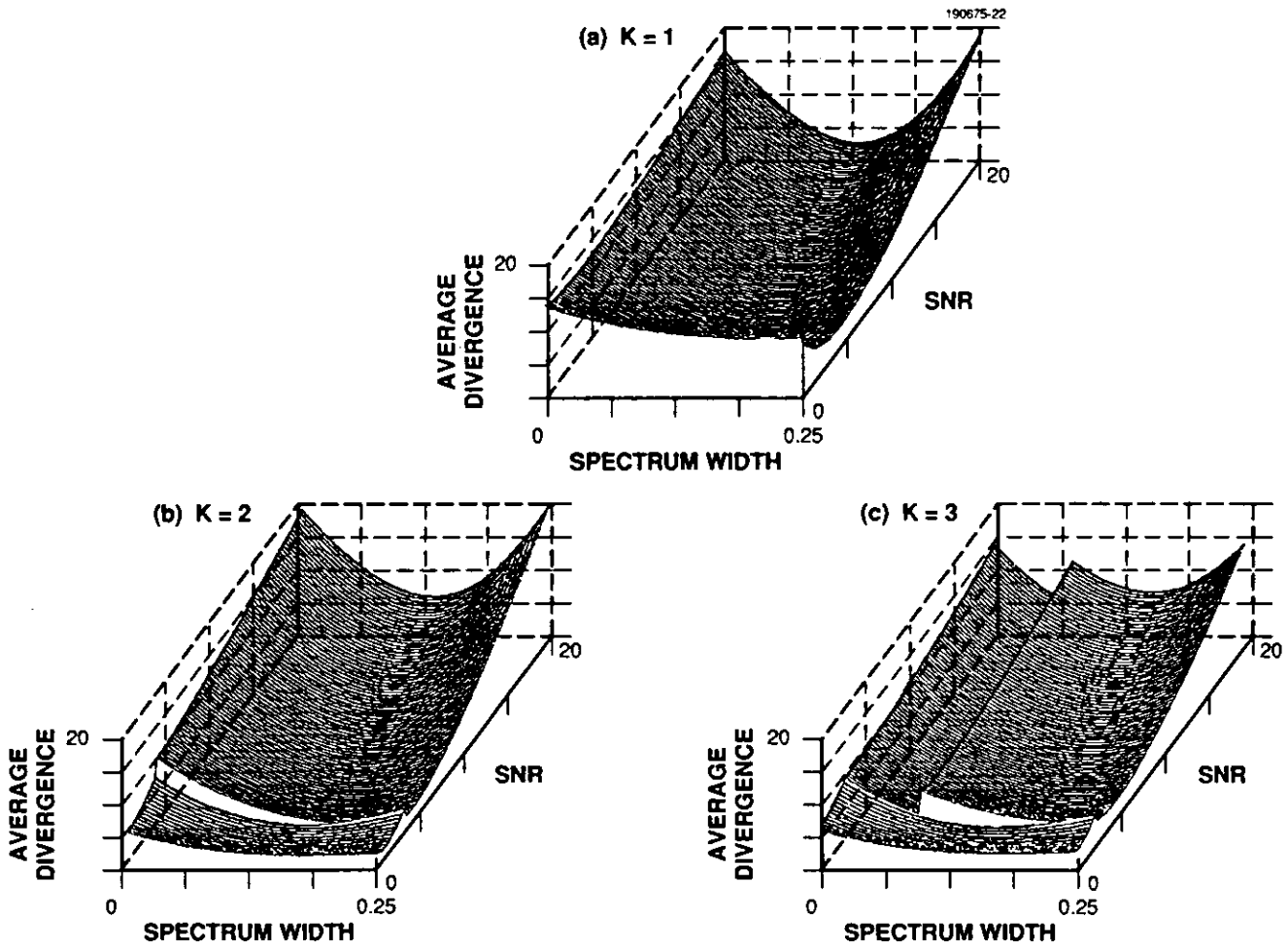

Figure 9. Divergence minimization for three simple partitions. Computed directed divergence between an optimal compartment representor [Equation (17)] and the feasible model densities of its compartment is plotted. (The divergence scale is arbitrary.)

zero as possible. The logical extension, of course, is to seek improvement by solving the situation for $K > 1$. Two simple extensions, a two and three compartment model, are also illustrated in Figure 9(b-c). Here, the divergence surface for each compartment is computed as per Equation (17); although not yet striking, the notion of flattening the composite surface is clear.

There is, however, a problem with the simple extension of Equation (17) to $K > 1$; Equation (17), as written, has the implicit assumption that one would (could) distinguish perfectly between the opposing hypotheses as to whether data Z were more consistent with parameters from the set Ψ_k vs alternatives from its complement $\Psi_k^c = \Psi - \Psi_k$. Clearly, for $K > 1$, Equation (17) must be modified to account for the accuracy of the decision process that matches the data to one of the subsets Ψ_k:

$$\overline{(\sigma, \eta)}_k \stackrel{\text{def}}{=} \underset{(\sigma,\eta)\in\Psi}{\arg\min} \left\{ Pr(\text{select } \Psi_k \mid \Psi_k) \cdot \int_{\Psi_k} I(q_\xi : \overline{p}_k)\, d\xi \right.$$
$$\left. + Pr(\text{select } \Psi_k \mid \Psi_k{}^c) \cdot \int_{\Psi_k{}^c} I(q_\xi : \overline{p}_k)\, d\xi \right\}. \tag{18}$$

In this way, the desire to optimally match $\overline{(\sigma, \eta)}_k$ to Ψ_k is balanced against the probability that the data will be incorrectly matched with Ψ_k.

6.2 Suboptimal Estimation of σ and η

The calculations for Equation (18) require specification of a set of decision rules and the corresponding statistics (estimators) to be used; however, once this is done all information required to solve Equation (18) is present and the selection of Γ_k, $(k = 0, \ldots, K-1)$, can be completed prior to the processing of any data. Therefore, although computationally formidable, the solution of Equation (18) is quite achievable. This section will focus on decision rules using easily computable suboptimal estimates for σ and η. For suboptimal estimates, an appeal to the assumed correlation structure of Equation (1) and (method of moments) estimates for $\hat{\sigma}$ and $\hat{\eta}$, derived from a weighted least-squares fit to the data, can be used. The least-squares equations

$$\sum_{m=0}^{M-1} w_m \ln|\hat{r}_m| = \sum_{m=0}^{M-1} w_m \ln[\mathcal{S} + \mathcal{N}\delta_m] - \frac{1}{2}\pi^2\sigma^2 \sum_{m=0}^{M-1} w_m m^2 \tag{19a}$$

and

$$\sum_{m=0}^{M-1} w_m m^2 \ln|\hat{r}_m| = \sum_{m=0}^{M-1} w_m m^2 \ln[\mathcal{S} + \mathcal{N}\delta_m] - \frac{1}{2}\pi^2\sigma^2 \sum_{m=0}^{M-1} w_m m^4 \tag{19b}$$

can be used to solve for σ and $\mathcal{S}$ (which provides an estimate for η); the lag estimates $\hat{r}_m$ are unweighted here and the (least-squares) weights w_m can be used to weight or select the lag estimates

to be used. If $w_m = \delta_{m-1}$, for example, $\hat{\sigma}$ becomes equivalent to the common single-lag spectrum width estimator [see Zrnić [9], his Equation (5.1)]. For the results of this report,

$$w_m \stackrel{\text{def}}{=} \begin{cases} 1 & \text{for } m < 4 \\ 0 & \text{otherwise} \end{cases}$$

was used for estimation of σ and η.

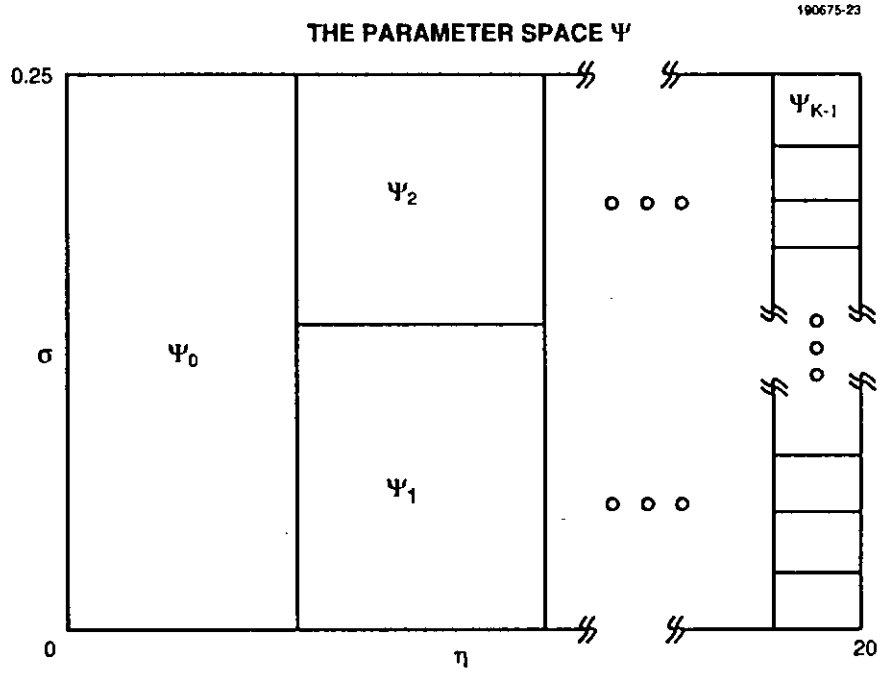

Figure 10. A partition design using simple decision rules.

6.3 Decision Rules and a Partition for Ψ

At this point it is necessary to be more specific regarding the description of the Ψ_k's. For the remainder, it is assumed that Ψ is the parallelepiped of the previous example: $(0, 0.25] \times (0, 20]$. To simplify definition of the Ψ_k's and to provide decisions based on simple rules, consider a partition derived from a sequence of threshold tests—first, for $\hat{\eta}$ and second, for $\hat{\sigma}$. The general design is illustrated in Figure 10. A threshold test of $\hat{\eta}$ against (yet undetermined) SNR values divides Ψ into

columns of (yet undetermined) width. Each column is further subdivided along spectrum width values (also to be determined) and a requirement is imposed whereby increased partitioning of a column with respect to spectrum width requires correspondingly high values for SNR. (The surface curvature of Figure 9(a) suggests that more detailed representation is desired for high SNR values than low.) To simplify implementation, each new column to the right is allowed one additional division with respect to spectrum width. The placement of SNR and spectrum width thresholds, the selection of $\overline{(\sigma, \eta)}_k$ values, and specification of K are all unknowns to be determined.

For K fixed, Equation (18) can be used to identify optimal values for all threshold boundaries and all $\overline{(\sigma, \eta)}_k$ pairs. A K-compartment summed divergence error can be defined by summing the divergence error in Equation (18) over each set in the partition of Ψ. The K-compartment summed divergence is monotonic (nonincreasing) in K. Certainly adding extra compartments in Figure 10 can only lower the total divergence error. For example, optimal threshold placement would force new compartments to become degenerate (collapse to nothing) if they could not improve the overall error value; lower resolution compartments to the left would get squeezed by higher resolution compartments from the right if the data and estimators for σ and η could support finer levels of partition. Hence, as K increases, the K-compartment summed divergence error (bounded below by zero) must converge. A stopping criteria can be established for selecting K by arguing diminishing returns with further increases in K. In this way, K, optimal threshold placement, and optimal set point values can be obtained.

Figure 11 plots the K-compartment summed divergence for the partition scheme illustrated in Figure 10. Evaluation of Equation (18) was approximated by using a Monte Carlo simulation to obtain a mean and standard deviation characterization for the $\hat{\sigma}$ and $\hat{\eta}$ of Equation (19), and a Gaussian approximation was used to evaluate $Pr(\text{select } \Psi_k | \Psi_k)$ and complement. Based on the results presented in Figure 11, $K = 15$ was selected for continued analysis. The corresponding thresholds and set point values for $K = 15$ are summarized in Table 2 and illustrated in Figure 12.

In Figure 12, optimal set point locations are indicated by a labeled (square) dot. The convex hull of the set points in Ψ is illustrated by shading. The most striking result of the optimization is that the best set point for a region Ψ_k is not necessarily contained within Ψ_k. The convex hull of the set points represents the constraint required of $\hat{\sigma}$ and $\hat{\eta}$ to balance the effect (on velocity estimation) of their uncertainty. Figure 13 plots the sequence of convex hulls for the values of K considered in Figure 11. As K increases, the convex hull expands but there is a fundamental limitation to its eventual extent. To increase the extent of the limiting hull requires more precise estimators for σ and η or, equivalently, more data. The limiting hull will always be strictly contained within Ψ: for the hull to be equivalent to (i.e., cover) Ψ implies that perfect estimation of σ and η is possible.

6.4 Adaptive Estimator Performance

The adaptive procedure defined by Table 2 (Figure 12) was used to process simulated data as in the previous sections. The results for narrow and wide spectrum width weather are presented in Figure 14.

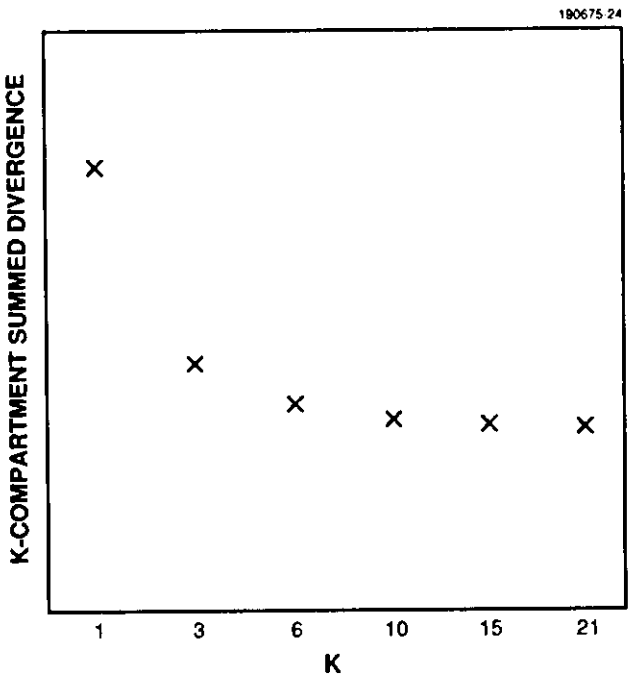

Figure 11. K-Compartment summed divergence.

For narrow spectrum width weather, Panels *a* and *b*, the adaptive method demonstrates a near uniform improvement on the performance of PP. The performance is not only better than that of PP, it is much closer to the optimum bound established in earlier sections. The slight deterioration at the higher SNR values may be due, in part, to the decision to use the 15 compartment partition. Figure 13 indicates that the 21 compartment partition added refinement specific to higher SNR values; therefore, some further improvement at higher SNR values may be possible. Note, also, that in parallel with approaching the standard error bound for velocity estimation, the adaptive estimator also approaches the bias performance of the optimal Bayes (bound) estimator.

For wide spectrum width weather, Panels *c* and *d*, the room for improvement was not as evident (except at low and very high SNR values). Nevertheless, the adaptive method demonstrated performance improvement relative to PP at low and high SNR values—those regions where improvement relative to optimal performance was most likely.

TABLE 2

Optimal Thresholds and Set Point Values ($K = 15$)

Compartment	Set Point	Upper η Threshold	Upper σ Threshold
0	(0.0983, 2.386)	1.367	-
1	(0.1095, 3.892)	3.545	0.0909
2	(0.1659, 3.794)		-
3	(0.1041, 6.440)	7.047	0.0549
4	(0.1452, 7.082)		0.1279
5	(0.1921, 6.960)		-
6	(0.0907, 9.554)	12.006	0.0380
7	(0.1317, 9.723)		0.0903
8	(0.1667, 10.203)		0.1476
9	(0.2039, 10.916)		-
10	(0.0811, 13.093)	-	0.0354
11	(0.1186, 13.822)		0.0821
12	(0.1554, 13.927)		0.1252
13	(0.1848, 15.046)		0.1844
14	(0.2155, 15.438)		-

The data of the previous figure was combined with measurements at additional weather spectrum widths to produce Figure 15, which illustrates estimator performance as a function of weather spectrum width for fixed levels of SNR. There is an almost uniform improvement in performance relative to PP and this improvement for a fixed SNR level is seen to apply across the range of spectrum widths considered. The improvement is most striking for the two low SNR levels (Panels *a* and *b*). Improvement *does not* require a corresponding performance loss at higher SNR levels—the adaptive method continues to improve performance there as well.

6.5 Summary

Applying adaptive procedures to situations with small-sample sizes is a difficult task. This section has shown that by adopting suitable constraints, an adaptive method could be developed for the small sample velocity estimation problem.

Performance for the adaptive (Bayes) estimator is close to the optimal performance bounds established earlier. Significant improvement relative to PP was established.

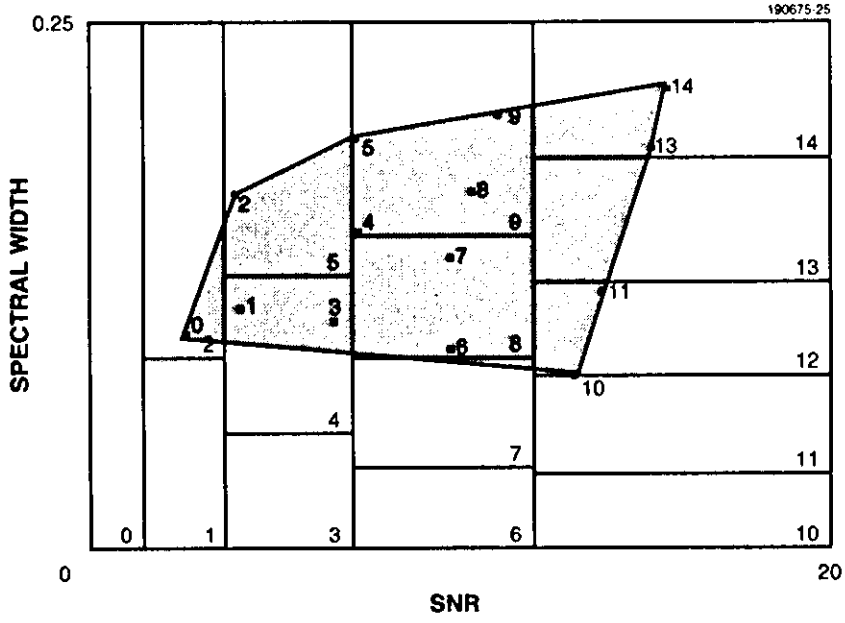

Figure 12. Optimum partition and set point locations.

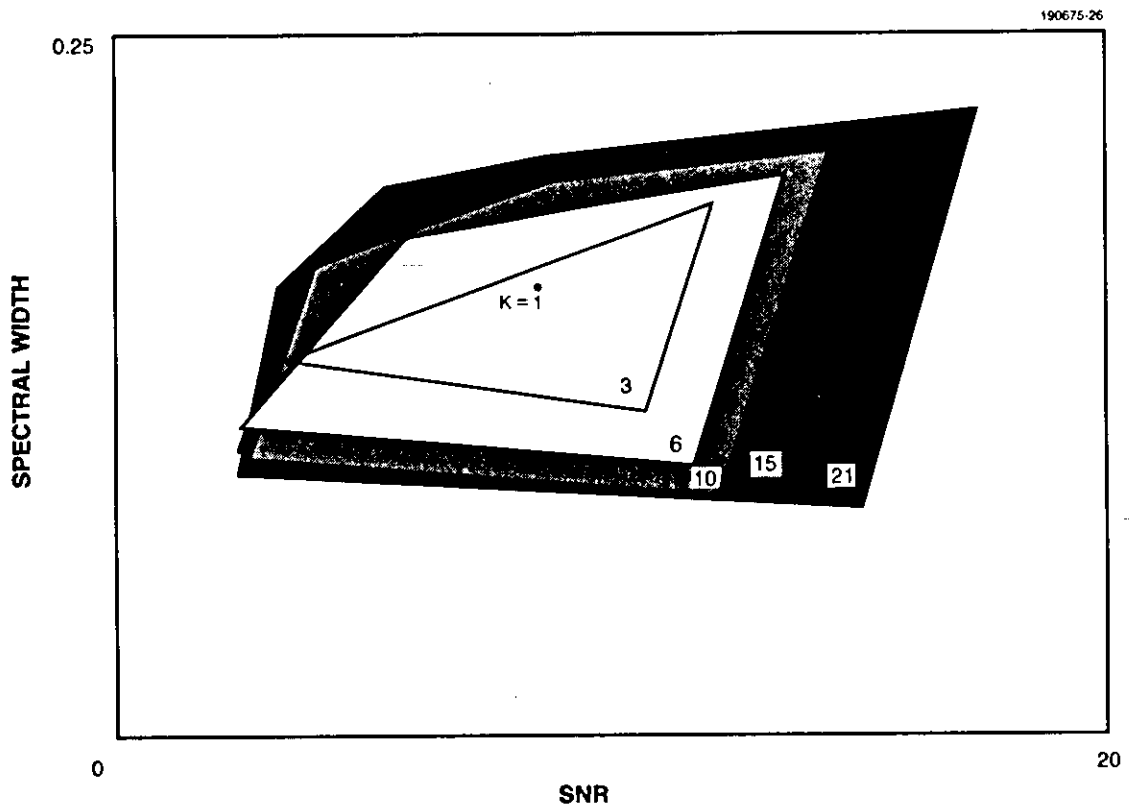

Figure 19. Set point convex hull with increasing K. Fundamental limitation on extent of set point convex hull is illustrated for a sequence with increasing K value.

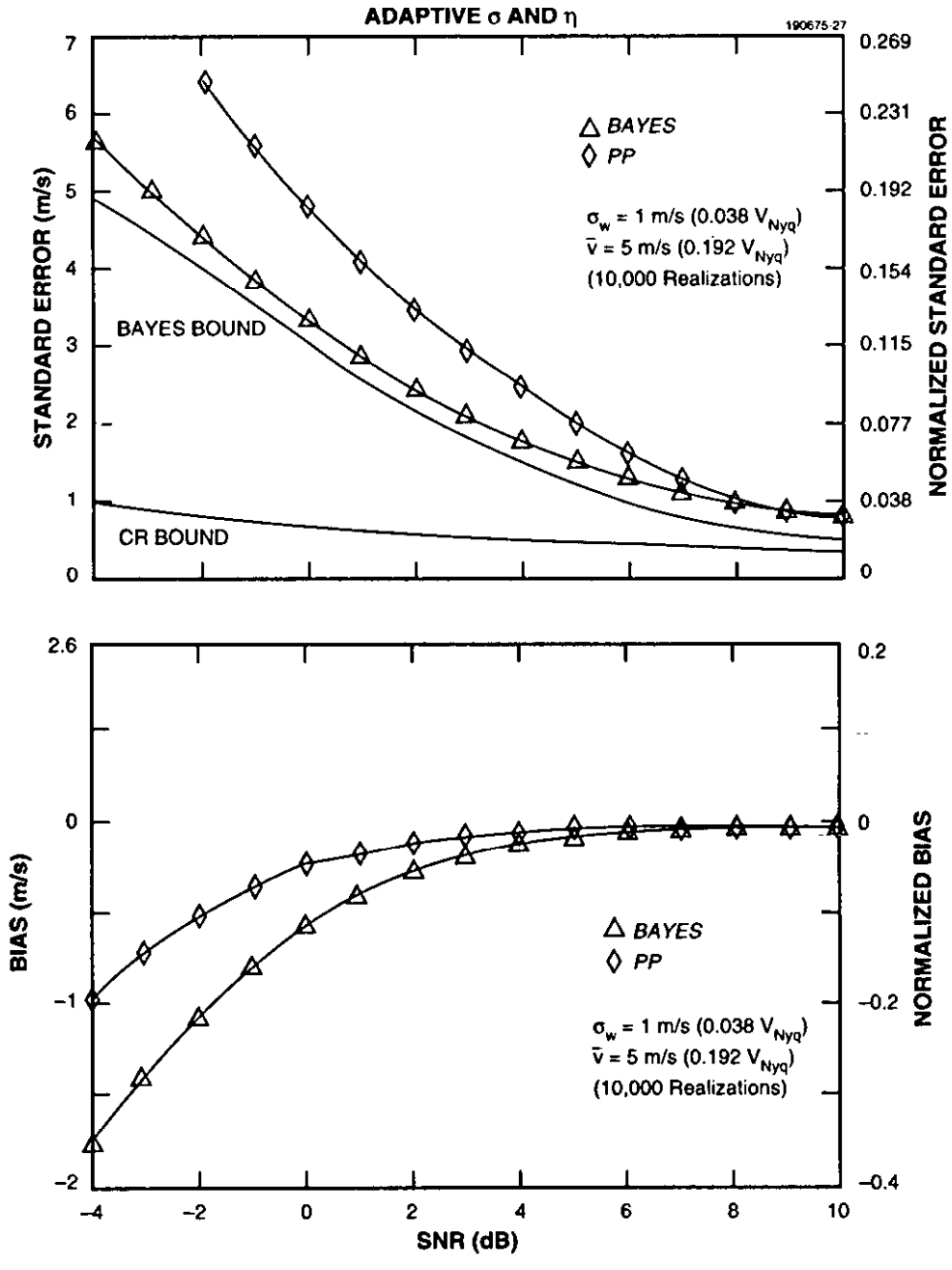

Figure 14. Bayes algorithm with adaptive σ and η: 15 compartment weighting coefficient set (I). (a) $\bar{v} = 5$ *m/s,* $\sigma_w = 0.038\, v_{Nyq}$.

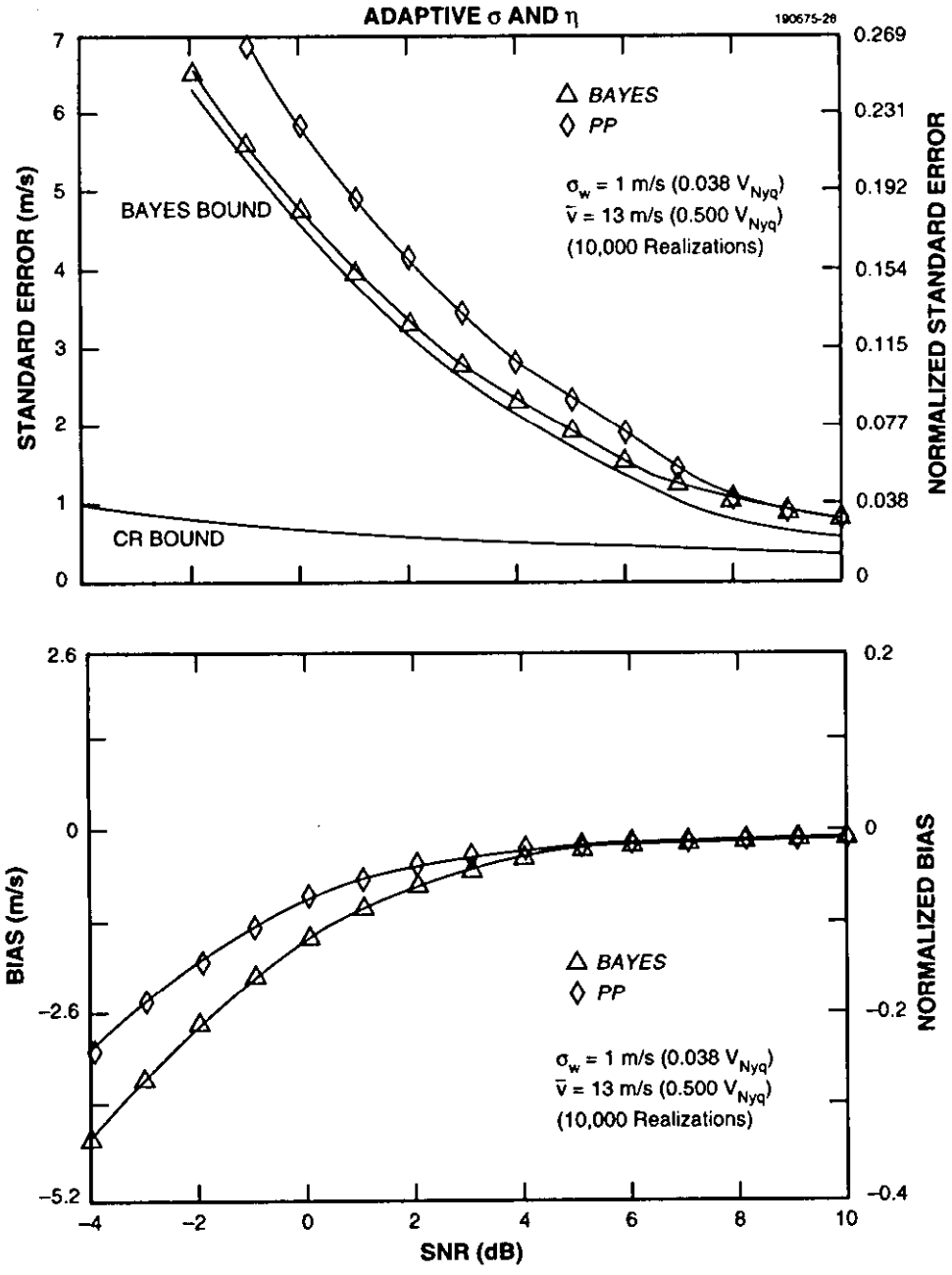

Figure 14. Bayes algorithm with adaptive σ and η: 15 compartment weighting coefficient set (1). (b) $\bar{v} = 13$ m/s, $\sigma_w = 0.038\ v_{Nyq}$.

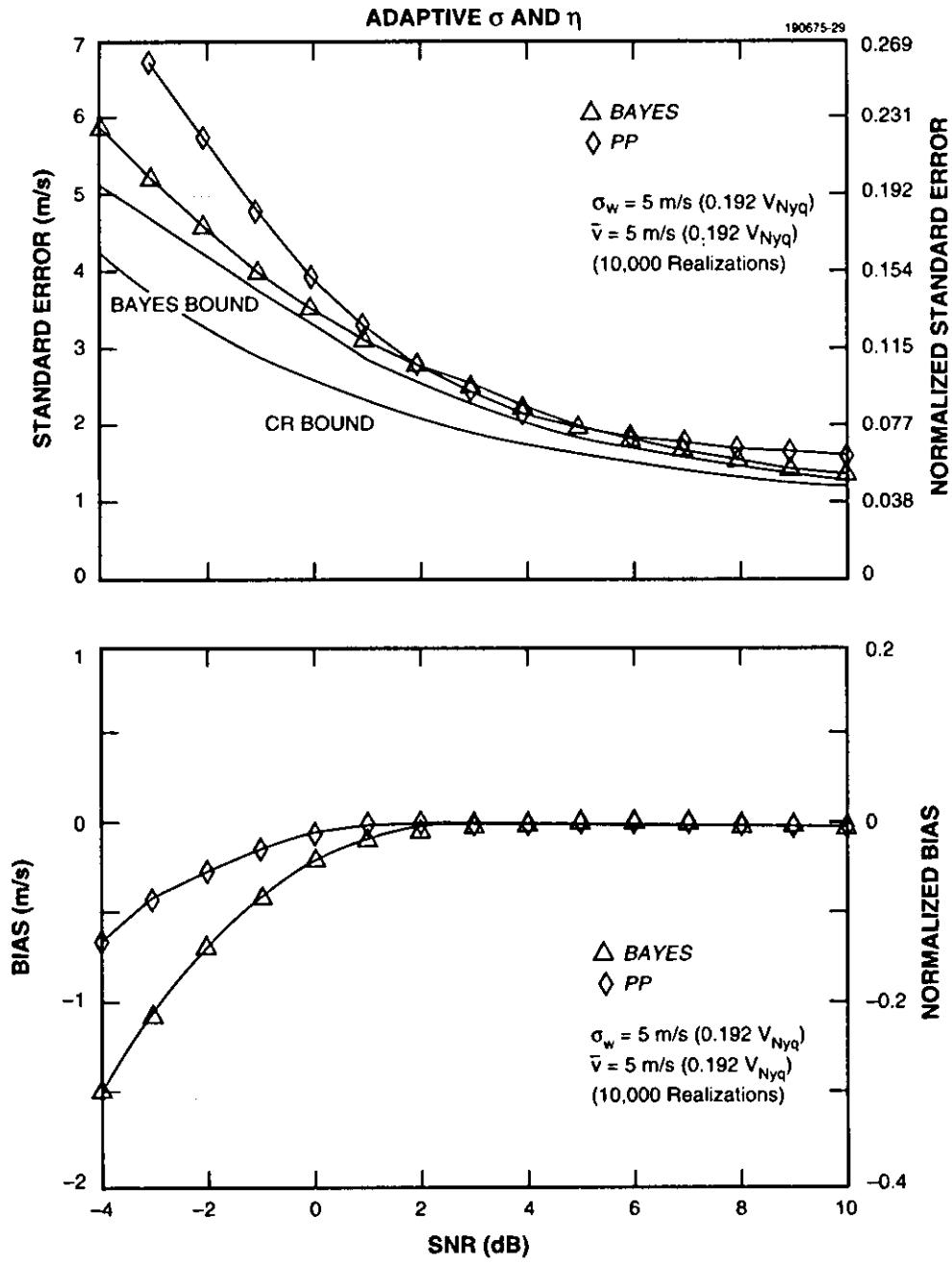

Figure 14. Bayes algorithm with adaptive σ and η: 15 compartment weighting coefficient set (I). (c) $\bar{v} = 5$ m/s, $\sigma_w = 0.192\, v_{Nyq}$.

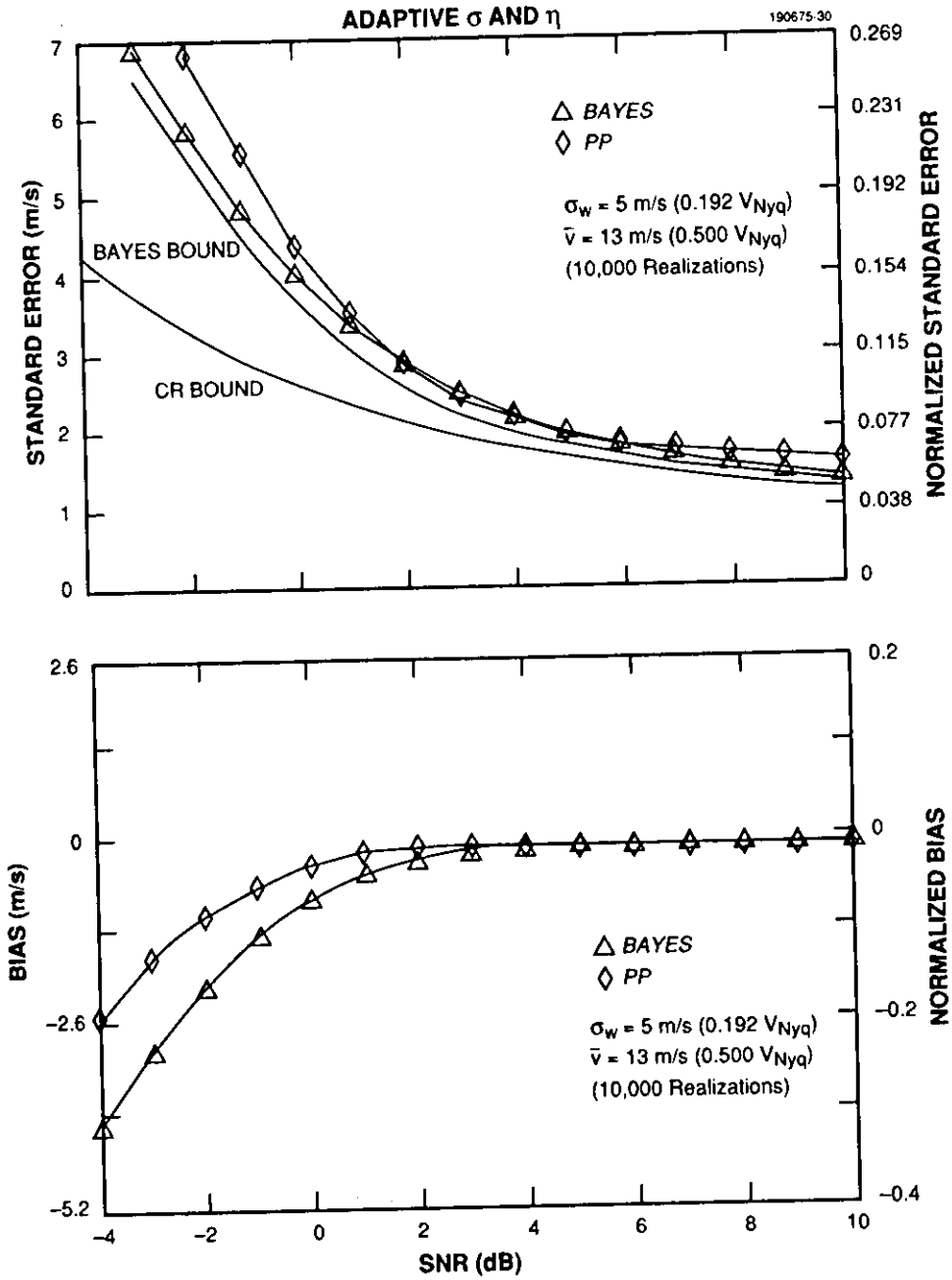

Figure 14. Bayes algorithm with adaptive σ and η: 15 compartment weighting coefficient set (I). (d) $\bar{v} = 13$ m/s, $\sigma_w = 0.192\, v_{Nyq}$.

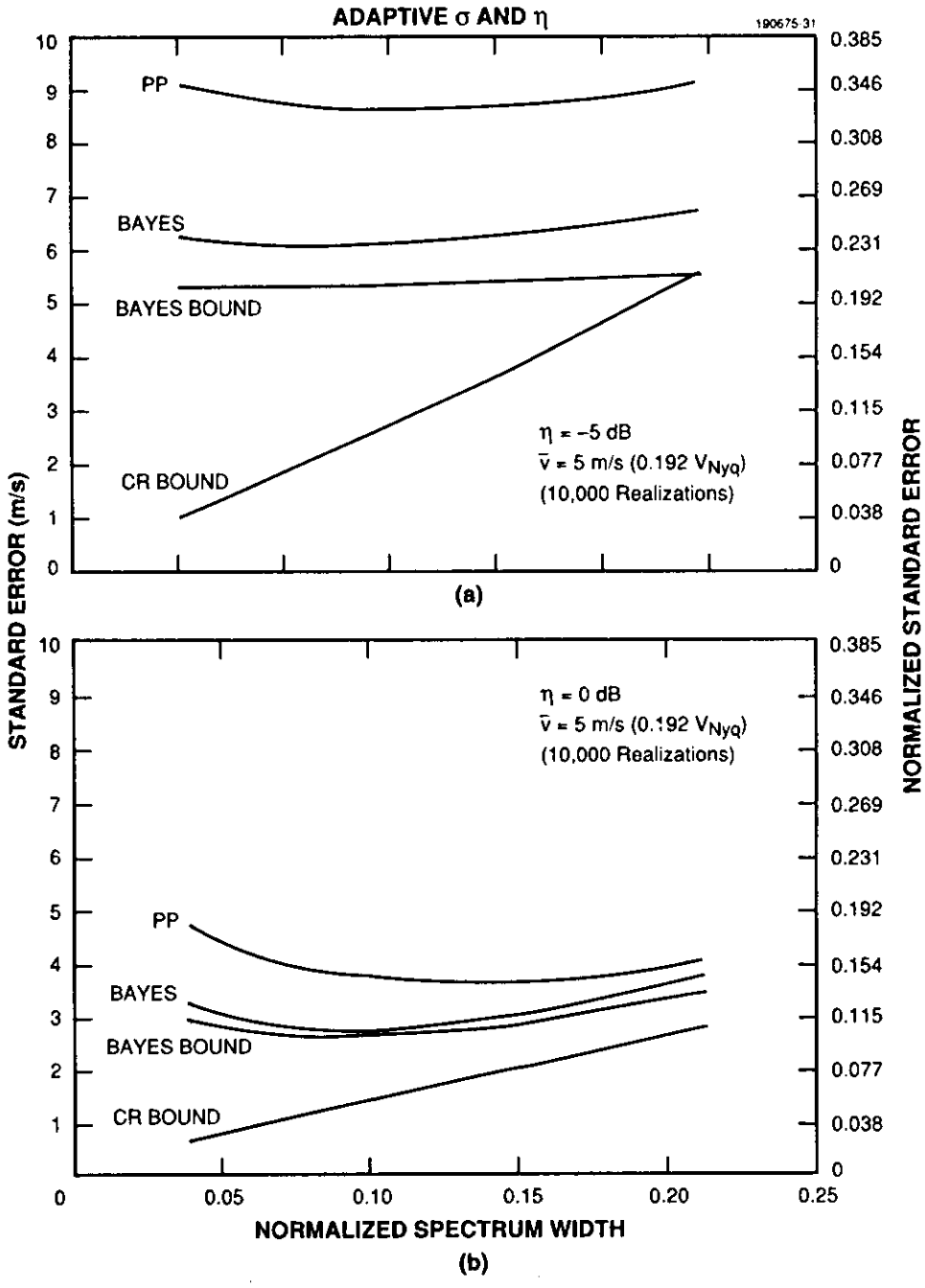

Figure 15. Bayes algorithm with adaptive σ and η: 15 compartment weighting coefficient set (II). (a) SNR = -5 dB. (b) SNR = 0 dB.

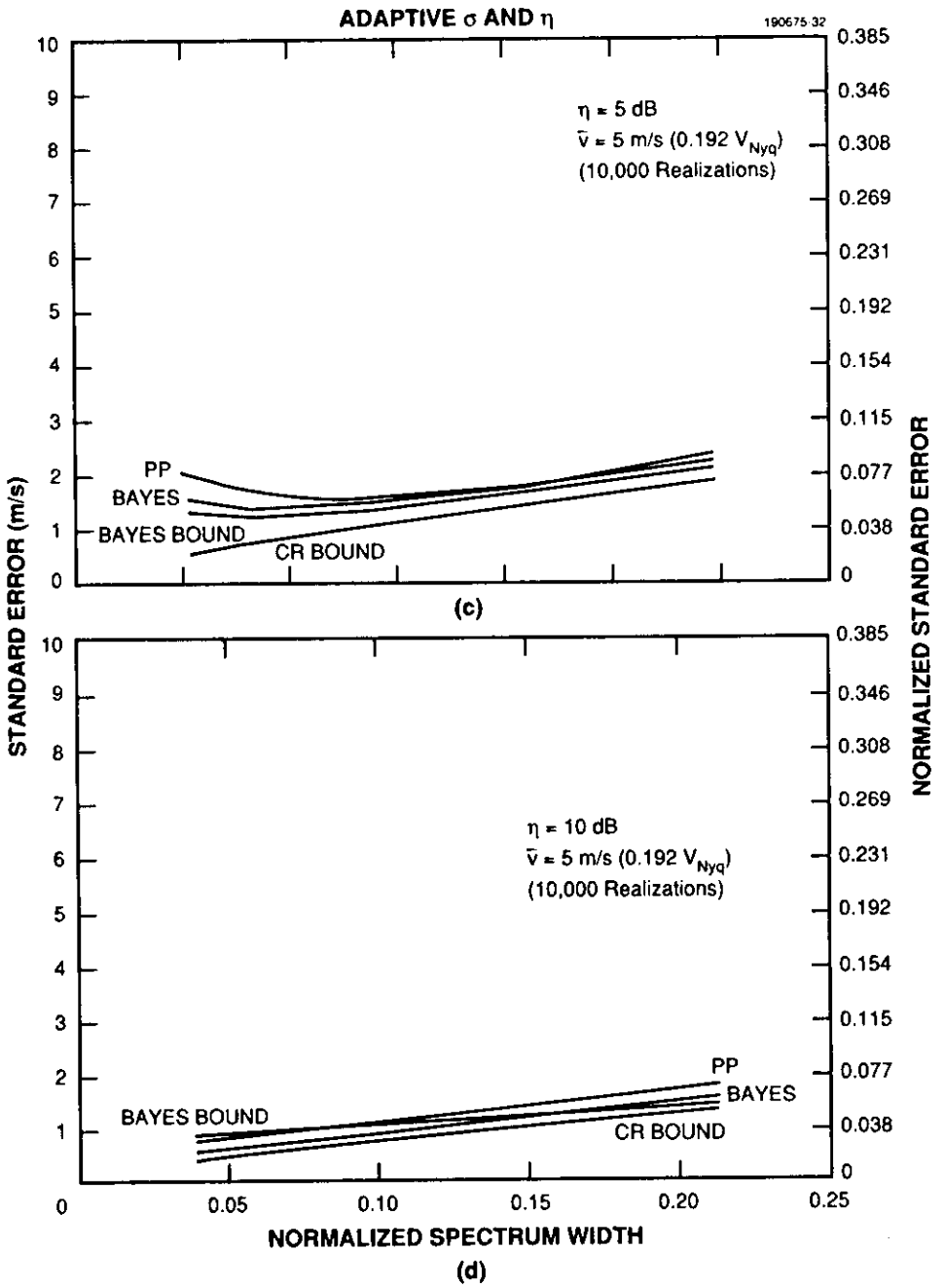

Figure 15. Bayes algorithm with adaptive σ and η: 15 compartment weighting coefficient set (II). (c) SNR = 10 dB. (d) SNR = 20 dB.

7. CONCLUSIONS

Each of the previous sections included a brief summary and these statements will not be repeated here. Some concluding remarks are, nevertheless, in order.

This report's primary objective has been to examine the challenge of Doppler velocity estimation when confronted with a small-sample size. Applying a generally accepted model for the measurement of (meteorologically generated) Doppler signals, lower bound performance limitations were first established. These bounds were shown to be clearly different (greater) than those provided by standard CR analysis, but room for improvement (relative to the standard PP estimator) was nevertheless indicated. Optimal velocity estimation, under the assumed model, by definition requires the joint estimation of a vector parameter. Previous attempts at such optimal estimation have been hampered by the technical difficulties that arise in solving the complicated system of equations that result. The potential of approximate methods that seek to treat some of the vector parameters as fixed quantities was explored. Sensitivity to these approximating assumptions was studied and insight into the relative importance of these nuisance parameters was provided. An adaptive method was proposed. The new method does not require excessive computations—nothing more extensive than previously proposed FFT methods. The final adaptive estimation scheme was shown to provide near optimal performance for a span of SNR and spectrum widths likely to be associated with meteorological signals.

APPENDIX A
Proof of Equation (4).

Proposition A.1 *Let* $\mathbf{Z}^T = [z_0\, z_1 \ldots z_{M-1}]$ *be a complex random vector with Gaussian density*

$$p(\mathbf{Z}|\Theta) = \pi^{-M}|\mathbf{R}|^{-1}e^{-\mathbf{Z}^\dagger\mathbf{R}^{-1}\mathbf{Z}},$$

where $\mathbf{R} = \mathbf{R}(\Theta)$ *is the* $M \times M$ *covariance matrix and* $\Theta^T = [\theta_0\, \theta_1 \ldots \theta_{n-1}]$ *is a real-valued parameter vector. Then, the Fisher Information matrix* $\mathbf{F} = [f_{i,j}]$ *defined by*

$$\mathbf{F} = E\left[\left(\frac{\partial \ln p(\mathbf{Z}|\Theta)}{\partial\Theta}\right)^T\left(\frac{\partial \ln p(\mathbf{Z}|\Theta)}{\partial\Theta}\right)\right]$$

can be obtained equivalently from

$$f_{i,j} = \mathrm{tr}\left\{\mathbf{R}^{-1}\frac{\partial\mathbf{R}}{\partial\theta_i}\mathbf{R}^{-1}\frac{\partial\mathbf{R}}{\partial\theta_j}\right\}. \tag{A.1}$$

Proof: Define the unit vectors of order n as

$$e_1 = \begin{bmatrix}1\\0\\\vdots\\0\end{bmatrix},\ e_2 = \begin{bmatrix}0\\1\\\vdots\\0\end{bmatrix},\ \ldots,\ e_n = \begin{bmatrix}0\\0\\\vdots\\1\end{bmatrix}$$

and the elementary matrix $E_{i,j}$ (of order $n \times n$) as

$$E_{i,j} = e_i e_j^T.$$

For an arbitrary matrix $A = [a_{i,j}]$, the following identities are standard (see Graham [10], for example):

$$a_{i,j} = e_i^T A e_j, \tag{A.2a}$$

$$A = \sum_i \sum_j a_{i,j} E_{i,j} = \sum_i \sum_j a_{i,j} e_i e_j^T, \tag{A.2b}$$

and

$$\mathrm{tr}A = \sum_i e_i^T A e_i. \tag{A.2c}$$

To prove Equation (A.1), begin by taking the logarithm of the density,

$$\ln p(\mathbf{Z}|\Theta) = -M \ln \pi - \ln |\mathbf{R}| - \mathbf{Z}^\dagger \mathbf{R}^{-1} \mathbf{Z},$$

and compute the derivative, term by term, with respect to Θ. Only the latter two terms are of importance. For the first of these, $\partial \ln |\mathbf{R}| / \partial \Theta$,

$$\begin{aligned}
\frac{\partial \ln |\mathbf{R}|}{\partial \theta_k} &= \sum_i \sum_j \frac{\partial \ln |\mathbf{R}|}{\partial r_{i,j}} \frac{\partial r_{i,j}}{\partial \theta_k} \\
&= \sum_i \sum_j \frac{1}{|\mathbf{R}|} \frac{\partial |\mathbf{R}|}{\partial r_{i,j}} \frac{\partial r_{i,j}}{\partial \theta_k} \\
&= \sum_i \sum_j r_{j,i}^{-1} \frac{\partial r_{i,j}}{\partial \theta_k} \quad \left\{ \text{since } \frac{\partial |\mathbf{R}|}{\partial \mathbf{R}} = |\mathbf{R}| \mathbf{R}^{-T} \right\} \\
&= \mathrm{tr} \left\{ \mathbf{R}^{-1} \frac{\partial \mathbf{R}}{\partial \theta_k} \right\},
\end{aligned} \tag{A.3}$$

where the last line follows from application of Equation (A.2):

$$\mathrm{tr} \left\{ \mathbf{R}^{-1} \frac{\partial \mathbf{R}}{\partial \theta_k} \right\} = \sum_i \sum_j e_j^T \mathbf{R}^{-1} e_i e_i^T \frac{\partial \mathbf{R}}{\partial \theta_k} e_j = \sum_i \sum_j r_{j,i}^{-1} \frac{\partial r_{i,j}}{\partial \theta_k}.$$

For the second term, note that

$$\mathbf{Z}^\dagger \mathbf{R}^{-1} \mathbf{Z} = \mathrm{tr}\{\mathbf{Z}^\dagger \mathbf{R}^{-1} \mathbf{Z}\} = \mathrm{tr}\{\mathbf{Z}\mathbf{Z}^\dagger \mathbf{R}^{-1}\}$$

from which it follows that

$$\frac{\partial \mathbf{Z}^\dagger \mathbf{R}^{-1} \mathbf{Z}}{\partial \Theta} = \frac{\partial \mathrm{tr}\{\mathbf{Z}\mathbf{Z}^\dagger \mathbf{R}^{-1}\}}{\partial \Theta}.$$

Continuing,

$$\begin{aligned}\frac{\partial \mathrm{tr}\{\mathbf{Z}\mathbf{Z}^\dagger\mathbf{R}^{-1}\}}{\partial\theta_k} &= \sum_i\sum_j \frac{\partial \mathrm{tr}\{\mathbf{Z}\mathbf{Z}^\dagger\mathbf{R}^{-1}\}}{\partial r_{i,j}}\frac{\partial r_{i,j}}{\partial\theta_k} \\ &= \sum_i\sum_j \mathrm{tr}\left\{\frac{\partial \mathbf{Z}\mathbf{Z}^\dagger\mathbf{R}^{-1}}{\partial r_{i,j}}\right\}\frac{\partial r_{i,j}}{\partial\theta_k} \\ &= \sum_i\sum_j -\mathrm{tr}\{\mathbf{Z}\mathbf{Z}^\dagger\mathbf{R}^{-1}E_{i,j}\mathbf{R}^{-1}\}\frac{\partial r_{i,j}}{\partial\theta_k},\end{aligned}$$

where the last line follows from the result

$$\frac{\partial \mathbf{R}^{-1}}{\partial r_{i,j}} = -\mathbf{R}^{-1}E_{i,j}\mathbf{R}^{-1}.$$

Additional rearrangement results in

$$\frac{\partial \mathrm{tr}\{\mathbf{Z}\mathbf{Z}^\dagger\mathbf{R}^{-1}\}}{\partial\theta_k} = -\mathrm{tr}\left\{\mathbf{R}^{-1}\frac{\partial \mathbf{R}}{\partial\theta_k}\mathbf{R}^{-1}\mathbf{Z}\mathbf{Z}^\dagger\right\}. \tag{A.4}$$

Combining Equations (A.3) and (A.4) provides the intermediate result

$$\frac{\partial \ln p(\mathbf{Z}|\Theta)}{\partial\theta_k} = \mathrm{tr}\left\{\mathbf{R}^{-1}\frac{\partial \mathbf{R}}{\partial\theta_k}\mathbf{R}^{-1}(\mathbf{Z}\mathbf{Z}^\dagger - \mathbf{R})\right\}. \tag{A.5}$$

The next step is to evaluate the expectation. From the definition of the Fisher Information matrix and Equation (A.5),

$$\begin{aligned}\mathbf{F}_{i,j} &= E\left[\frac{\partial \ln p(\mathbf{Z}|\Theta)}{\partial\theta_i}\frac{\partial \ln p(\mathbf{Z}|\Theta)}{\partial\theta_j}\right] \\ &= E\left[\mathrm{tr}\left\{\mathbf{R}^{-1}\frac{\partial \mathbf{R}}{\partial\theta_i}\mathbf{R}^{-1}(\mathbf{Z}\mathbf{Z}^\dagger - \mathbf{R})\right\}\mathrm{tr}\left\{\mathbf{R}^{-1}\frac{\partial \mathbf{R}}{\partial\theta_j}\mathbf{R}^{-1}(\mathbf{Z}\mathbf{Z}^\dagger - \mathbf{R})\right\}\right].\end{aligned}$$

Working with the terms inside the expectation and again using (A.2):

$$\mathrm{tr}\left\{\mathbf{R}^{-1}\frac{\partial \mathbf{R}}{\partial\theta_i}\mathbf{R}^{-1}(\mathbf{Z}\mathbf{Z}^\dagger - \mathbf{R})\right\}\mathrm{tr}\left\{\mathbf{R}^{-1}\frac{\partial \mathbf{R}}{\partial\theta_j}\mathbf{R}^{-1}(\mathbf{Z}\mathbf{Z}^\dagger - \mathbf{R})\right\}$$

$$= \operatorname{tr}\left\{\mathbf{R}^{-1}\frac{\partial \mathbf{R}}{\partial \theta_i}\mathbf{R}^{-1}(\mathbf{Z}\mathbf{Z}^\dagger - \mathbf{R})\right\}\operatorname{tr}\left\{(\mathbf{Z}\mathbf{Z}^\dagger - \mathbf{R})\mathbf{R}^{-1}\frac{\partial \mathbf{R}}{\partial \theta_j}\mathbf{R}^{-1}\right\}$$

$$= \sum_k\sum_l e_k^T\mathbf{R}^{-1}\frac{\partial \mathbf{R}}{\partial \theta_i}\mathbf{R}^{-1}(\mathbf{Z}\mathbf{Z}^\dagger - \mathbf{R})e_k\, e_l^T(\mathbf{Z}\mathbf{Z}^\dagger - \mathbf{R})\mathbf{R}^{-1}\frac{\partial \mathbf{R}}{\partial \theta_j}\mathbf{R}^{-1}e_l. \qquad \text{(A.6)}$$

Now,

$$(\mathbf{Z}\mathbf{Z}^\dagger - \mathbf{R})e_k e_l^T(\mathbf{Z}\mathbf{Z}^\dagger - \mathbf{R})$$

is an $M \times M$ matrix with nm^{th} entry given by

$$(z_n z_k^* - r_{n,k})(z_l z_m^* - r_{l,m})$$

for which the expectation is commonly known (see Miller [11], for example):

$$E\{(z_n z_k^* - r_{n,k})(z_l z_m^* - r_{l,m})\} = r_{n,m}r_{l,k}.$$

Therefore,

$$E\left\{(\mathbf{Z}\mathbf{Z}^\dagger - \mathbf{R})e_k e_l^T(\mathbf{Z}\mathbf{Z}^\dagger - \mathbf{R})\right\} = \mathbf{R}\, r_{l,k}. \qquad \text{(A.7)}$$

Finally, using Equations (A.6) and (A.7) one obtains

$$E\left\{\sum_k\sum_l e_k^T\mathbf{R}^{-1}\frac{\partial \mathbf{R}}{\partial \theta_i}\mathbf{R}^{-1}(\mathbf{Z}\mathbf{Z}^\dagger - \mathbf{R})e_k\, e_l^T(\mathbf{Z}\mathbf{Z}^\dagger - \mathbf{R})\mathbf{R}^{-1}\frac{\partial \mathbf{R}}{\partial \theta_j}\mathbf{R}^{-1}e_l\right\}$$

$$= \sum_k\sum_l e_k^T\mathbf{R}^{-1}\frac{\partial \mathbf{R}}{\partial \theta_i}\mathbf{R}^{-1}\frac{\partial \mathbf{R}}{\partial \theta_j}\mathbf{R}^{-1}e_l\, r_{l,k}$$

$$= \sum_k\sum_l \left[\mathbf{R}^{-1}\frac{\partial \mathbf{R}}{\partial \theta_i}\mathbf{R}^{-1}\frac{\partial \mathbf{R}}{\partial \theta_j}\mathbf{R}^{-1}\right]_{k,l} r_{l,k}$$

$$= \operatorname{tr}\left\{\mathbf{R}^{-1}\frac{\partial \mathbf{R}}{\partial \theta_i}\mathbf{R}^{-1}\frac{\partial \mathbf{R}}{\partial \theta_j}\right\},$$

which is the desired result.

REFERENCES

1. R.J. Doviak and D.S. Zrnić. *Doppler Radar and Weather Observations.* Academic Press, Inc., Orlando, Florida, 1984.
2. D.S. Zrnić. Simulation of weatherlike Doppler spectra and signals. *Journal of Applied Meteorology*, 14:619–620, 1975.
3. H.L. Van Trees. *Detection, Estimation, and Modulation Theory: Part I.* John Wiley and Sons, Inc., New York, New York, 1968.
4. P.R. Mahapatra and D.S. Zrnić. Practical algorithms for mean velocity estimation in pulse Doppler weather radars using a small number of samples. *IEEE Transactions on Geoscience and Remote Sensing*, GE-21(4):491–501, 1983.
5. K.S. Miller and M.M. Rochwarger. A covariance approach to spectral moment estimation. *IEEE Transactions on Information Theory*, IT-18(5):588–596, 1972.
6. P.T. May and R.G. Strauch. An examination of wind profiler signal processing algorithms. *Journal of Atmospheric and Oceanic Technology*, 6:731–735, 1989.
7. S. Kullback. *Information Theory and Statistics.* John Wiley and Sons, New York, New York, 1959.
8. P.E. Gill, W. Murray, and M.H. Wright. *Practical Optimization.* Academic Press, Inc., New York, New York, 1981.
9. D.S. Zrnić. Estimation of spectral moments for weather echoes. *IEEE Transactions on Geoscience Electronics*, GE-17(4):113–128, 1979.
10. A. Graham. *Kronecker Products and Matrix Calculus with Applications.* Halsted Press: John Wiley and Sons, New York, New York, 1981.
11. K.S. Miller. Complex gaussian processes. *SIAM Review*, 11(4):544–567, 1969.